2018

Estimation of a Partially Linear Regression in Triangular Systems

Xin Geng
IFPRI, x.geng@cgiar.org

Carlos Martins-Filho
University of Colorado, carlos.martins@colorado.edu

Feng Yao
West Virginia University, Feng.Yao@mail.wvu.edu

Department of Economics

Working Paper Series

Estimation of a Partially Linear Regression in Triangular Systems

Xin Geng

Carlos Martins-Filho

Feng Yao

Working Paper No. 18-05

ESTIMATION OF A PARTIALLY LINEAR REGRESSION IN TRIANGULAR SYSTEMS

XIN GENG*

IFPRI
2033 K Street NW
Washington, DC 20006-1002, USA
email: x.geng@cgiar.org
Voice: + 1 202 862 4636

CARLOS MARTINS-FILHO

Department of Economics
University of Colorado
Boulder, CO 80309-0256, USA
email: carlos.martins@colorado.edu
Voice: + 1 303 492 4599

&

IFPRI
2033 K Street NW
Washington, DC 20006-1002, USA
email: c.martins-filho@cgiar.org
Voice: + 1 202 862 8144

FENG YAO

Department of Economics
West Virginia University
Morgantown, WV 26505, USA
email: feng.yao@mail.wvu.edu
Voice: +1 304 293 7867

&

School of Economics and Trade
Guangdong University of Foreign Studies
Guangzhou, Guangdong 510006, P. R. China
email: 201470006@oamail.gdufs.edu.cn
Voice: + 86 20 3932 8858

March, 2018

Abstract. We propose kernel-based estimators for the components of a partially linear regression in a triangular system where endogenous regressors appear both in the linear and nonparametric components of the regression. Compared with other estimators currently available in the literature, e.g. the sieve estimators proposed in Ai and Chen (2003) or Otsu (2011), our estimators have explicit functional form and are much easier to implement. They rely on a set of assumptions introduced by Newey et al. (1999) that characterize what has become known as the "control function" approach for endogeneity in regression. We explore conditional moment restrictions that make this model suitable for additive regression estimation as in Kim et al. (1999) and Manzan and Zerom (2005). We establish consistency and $\sqrt{n}$ asymptotic normality of the estimator for the parameters in the linear component of the model, give a uniform rate of convergence, and establish the asymptotic normality for the estimator of the nonparametric component. In addition, for statistical inference, a consistent estimator for the covariance of the limiting distribution of the parametric estimator is provided. A small Monte Carlo study sheds light on the finite sample performance of our estimators and an empirical application illustrates their use.

Keywords. partially linear regression; endogeneity; semiparametric instrumental variable estimation.

JEL Classifcations. C14, C36.

AMS-MS Classifications. Primary: 62G07, 62G08, 62G20.

---

*We would like to thank seminar and conference participants at the European, China, and Asian Meetings of the Econometric Society, 3rd Conference of the International Society for Non-Parametric Statistics, University of Colorado at Boulder and IFPRI, for their helpful comments and suggestions. All remaining errors are ours.

1 Introduction

There exists a rapidly growing literature on the specification and estimation of semiparametric and nonparametric regression models with endogenous regressors.[1] This results from the understanding that fully specified parametric models generally lead to inconsistent estimators and faulty inference due to a high probability of model misspecification. In addition, the problem of regressor endogeneity is widely encountered in, but not restricted to, empirical models in Economics, mostly due to measurement error, omitted regressors, or simultaneity that arises in agents' optimization problems or the characterization of market equilibrium. Identification and estimation of these models have been conducted under two broad approaches: the instrumental variable (IV) approach (see, e.g., Newey and Powell, 2003; Ai and Chen, 2003; Otsu, 2011) or the control function (CF) approach (see, e.g., Newey et al., 1999; Pinkse, 2000; Blundell and Powell, 2003; Su and Ullah, 2008; Martins-Filho and Yao, 2012). As discussed in Newey et al. (1999) and Blundell et al. (2013) the desirability of these approaches rests on the suitability of different, and generally non-nested sets of assumptions, rendering their choice largely dependent on the specific stochastic framework encountered by the researcher.

It is now well known that following the IV approach is made difficult by the fact that, in this case, the nonparametric IV regression is typically an ill-posed problem, leading to estimators that converge at slower rates when compared to those obtained in the absence of endogeneity (see, e.g., Hall and Horowitz, 2005; Darolles et al., 2011; Chen and Christensen, 2015). In addition, computation of these estimators is numerically difficult due to the fact that they cannot be expressed by closed form algebraic expressions (see, e.g., Ai and Chen, 2003; Otsu, 2011). Alternatively, following the CF approach normally leads to multi-stage estimation procedures, where nonparametric generated regressors make it difficult to asymptotically characterize final stage estimators for both finite and infinite parameters of interest (see, e.g., Newey et al., 1999; Pinkse, 2000; Hahn and Ridder, 2013).

In this paper, we contribute to the CF approach by considering the estimation of a partially linear regression model where endogenous regressors appear in both the finite and infinite dimensional components of the model. Our proposed estimators are all kernel based and, therefore, easy to implement from a computational perspective. In addition, we obtain their consistency, give their asymptotic distributions, and provide estimators for their variances, allowing for easy asymptotically based inference. Specifically, we consider the following partially linear triangular model,

$$Y_i = \beta_0 + X_{2i}'\beta + m(X_{1i}, Z_{1i}) + \varepsilon_i, \tag{1}$$

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

$$\mathrm{E}(\varepsilon_i) = 0, \quad \mathrm{E}(U_i|Z_i) = 0, \quad \mathrm{E}(\varepsilon_i|Z_i, U_i) = \mathrm{E}(\varepsilon_i|U_i), \quad \text{for } i = 1, \cdots, n. \tag{3}$$

[1] See Chen and Qiu (2016) for a comprehensive review of the existing literature.

Y_i is a scalar regressand, $Z_{1i} \in \mathbb{R}^{D_{11}}$ is a subvector of $Z_i = (Z_{1i}', Z_{2i}')' \in \mathbb{R}^{D_1}$ with $D_1 = D_{11} + D_{12}$, X_{1i}, X_{2i} are non-overlapping subvectors of $X_i \in \mathbb{R}^{D_2}$ of dimensions D_{21} and D_{22} with $D_2 = D_{21} + D_{22}$ and ε_i is an unobserved scalar random error. The variables X_i are taken to be endogenous in that $\mathrm{E}(\varepsilon_i|X_i) \neq 0$, and the variables Z_i are exogenous as a result of the moment conditions in (3). U_i is a vector of unobserved random errors and $\Pi : \mathbb{R}^{D_1} \to \mathbb{R}^{D_2}$ is an unknown nuisance function. Our primary interest is in the estimation of the finite dimensional parameters $(\ \beta_0 \quad \beta'\ )$ and the infinite dimensional parameter $m(\cdot)$ in Equation (1). The partially linear structure of this equation reflects the often assumed linearity with respect to some of the regressors while retaining the flexibility of a nonparametric structure for other components of the regression. See, for motivation, Robinson (1988), Speckman (1988) and Härdle et al. (2000). For another parsimonious semiparametric specification see the single-index model in Birke et al. (2017).

Newey et al. (1999) proposed series estimators (power and splines) for a model where there is no intercept in equation (1), i.e., $\beta_0 = 0$ and the partially linear structure in (1) is generically modeled as $g(X_i, Z_{1i})$.[2] Otherwise, their model is identical to ours. The fact that in their case $\beta_0 = 0$ permits the relaxation of the assumption that $E(\varepsilon_i) = 0$, and given that our partially linear structure is a restriction on g, their estimation method can be adapted to the model described by (1)–(3) (see Section 6 of their paper). In Section 3 of this paper, we contrast the additional assumptions they make to characterize some of the asymptotic behavior of their estimators with those we make to obtain similar results.

Recently, Martins-Filho and Yao (2012) proposed kernel-based estimators for $(\ \beta_0 \quad \beta'\ )$ and $m(\cdot)$, but although their estimators appear to have good finite-sample properties, they have failed to provide a characterization of their asymptotic behavior. In fact, our theoretical work suggests that their estimators cannot be shown to be asymptotically normally distributed under standard parametric and nonparametric normalizations, respectively (see details given in Section 2). Alternatively, to our knowledge, besides the estimators proposed by Newey et al. (1999), there exist two estimation procedures following the IV approach that can be used to estimate the parameters in Equation (1): the sieve minimum distance estimator of Ai and Chen (2003) and the sieve conditional empirical likelihood estimator of Otsu (2011). These estimators are based on the moment condition $\mathrm{E}(\varepsilon_i|Z_i) = 0$, which is different from those given in (3). As mentioned above, strictly speaking, neither their condition nor the ones given in (3) are stronger than the other (see Newey et al., 1999). However, under the additional restrictions that U_i is independent of Z_i and $\mathrm{E}(\varepsilon_i) = 0$, the moment restrictions in (3) imply that $\mathrm{E}(\varepsilon_i|Z_i) = 0$, making the estimators developed in their papers suitable for our model.[3]

As will be shown in Section 2 and 3, the estimators proposed in this paper have a number of desirable characteristics. First, the estimator for the linear components of the semiparametric regression model given in (1) are $\sqrt{n}$ asymptotic normal. Second, we provide consistency, give the uniform convergence rate, and establish asymptotic normality, under standard nonparametric normalization, for the estimator of the nonparametric component in (1). In addition, we provide a consistent estimator for the covariance of the limiting distribution of the parametric estimator,

[2] Ozabaci et al. (2014) also considered a model similar to that in Newey et al. (1999), but in their formulation $\Pi(Z_i)$, $g(X_i, Z_{1i})$ and $E(\varepsilon_i|U_i)$ are all additive nonparametric functions of each of their arguments.

[3] It should be noted that the estimators of Ai and Chen (2003) and Otsu (2011) apply to more general models than ours, since their use is not constrained to the partially linear regression under the control function structure we adopt.

making our results directly usable for inference.

From a technical perspective, the results in this paper can be viewed as extensions of the asymptotic normality results of Manzan and Zerom (2005) to the case of a partially linear regression model with generated regressors appearing in the parametric and nonparametric component. In this sense, our work is also related to Li and Wooldridge (2002). Although the estimation procedure we consider is conceptually simple and easy to implement, its asymptotic characterization is non-trivial, requiring repeated analysis of U-statistics of high degree. This has been greatly facilitated by results in Yao and Martins-Filho (2015), which are used frequently in our proofs. The ancillary results required to obtain our theorems are, to our knowledge, novel and can be used in other contexts where generated regressors are encountered in various types of two stage kernel based estimators.

The rest of this paper is organized as follows. Section 2 describes the model in greater detail, considers identification and the moment conditions used in estimation, and provides a detailed algorithm for estimation. Section 3 gives asymptotic characterizations for our estimators and the assumptions we used to obtain our results. Where appropriate, we contrast our assumptions with those in Newey et al. (1999). Section 4 contains a small Monte Carlo study that sheds some light on the finite sample performance of our estimators and contrasts them to the series estimator proposed by Newey et al. (1999). Section 5 gives an empirical application using our methods to study the aid-policy-growth relationship, which has been the subject of much work in the Economic Development literature. Section 6 concludes. All proofs are given in the Appendix.

2 Moment conditions, identification, and estimation

2.1 Moment Conditions

We start by deriving a collection of conditional moments that emerge from the model described by equations (1)–(3). They are the bases for the estimators we propose in section 2.2. Given equations (2) and (3), we have that $\mathrm{E}(\varepsilon_i|X_{1i}, Z_i,U_i) = \mathrm{E}(\varepsilon_i|Z_i,U_i) = \mathrm{E}(\varepsilon_i|U_i)$, and $\mathrm{E}(X_{2i}|X_{1i},Z_i,U_i) = \mathrm{E}(X_{2i}|Z_i,U_i) = X_{2i}$. Letting $g(U_i) \equiv \mathrm{E}(\varepsilon_i|U_i) : \mathbb{R}^{D_2} \to \mathbb{R}$, and using (1), we can write

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

Letting $v_i = Y_i - \mathrm{E}(Y_i|X_{1i},Z_i,U_i)$, we have

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

where, by construction, $\mathrm{E}(v_i|Z_i,U_i) = 0$. Note that if β_0 and β were known, and U_i were observed, (5) could be viewed as an additive nonparametric regression model, with regressand $Y_i - \beta_0 - X_{2i}'\beta$. As is common in the additive nonparametric literature (see, *inter alia*, Linton and Härdle, 1996, Kim et al., 1999, Martins-Filho and Yang, 2007), we assume that $\mathrm{E}\left(m(X_{1i},Z_{1i})\right)=\mathrm{E}\left(g(U_i)\right)=0$, since each component in an additive nonparametric model can only be identified up to an additive constant.[4]

Using a suitable "instrument" function, we now obtain moment conditions that motivate our estimator for β_0 and β. For simplicity, in what follows, we put $W_i=(X'_{1i},Z'_{1i})'$. As in Kim et al. (1999), we define the "instrument" function as $\eta_i=\eta(W_i,U_i)\equiv\frac{f_W(W_i)f_U(U_i)}{\phi(W_i,U_i)}$, where f_W is the joint marginal density of elements in W_i, f_U the marginal density of U_i, and ϕ the joint density of W_i and U_i. Note that $\mathrm{E}\big(\eta(W_i,U_i)|W_i\big)=1$, $\mathrm{E}\big(\eta(W_i,U_i)g(U_i)|W_i\big)=0$, $\mathrm{E}\big(\eta(W_i,U_i)|U_i\big)=1$ and $\mathrm{E}\big(\eta(W_i,U_i)m(W_i)|U_i\big)=0$. By pre-multiplying both sides of (5) by η_i, and taking conditional expectations given W_i and U_i we have, respectively,

$$\mathrm{E}\left(\eta_i(Y_i-X'_{2i}\beta-\beta_0)\,|\,W_i\right)=m(W_i),\qquad \mathrm{E}\left(\eta_i(Y_i-X'_{2i}\beta-\beta_0)\,|\,U_i\right)\;=\;g(U_i). \tag{6}$$

It is apparent that if β_0 and β were known, and U_i were observed, $m(W_i)$ and $g(U_i)$ could be estimated based on the moment conditions (6) using an estimated sequence $\{\hat{\eta}_i\}_{i=1}^n$ constructed with nonparametric density estimators of f_W, f_U and ϕ evaluated at all data points. To address the fact that β_0 and β are unknown, note that $m(W_i)$ and $g(U_i)$ can be expressed as conditional expectations containing β, β_0 in (6). Substituting them back into (5) and rearranging, with $\beta_0=\mathrm{E}(\eta_i(Y_i-X'_{2i}\beta))$, we have

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

where $Y_i^*\equiv Y_i-\mathrm{E}(\eta_iY_i|W_i)-\mathrm{E}(\eta_iY_i|U_i)+\mathrm{E}(\eta_iY_i)$, and $X_{2i}^*\equiv X_{2i}-\mathrm{E}(\eta_iX_{2i}|W_i)-\mathrm{E}(\eta_iX_{2i}|U_i)+\mathrm{E}(\eta_iX_{2i})$.

It is important to note that Equation (7) provides infinitely many moment conditions to estimate β, since by pre-multiplying by any arbitrary measurable function $L(X_{1i},Z_i,U_i)$, we still have $\mathrm{E}\Big(L(X_{1i},Z_i,U_i)\,v_i\Big|X_{1i},Z_i,U_i\Big)=0$. Here, $L(X_{1i},Z_i,U_i)$ can be treated as a normalizing factor that should be suitably chosen to derive the asymptotic properties of an estimator for β. In our case, we choose $L(X_{1i},Z_i,U_i)=\sqrt{\eta_i}$, and consider

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

Letting $Y=(Y_1,\cdots,Y_n)'$, $X=(X_1,\cdots,X_n)'$, $Z=(Z_1,\cdots,Z_n)'$, we write $\sqrt{\eta}\,Y^*=\sqrt{\eta}\,X_2^*\beta+\sqrt{\eta}\,v$, where $Y^*=(Y_1^*,\cdots,$

[4] As in Robinson (1988), we note that $\mathrm{E}\left(m(X_{1i},Z_{1i})\right)=0$ can be relaxed if we set $\beta_0=0$.

$Y_n^*)'$, $X_2^* = (X_{21}^*, \cdots, X_{2n}^*)'$, $v = (v_1, \cdots, v_n)'$, $\sqrt{\eta} = \text{diag}\{\sqrt{\eta_i}\}_{i=1}^n$, and $\text{E}\big(\sqrt{\eta_i}\, v_i \big| X_{1i}, Z_i, U_i\big) = 0$. Note that since $\beta_0 = \text{E}\big(\eta_i(Y_i - X_{2i}'\beta)\big)$ and given $L(X_{1i}, Z_i, U_i) = \sqrt{\eta_i}$, we have $\text{E}(\eta_i Y_i^* | W_i) = \text{E}(\eta_i Y_i^* | U_i) = \text{E}(\eta_i X_{2i}^* | W_i) = \text{E}(\eta_i X_{2i}^* | U_i) = 0$. The choice of $L(\cdot)$ is critical in establishing the asymptotic properties of our estimators of β_0, β, and $m(\cdot)$. Besides using different estimators for the conditional expectations in Y_i^* and X_{2i}^*, Martins-Filho and Yao (2012) failed to suggest, or understand, the role of $L(\cdot)$ in obtaining asymptotic properties of the kernel-based estimators for this model. In fact, a more careful investigation of the consequences of choosing such a normalizing function in establishing the asymptotic properties of estimators for β_0, β, and $m(\cdot)$ remains an open and important topic of study, as it also has a direct impact on the structure of the variances of their asymptotic distributions.

We denote the additive components in Y_i^*, X_{2i}^* and the corresponding error terms by $m_1(W_i) \equiv \text{E}(\eta_i Y_i | W_i)$, $m_2(W_i) \equiv \text{E}(\eta_i X_{2i} | W_i)$, $m_3(W_i) \equiv \text{E}(\eta_i | W_i) = 1$, $g_1(U_i) \equiv \text{E}(\eta_i Y_i | U_i)$, $g_2(U_i) \equiv \text{E}(\eta_i X_{2i} | U_i)$, $g_3(U_i) \equiv \text{E}(\eta_i | U_i) = 1$, $\mu_1 \equiv \text{E}(\eta_i Y_i)$, $\mu_2 \equiv \text{E}(\eta_i X_{2i})$, $v_{m1i} \equiv \eta_i Y_i - m_1(W_i)$, $v_{m2i} \equiv \eta_i X_{2i} - m_2(W_i)$, $v_{m3i} \equiv \eta_i - 1$, $v_{g1i} \equiv \eta_i Y_i - g_1(U_i)$, $v_{g2i} \equiv \eta_i X_{2i} - g_2(U_i)$, and $v_{g3i} \equiv \eta_i - 1$. Given the moment condition associated with $m(W_i)$ in Equation (6), we let $v_{mi} \equiv \eta_i(Y_i - X_{2i}'\beta - \beta_0) - m(W_i) = v_{m1i} - v_{m2i}'\beta - v_{m3i}\beta_0$.

The regressors $\sqrt{\eta_i} X_{2i}^*$ in Equation (8) satisfy $\text{E}\big(\sqrt{\eta_i} X_{2i}^* v_i\big) = 0$, suggesting an estimator of β that is obtained by inserting estimators of $\sqrt{\eta_i} Y_i^*$ and $\sqrt{\eta_i} X_{2i}^*$ prior to an application of a standard rule, such as no-intercept ordinary least squares (OLS) method. Note that by (6), we have $m(W_i) = m_1(W_i) - m_2'(W_i)\beta - m_3(W_i)\beta_0$, and $g(U_i) = g_1(U_i) - g_2'(U_i)\beta - g_3(U_i)\beta_0$. Thus, to estimate Y_i^*, X_{2i}^*, $m(W_i)$, and $g(U_i)$, we need only estimate each of their additive components separately. The main technical difficulty rests in the fact that U_i must be substituted by a generated regressor $\hat{U}_i$ in the estimation of all conditional moments involving U_i and η_i. Kernel-based nonparametric regression estimators are employed throughout this paper, and for identification purposes, existence and nonsingularity of $\Phi_0 \equiv \text{E}\big(\eta_i X_{2i}^* X_{2i}^{*\prime}\big)$ needs to be assumed.

2.2 Estimation

Based on the moment conditions given in section 2.1, we now describe in detail our proposed estimation procedure. Since U_i is not observed, the first step in the estimation generates $\hat{U}_i$. We obtain a Nadaraya-Watson (NW) estimator for $\Pi(Z_i)$ from (2), with the j^{th} element defined as

$$\hat{\Pi}_j(Z_i) = \operatorname*{argmin}_{\theta} \frac{1}{nh_1^{D_1}} \sum_{t=1}^{n} (X_{t,j} - \theta)^2 K_1\left(\frac{Z_t - Z_i}{h_1}\right) \text{ for } j = 1, \cdots, D_2,$$

where $X_{t,j}$ is the j^{th} element of X_t, $h_1 > 0$ is the associated bandwidth, and $K_1 \colon \mathbb{R}^{D_1} \to \mathbb{R}$ is a multivariate kernel function. To associate the relevant subvector of $\Pi(Z_i)$ with X_{2i}, we define $\Pi(Z_i) \equiv (\Pi_1'(Z_i), \Pi_2'(Z_i))'$, where $\Pi_2(Z_i) \equiv (\Pi_{21}(Z_i), \cdots, \Pi_{2D_{22}}(Z_i))' = X_{2i} - U_{2i}$. $\Pi_1(Z_i)$ is defined similarly. Denote the estimates by $\hat{\Pi}(Z_i) = \left(\hat{\Pi}_1'(Z_i), \hat{\Pi}_2'(Z_i)\right)' \equiv \left(\hat{\Pi}_1(Z_i), \cdots, \hat{\Pi}_{D_2}(Z_i)\right)'$ and calculate the nonparametric residuals $\hat{U}_i \equiv \left(\hat{U}_{i1}, \cdots, \hat{U}_{iD_2}\right)'$, where $\hat{U}_{ij} \equiv X_{i,j} - \hat{\Pi}_j(Z_i)$, for $j = 1, \cdots, D_2$ and $i = 1, \cdots, n$.

In the second step, we estimate η_i (instrument functions) from section 2.1 using W_t, and the generated regressors $\hat{U}_t$ obtained in the first step. We first obtain Rosenblatt-Parzen density estimators for f_U, f_W, and ϕ:

$$\begin{aligned}
\hat{f}_U(u) &= \frac{1}{nh_2^{D_2}} \sum_{t=1}^{n} K_2\left(\frac{\hat{U}_t - u}{h_2}\right), \qquad \hat{f}_W(w) = \frac{1}{nh_3^{D_3}} \sum_{t=1}^{n} K_3\left(\frac{W_t - w}{h_3}\right), \\
\hat{\phi}(w,u) &= \frac{1}{nh_4^{D_4}} \sum_{t=1}^{n} K_4\left(\frac{(W_t' \quad \hat{U}_t')' - (w' \quad u')'}{h_4}\right),
\end{aligned}$$

where $K_2 \colon \mathbb{R}^{D_2} \to \mathbb{R}$, $K_3 \colon \mathbb{R}^{D_3} \to \mathbb{R}$, and $K_4 \colon \mathbb{R}^{D_4} \to \mathbb{R}$ are multivariate kernel functions, $D_3 \equiv D_{11} + D_{21}$, $D_4 \equiv D_2 + D_3$, and $h_i > 0$ is the associated bandwidth for $i = 2, 3, 4$. Then, a natural estimator for η_i is $\hat{\eta}_i = \hat{\eta}(W_i, \hat{U}_i) \equiv \frac{\hat{f}_W(W_i)\hat{f}_U(\hat{U}_i)}{\hat{\phi}(W_i, \hat{U}_i)}$.

In the third step we obtain NW estimators for the conditional expectations in the expressions for Y_i^*, X_{2i}^* as follows:

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

Estimators of the unconditional expectations μ_1 and μ_2 are given by $\hat{\mu}_1 = \frac{1}{n}\sum_{t=1}^{n} \hat{\eta}_t Y_t$, and $\hat{\mu}_2 = \frac{1}{n}\sum_{t=1}^{n} \hat{\eta}_t X_{2t}$. Thus, we define estimators of Y_i^* and X_{2i}^* respectively as $\hat{Y}_i = Y_i - \hat{m}_1(W_i) - \hat{g}_1(\hat{U}_i) + \hat{\mu}_1$, $\hat{X}_{2i} = X_{2i} - \hat{m}_2(W_i) - \hat{g}_2(\hat{U}_i) + \hat{\mu}_2$, for $i = 1, \cdots, n$.

In the fourth step, using the estimators $\hat{\eta}_i$, $\hat{Y}_i$, and $\hat{X}_{2i}$ derived in the previous steps, instead of η_i, Y_i^*, and X_{2i}^* in

(8), we have a feasible no-intercept OLS estimator of β:

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

where $\hat{Y} = (\hat{Y}_1, \cdots, \hat{Y}_n)'$, $\hat{X}_2 = (\hat{X}_{21}, \cdots, \hat{X}_{2n})'$, and $\hat{\eta} = \text{diag}\{\hat{\eta}_i\}_{i=1}^n$. Given that $\beta_0 = \text{E}(Y_i - X_{2i}'\beta)$ and the estimator $\hat{\beta}$, an estimator of β_0 is $\hat{\beta}_0 = \bar{Y} - \bar{X}_2'\hat{\beta}$, where $\bar{Y} \equiv \frac{1}{n}\sum_{t=1}^n Y_t$, and $\bar{X}_2 \equiv \frac{1}{n}\sum_{t=1}^n X_{2t}$.

Finally, the last step provides an estimator for m. Given Equation (6) and the estimators $\hat{\beta}_0$ and $\hat{\beta}$, we propose the following estimators have for $m(W_i)$ and $g(U_i)$,

$$\hat{m}(W_i) = \hat{m}_1(W_i) - \hat{m}_2'(W_i)\hat{\beta} - \hat{m}_3(W_i)\hat{\beta}_0, \qquad \hat{g}(\hat{U}_i) = \hat{g}_1(\hat{U}_i) - \hat{g}_2'(\hat{U}_i)\hat{\beta} - \hat{g}_3(\hat{U}_i)\hat{\beta}_0, \tag{11}$$

where $\hat{m}_3(W_i)$ and $\hat{g}_3(\hat{U}_i)$ are NW estimators for $m_3(W_i)$ and $g_3(U_i)$ defined similarly as $\hat{m}_1(W_i)$ and $\hat{g}_1(\hat{U}_i)$ in (9) except that $\hat{\eta}_t$ is used, instead of $\hat{\eta}_t Y_t$, as regressand.

3 Asymptotic characterizations of $\hat{\beta}$ and $\hat{m}(\cdot)$

In this section, we study the asymptotic properties of the estimators $\hat{\beta}$ and $\hat{m}(\cdot)$ defined 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 $\{\hat{U}_i\}_{i=1}^n$. Third, we establish $\sqrt{n}$ asymptotic normality of $\hat{\beta} - \beta$. Lastly, we use the asymptotic normality of $\sqrt{n}(\hat{\beta} - \beta)$ to establish the asymptotic distribution of $\hat{m}$ under suitable centering and normalization.

3.1 Assumptions

First we provide a list of general assumptions that will be adopted in our theorems and introduce notation. In what follows, C denotes a generic constant in $(0, \infty)$ 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. k_i is symmetric about zero, 4-times continuously differentiable and satisfies: a) $\int k_i(x)\text{d}x = 1$; b) $|k_i^{(j)}(x)||x|^{5+a} \to$

0 as $|x| \to \infty$, $j = 0, \cdots, 4$, for some $a > 0$; c) k_i is a kernel of order s_i, i.e., $\int k_i(x) x^j \mathrm{d}x = 0$ for $j = 1, \cdots, s_i - 1$, and $\int |k_i(x)||x|^{s_i} \mathrm{d}x < C$. We let $s \equiv \max\{s_i\}_{i=1}^4$ and $\mu_{k_i,s_i} \equiv \int k_i(x) x^{s_i} \mathrm{d}x$.

Our use of "higher-order" kernels is needed to attain suitable orders for the biases of our nonparametric estimators. Since global differentiability of the kernel functions is required in using Taylor's Theorem, in the following theorems, kernels that have compact support are excluded. It is easy to construct kernels that satisfy the conditions in A1. For example, kernels of even order $s \geq 2$, can be defined as

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

where $\phi(x) = (2\pi)^{-1/2} \exp(-\frac{1}{2}x^2)$ for suitably chosen c_j. In particular, given that we can evaluate the moments $m_{2j} = \int x^{2j} \phi(x) \mathrm{d}x$, $0 \leq j \leq \frac{1}{2}(s-2)$, we choose $\{c_j\}_{j=0}^{\frac{1}{2}(s-2)}$ that satisfy the linear system of $s/2$ simultaneous equations $\sum_{j=0}^{\frac{1}{2}(s-2)} c_j m_{2(i+j)} = \delta_{i0}$, $0 \leq i \leq \frac{1}{2}(s-2)$, where δ_{i0} is Kronecker's delta. For example, $k_2(x) = \phi(x)$, $k_4(x) = \left(\frac{3}{2} - \frac{1}{2}x^2\right) \phi(x)$ and $k_6(x) = \left(\frac{15}{8} - \frac{5}{4}x^2 + \frac{1}{8}x^4\right) \phi(x)$. Note that these kernels are continuously differentiable of any order everywhere, and when multiplied by any polynomial function they are all uniformly bounded and absolutely integrable, as their tails decay exponentially. We show in Lemma 1 that product kernels satisfying A1 are locally Lipschitz continuous, which is necessary for Lemma 3.

Assumption A2. The components of the sequence $\{(X_i', Z_i', Y_i)\}_{i=1}^n$ of random vectors by described in (1) - (3) are independent and identically distributed (IID) random vectors. The density functions $f_W(W_i)$, $f_Z(Z_i)$, $\phi(W_i, U_i)$, $f_{UZ}(U_i, Z_i)$ and $f_U(U_i)$ are uniformly bounded away from zero and infinity on arbitrary convex compact subsets of their domains. Here, $f_{UZ}(\cdot)$ is the joint density function of (U_i, Z_i).

The existence, boundedness properties and compactness of the support of the densities in assumption A2 are common regularity conditions imposed to derive properties of kernel based nonparametric estimators and largely overlap with Assumption 2 in Newey et al. (1999).

Assumption A3. (i) $\mathrm{E}\big(m(W_i)\big) = \mathrm{E}\big(g(U_i)\big) = 0$, (ii) $\mathrm{E}\big(v_i^2 \big| Z_i, U_i\big) = \sigma_v^2 < \infty$, $\mathrm{E}\big(U_{ij}^2 \big| Z_i\big) = \sigma_{Uj}^2 < \infty$, $\mathrm{E}\big(v_{m1i}^2 \big| W_i\big) = \sigma_{vm1}^2 < \infty$, $\mathrm{E}\big(v_{m2i,j}^2 \big| W_i\big) = \sigma_{vm2}^2 < \infty$, $\mathrm{E}\big(v_{g1i}^2 \big| U_i\big) = \sigma_{vg1}^2 < \infty$, $\mathrm{E}\big(v_{g2i,j}^2 \big| U_i\big) = \sigma_{vg2}^2 < \infty$, and (iii) the following Cramer's conditions: $\mathrm{E}\big|X_{2i,j}\big|^p \leq C^{p-2} p! \mathrm{E}\big|X_{2,j}\big|^2 < \infty$, $\mathrm{E}\big(|U_{ij}|^p \big| Z_i\big) \leq C^{p-2} p! \sigma_{Uj}^2$, for some $C > 0$, all i, $p = 3, 4, \cdots$, and $j = 1, \cdots, D_2$.

A3 (i) is assumed without loss of generality and is used in identification of the additive structure in Equation (1). In A3 (ii), it is not essential to assume the second conditional moment of the error terms are independent of the conditioning variables; however, the boundedness of the second moment is crucial here, as in Assumptions 1 and 5 in Newey et al. (1999). The Cramer's conditions in A3 (iii) are imposed due to the use, in Lemma 2, of Bernstein's Inequality to establish the uniform order in probability of some specific averages. In particular, Lemma 2 is critical in handling the fact that U_i is estimated by $\hat{U}_i$, which is used in defining $\hat{f}_U$, $\hat{\phi}$ and $\hat{\eta}_i$. If U_i were observed, Cramer's conditions could be relaxed.

Assumption A4. Let C^k denote the class of functions such that each of its elements: (i) is k-times partially continuously differentiable, and (ii) all their partial derivatives up to order k are uniformly bounded. For $d = 1, \cdots, D_2$, and $k = 1, 2$, $\Pi_d(\cdot), \phi(\cdot), f_{UZ}(\cdot), m(\cdot), g(\cdot), m_k(\cdot), g_k(\cdot) \in C^{s+1}$, where s is defined in assumption A1.

Assumption A4 assumes smoothness of the regression functions and uniform bounds of their partial derivatives. This assumption, together with kernels of suitable order, as required in A1, gives desired orders for the biases. We note that in our assumption A1 $s \equiv \max\{s_i\}_{i=1}^4$, and for convenience A4 requires all functions to be in C^{s+1}. This is sufficient for our theorems, but not necessary, expressing only the highest degree of smoothness needed. Depending on the context lower degrees of smoothness can be assumed.[5]

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

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

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

(iii) $h_3 = n^{-\frac{1}{2s_3 + D_3}}$, with $\frac{1}{2} < \frac{s_3}{D_3} < \min_{\{i=2,4\}} \frac{s_i}{D_i}$.

Assumption A5 provides the order of all the bandwidths. The fact that using residual estimates $\{\hat{U}_i\}_{i=1}^n$, instead of $\{U_i\}_{i=1}^n$, has no impact on the first-order asymptotic properties of our estimator relies on undersmoothing in the first stage when regressing X on Z nonparametrically, and on $\Pi(z)$ being sufficiently smooth. For h_2, h_3 and h_4, the orders are chosen optimally by minimizing the mean squared error of traditional NW kernel estimators. The second inequality in A5 (iii) implies that $L_{in}/L_{3n} \to 0$ for $i = 2, 4$ to ensure that using estimated densities for $f_U(\cdot)$ and $\phi(\cdot)$ does not result in any asymptotic consequences in deriving the distribution of $\hat{m}$.

[5] For example, in Section 4, where specific data generating processes (DGP) are considered, it suffices to have $\Pi_d(\cdot) \in C^6$, $\phi(\cdot) \in C^4$, $f_{UZ}(\cdot) \in C^5$, $m(\cdot), m_k(\cdot) \in C^2$, $g, g_k(\cdot) \in C^4$.

3.2 Theorems

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

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

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

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

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

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

Note that in Theorem 1 we establish the uniform convergence rate of $\hat{f}_U(\hat{U}_i)$ and $\hat{\phi}(W_i, \hat{U}_i)$ over $\mathscr{G}_Z \times \mathscr{G}_U$ and $\mathscr{G}_Z \times \mathscr{G}_U \times \mathscr{G}_W$ separately. This is due to the fact that $\hat{U}_i$ is an estimated residual given by $\hat{U}_i = X_i - \hat{\Pi}(Z_i)$ and the uniform convergence rate of $\hat{\Pi}(Z_i)$ given in (13) is taken over a compact set $\mathscr{G}_Z$. Theorem 1 and A2 together imply that $|\hat{\eta}_i - \eta_i| = O_p(L_n)$ uniformly, where $L_n \equiv \sum_{i=2}^4 L_{in}$, and consequently we have $|\hat{\mu}_k - \mu_k| = O_p(L_n)$ for $k = 1, 2$. With this result, we are ready to provide the uniform convergence rate of the estimators given in (9).

Theorem 2. *Under A1–A5, for arbitrary convex and compact subsets $\mathscr{G}_Z$, $\mathscr{G}_U$ and $\mathscr{G}_W$, for $k = 1, 2, 3$, we have,*

$$\sup_{\{Z_i, U_i, W_i\} \in \mathscr{G}_Z \times \mathscr{G}_U \times \mathscr{G}_W} \left| \hat{g}_k(\hat{U}_i) - g_k(U_i) \right| = O_p\left(L_n + \frac{L_{1n}}{h_2}\right), \quad \sup_{\{Z_i, U_i, W_i\} \in \mathscr{G}_Z \times \mathscr{G}_U \times \mathscr{G}_M} \left| \hat{m}_k(W_i) - m_k(W_i) \right| = O_p(L_n). \tag{15}$$

The rates of uniform convergence in probability of $\hat{g}_k$ to g_k and $\hat{m}_k$ to m_k, and by consequence, those of $\hat{g}$ to g and $\hat{m}$ to m depend fundamentally on the degree of smoothness of the functions appearing in A4 and the dimensions of the vectors X_i and Z_i. Given D_i for $i = 1, \cdots, 4$ and assumption A5, it is possible to obtain the necessary smoothness in

A4 that assures the results in Theorem 2. Furthermore, the given rate of convergence can be calculated as a function of n. Similarly, given Assumptions 3 and 8 in Newey et al. (1999), the rate of convergence in their Theorem 4.3 can also be calculated. An important difference between our results and theirs is that, in our case, the rate is obtained taking into account the randomness of $\hat{U}_i$ and the estimation of g (λ in their notation), whereas they take $U = \bar{u}$ as fixed and the true g to be known.

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

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

As we can see in (16), there are two components that need to be studied to establish the asymptotic properties of $\sqrt{n}(\hat{\beta} - \beta)$. We need to (i) establish the asymptotic behavior of the matrix $\frac{1}{n}\hat{X}_2'\hat{\eta}\hat{X}_2$, and (ii) establish the asymptotic normality of the term $\frac{1}{\sqrt{n}}\hat{X}_2'\hat{\eta}(\hat{Y} - \hat{X}_2\beta)$. Uniform orders of NW estimators derived in Theorem 2 will help take care of (i). However, to establish $\sqrt{n}$ asymptotic normality for the second term, we need to investigate the behavior of U-statistics up to degree 3. Yao and Martins-Filho (2015) provides a direct and convenient method to characterize the asymptotic magnitude of each component in the H-decomposition (see Hoeffding, 1948) of a U-statistic, and many places in our proofs are built on their results. The next theorem establishes the asymptotic distribution of $\hat{\beta}$ after suitable centering and under $\sqrt{n}$-normalization.

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

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

where the matrices Φ_0, Φ_1, Φ_2 have typical elements given by

$$\begin{aligned}
\Phi_{0_{(j,k)}} &= \mathrm{E}\left[\eta_t\left(X_{2t,j} - m_{2j}(W_t) - g_{2j}(U_t) + \mu_{2j}\right)\left(X_{2t,k} - m_{2k}(W_t) - g_{2k}(U_t) + \mu_{2k}\right)\right]; \\
\Phi_{1_{(j,k)}} &= \mathrm{E}\left[\eta_t^2\left(X_{2t,j} - m_{2j}(W_t) - g_{2j}(U_t) + \mu_{2j}\right)\left(X_{2t,k} - m_{2k}(W_t) - g_{2k}(U_t) + \mu_{2k}\right)\right]\sigma_v^2;
\end{aligned}$$

$$\Phi_{2_{(j,k)}} = \mathrm{E}\Bigg[\sum_{d=1}^{D_2}\sum_{\delta=1}^{D_2}\mathrm{E}\Big(\big(\mathit{\Pi}_{2j}(Z_i)-\mathit{\Pi}_{2j}(Z_t)\big)D_d g(U_t)\eta_t\big|Z_i\Big) \times \mathrm{E}\Big(\big(\mathit{\Pi}_{2k}(Z_i)-\mathit{\Pi}_{2k}(Z_t)\big)D_\delta g(U_t)\eta_t\big|Z_i\Big)U_{id}U_{i\delta}\Bigg], \quad \text{for} \quad j,k=1,\cdots,D_{22}.$$

Remarks. 1. It follows directly from Theorem 3 that $\hat{\beta}$ is consistent and asymptotically unbiased. The explicit structure for the covariance of the limiting distribution allows for asymptotically valid inference and hypothesis testing when a consistent estimator for the covariance is available. Given the structure of its component covariance matrices, we provide consistent estimators for Φ_i, $i=1,2,3$ as follows,

$$\hat{\Phi}_0 = \frac{1}{n}\hat{X}_2'\hat{\eta}\hat{X}_2, \qquad \hat{\Phi}_1 = \frac{1}{n}\hat{X}_2'\hat{\eta}^2\hat{X}_2\hat{\sigma}_v^2, \qquad \hat{\Phi}_2 = \frac{1}{n}Q'Q, \tag{18}$$

where $\hat{\sigma}_v^2 \equiv \frac{1}{n}\hat{v}'\hat{v}$, $\hat{v} \equiv Y - X_2\hat{\beta} - \hat{\beta}_0 - \hat{m} - \hat{g}$, $Q \equiv (Q_1,\cdots,Q_n)'$, $Q_i \equiv \frac{1}{n}(\mathbf{1}_n\hat{\Pi}_2'(Z_i) - \hat{\Pi}_2)'\hat{\eta}D\hat{g}\hat{U}_i$, $\hat{\Pi}_2(Z_i) \equiv (\hat{\mathit{\Pi}}_{21}(Z_i), \cdots, \hat{\mathit{\Pi}}_{2D_{22}}(Z_i))'$, $\hat{\Pi}_2 \equiv (\hat{\Pi}_2(Z_1),\cdots,\hat{\Pi}_2(Z_n))'$, $\mathbf{1}_n \equiv (1,\cdots,1)'_{n\times 1}$, $D\hat{g} \equiv (\hat{D}_1\hat{g},\cdots,D_{D_2}\hat{g})$, $D_d\hat{g} \equiv (D_d\hat{g}(\hat{U}_1),\cdots, D_d\hat{g}(\hat{U}_n))'$, and $D_d\hat{g}(\hat{U}_i)$ is the partial derivative of the estimator $\hat{g}(u)$ with respect to u_d evaluated at $\hat{U}_i$. Given Equation (9) and (11), by taking partial derivatives, we have $D_d\hat{g}(\hat{U}_i)$ given by

$$D_d\hat{g}(\hat{U}_i) = -\frac{1}{nh_2^{D_2+1}}\frac{1}{\hat{f}_U(\hat{U}_i)}\sum_{t=1}^{n}D_dK_2\left(\frac{\hat{U}_t-\hat{U}_i}{h_2}\right)\left[\hat{\eta}_t(Y_t - X_{2t}\hat{\beta}) - (\hat{g}_1(\hat{U}_i) - \hat{g}_2'(\hat{U}_i)\hat{\beta})\right].$$

2. The covariance $\Phi_0^{-1}(\Phi_1+\Phi_2)\Phi_0^{-1}$ differs from what one would obtain if U_i were observed. Hence, there is an asymptotic cost in using $\hat{U}_i$ in estimation. It manifests itself via the presence of Φ_2, which would be zero if U_i were observed. Furthermore, the covariance matrix of the limiting distribution does not meet the semiparametric efficiency bound of Chamberlain (1992), a characteristic that our estimator shares with that proposed in Li and Wooldridge (2002).[6]

3. Given Theorems 2, 3 and (11), we have the uniform convergence rate of $\hat{g}(\hat{U}_i)$ at $O_p\left(L_n + \frac{L_{1n}}{h_2}\right)$, which is generally worse than that of the traditional NW estimator due to the presence of h_2 in the second term.

The following theorem gives asymptotic normality of $\hat{m}(\cdot)$ at the typical nonparametric rate, in our case, $\sqrt{nh_3^{D_3}}$.

[6]See Li (2000) and Manzan and Zerom (2005) for estimators that satisfy a semiparametric efficiency bound when all regressors are observed, i.e., in the absence of generated regressors.

Theorem 4. *Let $D_j^k f(x) \equiv \frac{\partial^k}{\partial_j \cdots \partial_j} f(x)$ and $D_j^0 f(x) \equiv f(x)$, $\forall k \geq 1, 1 \leq j \leq k$. Under A1–A5, and assume $\mathrm{E}\left(v_{mi}^2 \middle| W_i\right) = \sigma_{vm}^2 < \infty$, $\mathrm{E}\left(|v_{mt}|^{2+\delta}|W_t\right) \leq C < \infty$ for some $\delta > 0$, we have*

$$\sqrt{nh_3^{D_3}}\left(\hat{m}(w) - m(w) - b_m(w)\right) \xrightarrow{d} \mathscr{N}\left(0, \Phi_3 + \Phi_4\right),$$

where

$$b_m(w) = h_3^{s_3} \frac{\mu_{k_3,s_3}}{f_W(w)} \sum_{k=0}^{s_3} \frac{1}{k!(s_3-k)!} \sum_{j=1}^{D_3} D_j^k m(w) D_j^{s_3-k} f_W(w) + o_p(h_3^{s_3}),$$

$$\Phi_3 = \frac{\sigma_{vm}^2}{f_W(w)} \int K_3^2(\gamma)\mathrm{d}\gamma, \quad \Phi_4 = m^2(w) f_W(w) \int \left(\int K_3(\gamma_1) K_3(\gamma_1+\gamma_2)\mathrm{d}\gamma_1\right)^2 \mathrm{d}\gamma_2.$$

Remarks. 1. Given the order and structure of the bias, it follows immediately from Theorem 4 that $\hat{m}(w) - m(w) = o_p(1)$.

2. The fact that η_i, β_0, and β have to be estimated is costly asymptotically. In particular, the variance of the limiting distribution contains the strictly positive term Φ_4 added to Φ_3. Φ_3 can be immediately recognized as the covariance of the limiting distribution of an "oracle" Nadaraya-Watson estimator constructed under the assumption that η_i, β_0, and β are known. Hence, $\hat{m}(\cdot)$ is not oracle efficient. It may be possible to eliminate Φ_4 by considering a new estimator that explores a one-step backfitting procedure using $\hat{g}(\cdot)$. We leave this modification for future research.

4 Monte Carlo Study

In this section, we provide some experimental evidence on the finite sample behavior of our estimators $\left(\hat{\beta}, \hat{m}(\cdot)\right)$ and contrast it to that of some alternative estimation procedures. We consider the following data generating processes (DGPs):

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

$$\mathrm{DGP}_2: \quad Y_i = \frac{\exp(X_{1i})}{1 + 3\exp(X_{1i})} + X_{2i}'\beta + \beta_0 + \varepsilon_i,$$

for $i=1,\cdots,n$. The sample size n is set at 100 and 400. In both DGPs, Z_{1i} and Z_{2i} are generated independently from $N(0,1)$, and we construct $X_{1i}=Z_{1i}+Z_{2i}+U_{1i}$ and $X_{2i}=Z_{1i}^2+Z_{2i}^2+U_{2i}$. ε_i and $U_i=(U_{1i},U_{2i})$ are generated as

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

where the values $\theta=0.3,\ 0.6$, and 0.9 indicate weak, moderate, and strong endogeneity, respectively. It is easy to verify that $\mathrm{E}(\varepsilon_i|Z_i)=0$, $\mathrm{E}(U_i|Z_i)=0$, and thus $\mathrm{E}(\varepsilon_i|U_i,Z_i)=\mathrm{E}(\varepsilon_i|U_i)=\dfrac{\theta}{1+\theta^2}(U_{1i}+U_{2i})$. We set the parameters $\beta=1,\beta_0=1$, and perform 1000 repetitions for each experiment design.

The implementation of our estimators requires a choice of kernel function $K_i(\cdot)$ for $i=1,\cdots,4$ and bandwidth sequences. For all kernels, we use products of a univariate Gaussian kernel of appropriate orders, as we discussed in assumption A1. For both DPGs we have $D_1=D_2=2$, $D_3=1$ and $D_4=3$, and setting $s_1=5$, $s_2=3$, $s_3=1$, $s_4=4$, we choose bandwidths in accordance to A5 by setting $h_1=1.25\hat{\sigma}(Z_i)n^{-\delta}$ for $\delta=1/9$ and $h_i=1.25\hat{\sigma}(M_i)n^{-1/(2s_i+D_i)}$, for $i=2,3,4$, where $\hat{\sigma}(M_i)$ is the sample standard deviation of the variable M_i, with $M_2=\hat{U}_i$, $M_3=(X_{1i},Z_{1i})$, and $M_4=(X_{1i},Z_{1i},\hat{U}_i)$.

We also implement the series estimators proposed by Newey et al. (1999), which we denote by $(\hat{\beta}_{SP},\hat{m}_{SP})$. It should be noted that their estimator was developed for a model where $\beta_0=0$, and the use of a trimming function $w(\tau)$ (in their notation), prevents the use of our assumption $E(\varepsilon)=0$. Thus, we adapt their estimation procedure to the DGPs under consideration and use B-splines throughout the implementation. We use the same number of knots to estimate Π, m and g and follow their constraints on how fast the number of knots diverge to infinity to obtain the convergence results in their Theorem 5.1. Specifically, given D_i for $i=1,\cdots,4$ in the DGPs we must select B-splines of order 7 with $s_1>6$. Hence, the smallest degree of differentiability permitted for Π is $s_1=7$, more than we need to assume to attain the uniform rates of convergence for our nonparametric estimator of m. The higher degree of smoothness they must assume provides some benefits, specifically, for the DGPs considered here, the rate of uniform convergence in probability of our estimator is $n^{-1/3}$ while theirs is $n^{-5/14}$.

Table 1: Finite sample performance

	$\theta = 0.3$					$\theta = 0.6$					$\theta = 0.9$				
	B	S	R	D	M	B	S	R	D	M	B	S	R	D	M
DGP_1	n = 100														
$(\hat{\beta}, \hat{m})$	0.057	0.058	0.081	0.059	0.280	0.078	0.058	0.097	0.078	0.279	0.098	0.056	0.113	0.096	0.310
$(\hat{\beta}_{SP}, \hat{m}_{SP})$	0.062	0.089	0.109	0.073	0.609	0.125	0.088	0.153	0.122	0.587	0.172	0.085	0.192	0.172	0.580
$(\hat{\beta}_{Rob}, \hat{m}_{Rob})$	0.076	0.052	0.092	0.074	0.533	0.139	0.054	0.149	0.135	0.557	0.181	0.054	0.189	0.181	0.591
$(\hat{\beta}_{2SLS}, m)$	0.029	0.798	0.798	0.164		0.058	0.506	0.509	0.167		0.065	0.507	0.511	0.171	
$(\hat{\beta}_{IV}, m)$	0.005	0.053	0.053	0.035		0.013	0.054	0.056	0.038		0.017	0.051	0.054	0.038	
	n = 400														
$(\hat{\beta}, \hat{m})$	0.046	0.029	0.054	0.044	0.277	0.061	0.029	0.067	0.060	0.270	0.075	0.029	0.080	0.074	0.303
$(\hat{\beta}_{SP}, \hat{m}_{SP})$	0.017	0.034	0.039	0.025	0.511	0.032	0.030	0.044	0.034	0.508	0.043	0.029	0.052	0.043	0.505
$(\hat{\beta}_{Rob}, \hat{m}_{Rob})$	0.073	0.025	0.078	0.073	0.520	0.133	0.026	0.136	0.133	0.554	0.173	0.026	0.175	0.172	0.582
$(\hat{\beta}_{2SLS}, m)$	0.018	0.444	0.444	0.159		0.076	0.812	0.815	0.151		0.062	0.404	0.409	0.166	
$(\hat{\beta}_{IV}, m)$	0.002	0.026	0.026	0.017		0.005	0.025	0.025	0.017		0.007	0.025	0.026	0.018	
DGP_2	n = 100														
$(\hat{\beta}, \hat{m})$	0.096	0.058	0.112	0.095	0.182	0.119	0.055	0.131	0.116	0.211	0.144	0.056	0.154	0.144	0.268
$(\hat{\beta}_{SP}, \hat{m}_{SP})$	0.062	0.088	0.108	0.072	0.340	0.123	0.090	0.152	0.122	0.408	0.171	0.082	0.190	0.171	0.311
$(\hat{\beta}_{Rob}, \hat{m}_{Rob})$	0.071	0.052	0.088	0.072	0.243	0.132	0.053	0.143	0.133	0.270	0.175	0.053	0.183	0.176	0.303
$(\hat{\beta}_{2SLS}, m)$	0.031	0.475	0.475	0.156		0.056	0.592	0.594	0.173		0.074	0.718	0.721	0.181	
$(\hat{\beta}_{IV}, m)$	0.003	0.053	0.053	0.036		0.011	0.052	0.054	0.037		0.018	0.053	0.056	0.038	
	n = 400														
$(\hat{\beta}, \hat{m})$	0.077	0.031	0.083	0.075	0.125	0.094	0.034	0.100	0.092	0.163	0.115	0.032	0.119	0.113	0.236
$(\hat{\beta}_{SP}, \hat{m}_{SP})$	0.019	0.033	0.038	0.025	0.319	0.032	0.031	0.045	0.034	0.240	0.043	0.029	0.052	0.044	0.231
$(\hat{\beta}_{Rob}, \hat{m}_{Rob})$	0.073	0.025	0.077	0.072	0.229	0.131	0.027	0.134	0.131	0.258	0.172	0.027	0.174	0.173	0.301
$(\hat{\beta}_{2SLS}, m)$	0.024	0.499	0.499	0.153		0.061	0.456	0.460	0.156		0.090	0.814	0.818	0.165	
$(\hat{\beta}_{IV}, m)$	0.003	0.026	0.026	0.018		0.005	0.025	0.026	0.018		0.007	0.025	0.026	0.017	

Note: The mean of root mean squared error (M) is intended to be left blank for $(\hat{\beta}_{2SLS}, m)$ and $(\hat{\beta}_{IV}, m)$ since m is treated as known and will not be estimated in these cases.

In Table 1, we provide results on bias (B), standard deviation (S), root mean squared error (R), and median of root squared error (D) for the estimation of β, and the mean of root mean squared error (M) for estimating m obtained by averaging across the realized values of (X_{1i}, Z_{1i}). We give results for $\left(\hat{\beta}, \hat{m}\right)$ and for comparison, we also provide results for the oracle estimators for β and β_0 by taking $m(\cdot)$ as given using two different methods. $\hat{\beta}_{2SLS}$ is derived using the traditional two stage least square (2SLS) method for linear models, while $\hat{\beta}_{IV}$ is based on IV estimation using the nonparametric proxies $\hat{\Pi}_2$ as in section 2.2. Lastly, we provide results for the estimators proposed by Robinson (1988), denoted here by $(\hat{\beta}_{Rob}, \hat{m}_{Rob})$, which ignore the endogeneity of X_i. To avoid any extreme estimates or boundary bias in the nonparametric estimation, results on M for estimators of $m(\cdot)$ are only shown by the mean of $10-90\%$ quantile range of sample estimates.[7]

As shown in Table 1, all of the estimators' performances, in terms of the aforementioned measures, improve with the sample size (e.g., for DGP_1, when $\theta = 0.3$, root mean squared error of $\hat{\beta}$ drops nearly 40% from 0.081 to 0.054 when we increase the sample size from 100 to 400). For all DGPs, sample sizes and values of θ, our nonparametric estimators of m outperforms $\hat{m}_{SP}$ and, as expected, $\hat{m}_{Rob}$. The performance of $\hat{\beta}$ relative to that of $\hat{\beta}_{SP}$ is more nuanced. For DGP_1 and $n = 100$ it exhibits smaller B, S, R and D than $\hat{\beta}_{SP}$ for all θ. For $n = 400$ these relative results are reversed except for S where the estimators have similar performance. For DGP_2, $\hat{\beta}_{SP}$ outperforms $\hat{\beta}$ for all θ and all performance measures.

We note that $\hat{\beta}$ and $\hat{\beta}_{SP}$ seem to adequately account for the endogeneity problem since, given the same DGP and sample size, the performance of these estimators regarding bias (B) does not change significantly as the degree of endogeneity (θ) increases, contrasting with the estimator $\hat{\beta}_{Rob}$. In this case, as θ increases from 0.3 to 0.9, the bias more than doubles. The performance of $\hat{\beta}_{2SLS}$ is the worst among the five estimators, even though it is derived assuming $m(\cdot)$ is known. This result is not surprising since in 2SLS estimation we specify a linear structure when approximating the endogenous variables, which in our DGPs it is not. This illustrates the importance of nonparametric estimation when we are not able to specify the functional forms of interest. $\hat{\beta}_{IV}$ avoids that potential misspecification and gives the best performance among all estimators for β in every aspect, exactly as we expected.

To give a more visual description of the distribution of root squared error (RSE) for estimators of β across the simulated samples, we estimate and plot its density for each linear estimator with $n = 100$ for DGP_1 in the left panel

[7] Especially for the second DGP since it has a lower bound of zero for the range of the nonparametric component.

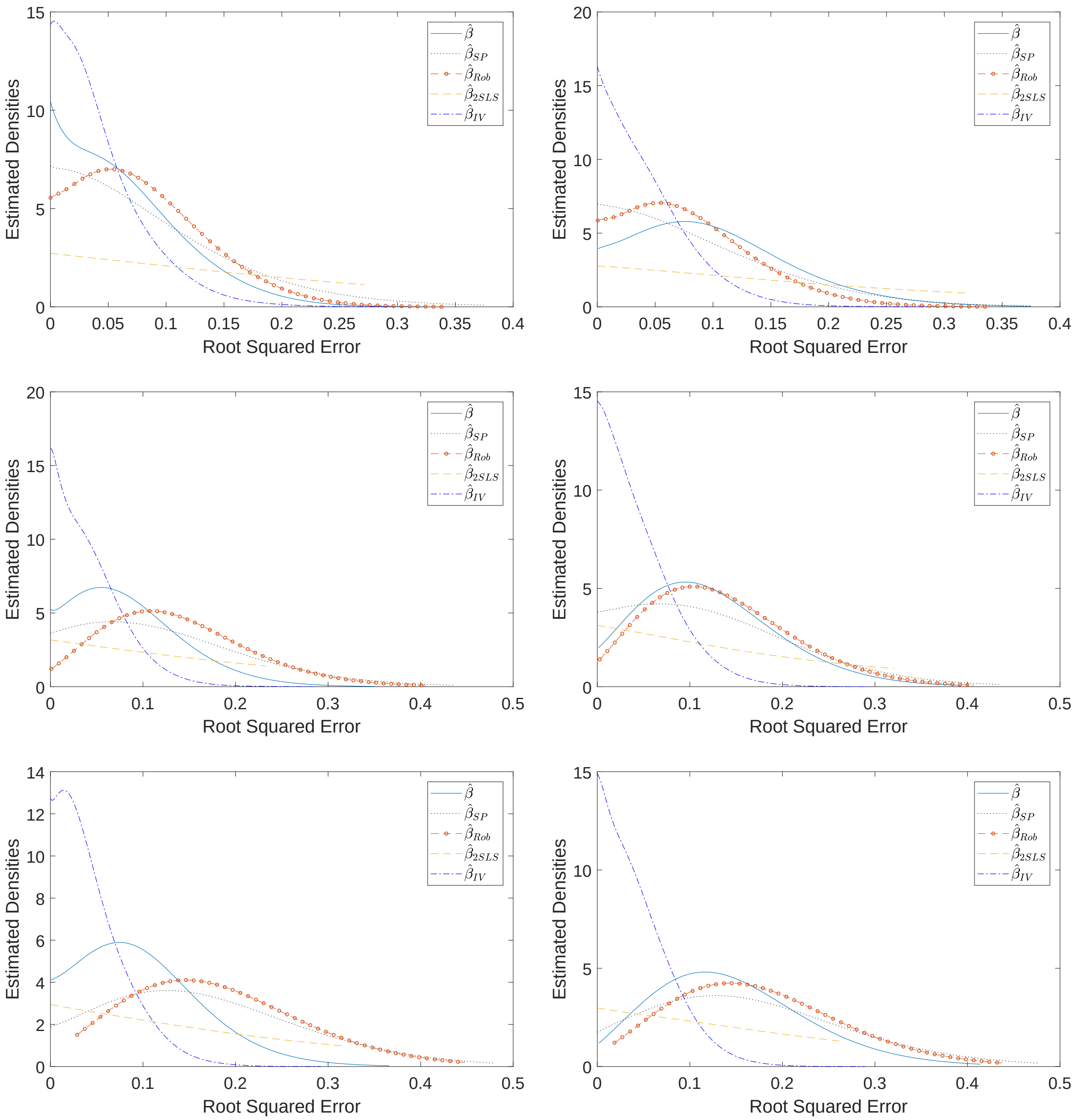


Figure 1: Estimated densities for RMS of estimators of β, $n = 100$, DGP$_1$(left panels) and DGP$_2$ (right panels). $\theta = 0.3$ (top panels), $\theta = 0.6$ (middle panels) and $\theta = 0.9$ (bottom panels)

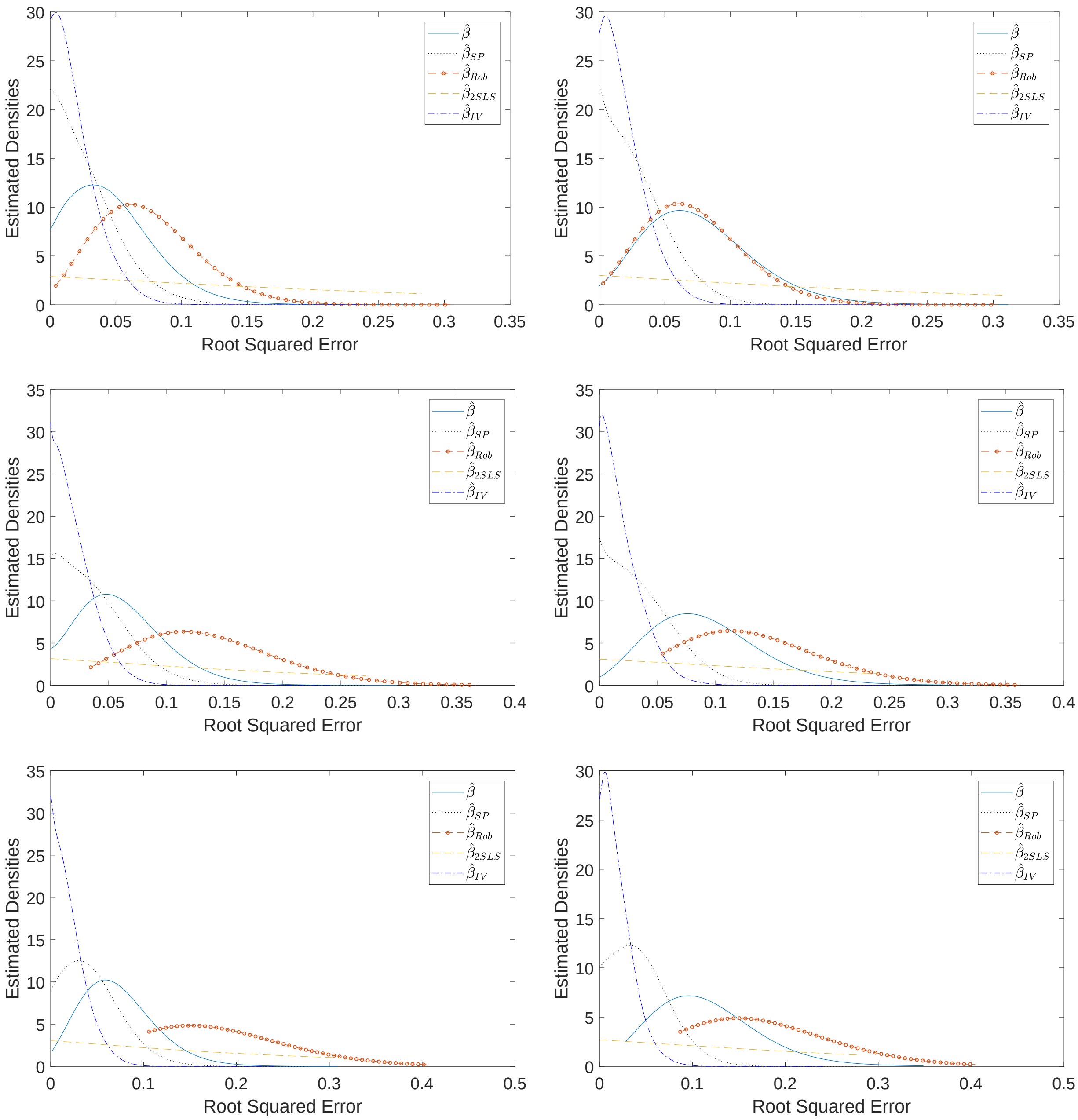


Figure 2: Estimated densities for RMS of estimators of β, $n = 400$, DGP_1(left panels) and DGP_2 (right panels). $\theta = 0.3$ (top panels), $\theta = 0.6$ (middle panels) and $\theta = 0.9$ (bottom panels)

of Figure 1, and DGP_2 in the right panel. The same is done with $n = 400$ in Figure 2. The density estimation is performed using the gamma kernel density estimator proposed by Chen (2000) to avoid any boundary bias. Top, middle, and bottom panels correspond to different degrees of endogeneity, $\theta = 0.3,\ 0.6$, and 0.9, respectively. It is apparent that the estimated densities for the RSE of estimators $\hat{\beta}_{IV}$ (dashed-dotted graph) are closest to the vertical axis, most concentrated around zero and exhibit thinnest tails to the right across all the panels in both figures. In Figure 1 the density associated with our estimator $\hat{\beta}$ (solid graph) is closer to the vertical axis and has thinner tails especially when $\theta = 0.6$ or 0.9. In Figure 2, it is $\hat{\beta}_{SP}$ (dotted line) that is closer to to the vertical axis with thinner tails. The densities associated with the other estimators exhibit particularly bad behavior, especially for large θ.

5 Empirical application: aid-policy-growth relationship

In this section we illustrate the use of our model and the ease of conducting estimation through a simple application. Specifically, we study the impact of foreign aid and policy on economic growth in developing countries. Prominent in this literature is Burnside and Dollar (2000) (henceforth BD). They find that aid is only effective in a good policy environment.[8] This paper was extraordinarily influential at the time and continues to be so due to its clear recommendation: foreign aid should be distributed to countries with good policy environments. However, following BD, an extensive study of the effect of aid has been conducted and results seem to vary greatly with model specifications and samples used.[9] Although alternative tightly parametrized specifications might be useful, Easterly et al. (2004) points out an essential problem and calls for more flexible regression models: *"This literature has the usual limitations of choosing a specification without clear guidance from theory, which often means there are more plausible specifications than there are data points in the sample."*

Therefore, without imposing any prior restrictive functional forms on aid and policy, our model is fully flexible and well suited in this context. More importantly, it controls for the potential endogeneity in the nonparametric and linear parts. For simplicity and ease of comparison, we adopt most of the variables from BD and consider the following

[8] They estimate a 2SLS model and find a significantly positive interaction term between aid and a policy index, controlling the potential endogeneity of aid by using a series of instruments.

[9] There are three mainstream views: 1) BD's Policy View: aid promotes growth but only with a "good policy" environment; see also Collier and Dehn (2001), Collier and Dollar (2002), and Burnside and Dollar (2004); 2) Diminishing Returns View: irrespective of policy, aid promotes growth but with diminishing returns; see Hansen and Tarp (2001) among others; 3) the "Null" View: Boone (1996) finds no relationship between aid and investment, the basic ingredient of growth drivers, excluding those with most aid; see also Rajan and Subramanian (2008).

empirical model:

$$Y_i = m(X_i, Z_{1i}) + Z_{2i}'\beta_1 + \beta_0 + \varepsilon_i, \tag{1'}$$

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

where Y_i is per-capita real GDP growth rate (*gdpg*), X_i is international aid (effective development assistance) provided to a country as a percentage of its GDP (*aid*), Z_{1i} is an index of quality of the policy environment (*policy*),[10] and Z_{2i} is a set of other control variables.[11] Note that Equation (1') is different from (1) in that it now includes a vector of exogenous variables rather than endogenous in the linear part.[12] In line with BD, policy and all the other variables in Z_i are considered exogenous. Aid might be endogenous due to the facts that donors might respond to negative growth shocks by providing more assistance, or countries with positive growth shocks (for example, newly discovered oil fields) might receive special favors from some donors due to strategic or commercial interests. Although the focus of this application lies in the aid-policy-growth relationship (estimation of the nonparametric part), the theoretical model is able to accommodate any endogeneity stemming from covariates in the linear part with suitable instruments.[13] Here in order to keep things simple and comparable with baseline results from BD, we stick to the above empirical model.

Based on the same dataset from BD with a total of 275 observations,[14] we provide all our graphical results in Figure 3. Figure 3a on the left presents a three-dimensional (3D) surface plot of the fitted growth against aid and policy.[15] The surface is smooth and varies significantly with different combinations of aid and policy. The most obvious feature is the high peak when both aid and policy are at high levels, which directly leads to BD's famous Policy View since effect of aid is greatly boosted by "good" policies. This is largely due to Botswana (1978-1989)[16]

---

[10]Variable *policy* is constructed by BD from measures of budget balance, inflation, and the Sachs-Warner openness index.

[11]$Z_i = (Z_{1i}', Z_{2i}', Z_{3i}')'$ represents the set of all exogenous variables where Z_{2i} consists of an index of institutional quality (*icrge*), log of initial real per-capita GDP for the period (*lgdp*), a measure of ethnic fractionalization (*ethnf*), a measure of assassination (*assas*), ethnic fractionalization × assassinations (*ethnf* × *assas*), and a measure of financial depth, money supply as a percentage of GDP lagged one period (*m21*); and Z_{3i} is a set of excluded instrumental variables including log of population (*lpop*) and arms import as a percentage of total imports lagged one period (*arms1*).

[12]The estimation procedure and Theorem 1–3 continue to hold since exogeneity of the added regressors creates no added difficulties for the asymptotic characterization of our proposed estimator.

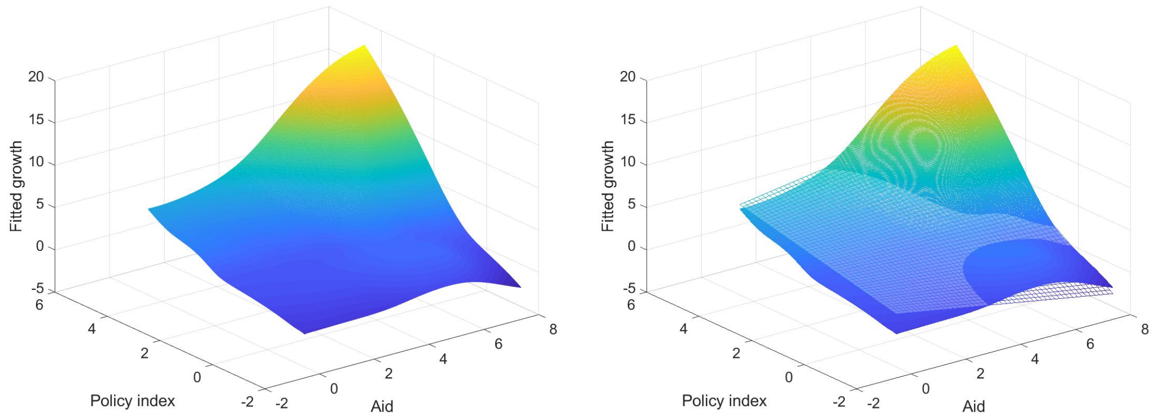
[13]For example, we also find that institutional quality could be endogenous given that faster economic growth may produce higher levels of institutional quality (see Aron, 2000) and there might be some unobserved factors that jointly determine both high levels of institutional quality and economic growth (see Easterly et al., 2006). A plausible instrument for it is Gini index, a measure of social cohesion that, in part, determines the institutional quality. See Easterly et al. (2006) for more details. We leave this for future work.

[14]The dataset is publicly available at `www.cgdev.org/publication/aid-policies-and-growth-data-set`.

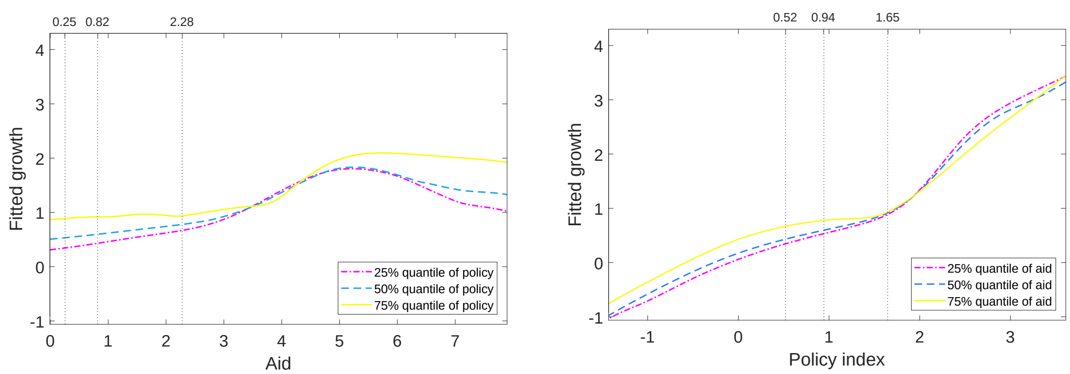
[15]We plot on where aid and policy most concentrated, that is, aid GDP ratio from -0.5% and 8% (more than 98% observations) and policy above -1.5 (more than 97% observations).

[16]Botswana is well known as the "African Exception" due to its high economic growth and democracy. Its record consistently stands in stark contrast to virtually all other parts of Sub-Saharan Africa.

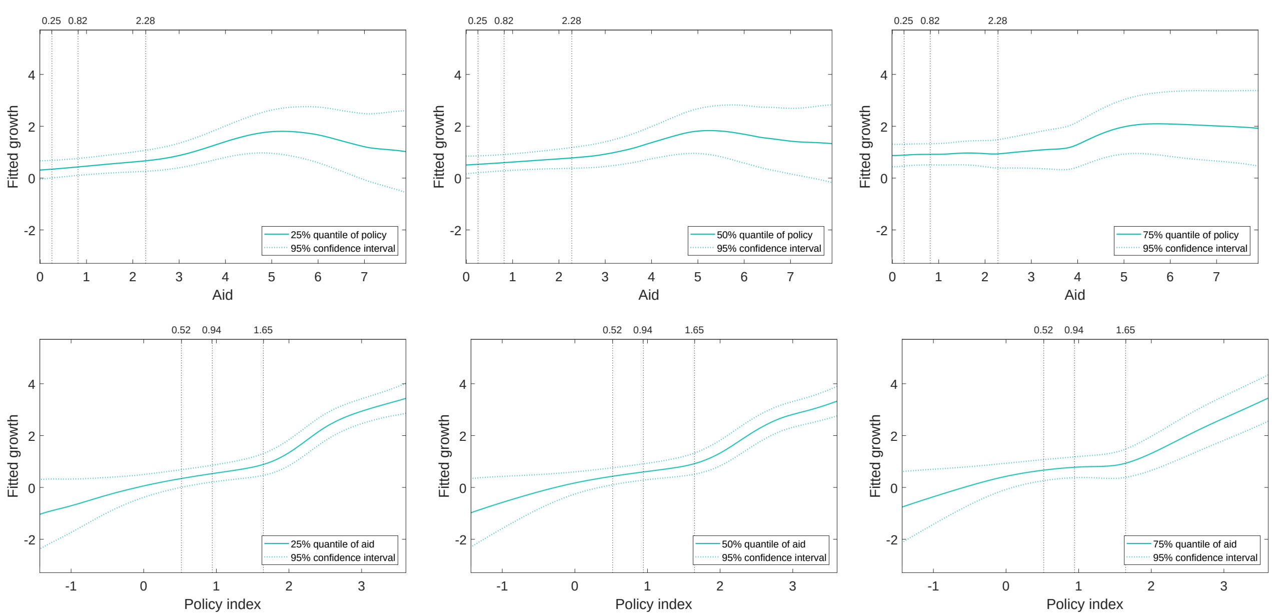
Figure 3: Aid and policy effects on growth



(a) Joint aid-policy-growth relationship



(b) Individual aid-growth and policy-growth relationship



(c) Individual aid-growth and policy-growth relationship with 95% confidence interval which has consistently high levels of aid, policy, and growth rate. To give a better comparison with BD, Figure 3a on the right stacks the 3D plot with the fitted growth predicted in BD under a linear 2SLS model.[17] Due to the linearity restriction, it is a flat plane without any fluctuation, which is roughly an average of our fitted surface. One of the most important features it misses is that aid appears to have varying effects at different range. In particular, it is growth-enhancing at high levels while the linear model simply averages it out. Taking a closer look into the individual effects of aid and policy, we slice the surface along aid with policy fixed at its 25%, 50%, and 75% quantile in Figure 3b on the left. To make the plot more informative, we also mark the 25%, 50%, and 75% quantile of aid on the top axis and draw three vertical dotted lines. In general, the effect of aid is not obvious, except at very high levels (above 3% aid GDP ratio).[18] In contrast, we can see from the right figure that a good policy environment is indeed growth-enhancing across its entire range with a larger effect at high levels (above its 75% quantile). For statistical inference, we add a 95% confidence interval in Figure 3c for each aid-growth or policy-growth curve of Figure 3b. As expected, the confidence band varies greatly with aid or policy distribution, that is, it widens where the data is scarce.

In sum, we find that aid in general does not promote growth, expect at high levels (above 3% aid GDP ratio) while policy has a consistent and positive effect. Our findings do not support BD's conclusion—policy increases aid effectiveness in growth. In BD, aid effectiveness is assumed to be only dependent on policy, not even on itself. Figure 3b on the left provides a plausible explanation. We see that the effect of aid does vary with itself, but it will be averaged out in BD's setup for countries with a not so good policy environment (25% and 50% quantile) due to the drop in curves when aid GDP ratio is above 5%, while for countries with a very good policy environment (75% quantile), we do not see such drop. The positive interaction term in BD only captures the increasing averaged effect of aid with policy but misses the whole picture. In fact, for the majority range of aid, its effectiveness (slope of the curves) actually decreases with policy although the difference seems not significant.

[17] Coefficient estimates are reported in Column (5) (2SLS) of Table 4 in BD, where aid has a coefficient -0.32, policy 0.74^{**}, and their interaction 0.18^{*}. ** and * represent 5% and 10% significance levels, respectively.

[18] We also implement estimators in Robinson (1988) without controlling for any endogeneity, and find that this positive effect at high levels is cut in half, suggesting that aid might be endogenous in that it is more likely to be given due to assistance purpose.

6 Summary and conclusion

In this paper we contribute to the growing literature on the estimation of semiparametric and nonparametric regression models with endogenous regression. Adopting the control function approach, we propose easily computable kernel-based estimators for the finite and infinite dimensional parameters of a partially linear regression model and establish their asymptotic distributions. Two critical steps are needed to establish these results: first, the choice of the normalizing function $L(\cdot)$ appearing in Section 2.1, and second the repeated use of the results on U-Statistics obtained in Yao and Martins-Filho (2015). Besides its role in assuring asymptotic normality of the proposed estimators, the choice of $L(\cdot)$ generates a class of estimators with different variances for their asymptotic distributions. A simple empirical investigation of the aid-policy-growth relationship is provided to illustrate the ease of implementation of our method. Future research should be done on selecting optimal (minimal variance) estimators from this class. In fact, further investigation of the efficiency properties of these estimators may shed light on how to construct oracle efficient estimators for $m(\cdot)$ and semiparametric efficient estimators for β.

References

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

Aron, J., 2000. Growth and Institutions: A Review of the Evidence. The World Bank Research Observer 15, 99–135.

Birke, M., van Bellegem, S., van Keilegom, I., 2017. Semi-parametric estimation in single-index model with endogenous variables. Scandinavian Journal of Statistics 44, 168–191.

Blundell, R., Kristensen, D., Matzkin, R., 2013. Control functions and simultaneous equations methods. American Economic Review: Papers and Proceedings 103, 563–569.

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

Boone, P., 1996. Politics and the effectiveness of foreign aid. European Economic Review 40 (2), 289–329.

Burnside, C., Dollar, D., 2000. Aid, Policies, and Growth. American Economics Review 90 (4), 847–868.

Burnside, C., Dollar, D., 2004. Aid, Policies and Growth: Revisiting the Evidence. The World Bank Policy Research Working Paper No. 3251.

Chamberlain, G., may 1992. Efficiency Bounds for Semiparametric Regression. Econometrica 60, 567–596.

Chen, S. X., 2000. Probability density function estimation using gamma kernels. Annals of the Institute of Statistical Mathematics 52 (3), 471–480.

Chen, X., Christensen, T., 2015. Optimal sup-norm rates, adaptivity and inference in nonparametric instrumental variable estimation. Discussion paper 1923R, Cowles Foundation.

Chen, X., Qiu, Y. J., 2016. Methods for nonparametric and semiparametric regressions with endogeneity: a gentle guide. Annual Review of Economics 8, 259–290.

Collier, P., Dehn, J., 2001. Aid, shocks, and growth. The World Bank Policy Research Working Paper No. 2688.

Collier, P., Dollar, D., 2002. Aid allocation and poverty reduction. European Economic Review 46, 1475–1500.

Darolles, S., Fan, Y., Florens, J.-P., Renault, E., 2011. Nonparametric instrumental regression. Econometrica 79, 1541–1566.

Easterly, W., Levine, R., Roodman, D., 2004. Aid, policies, and growth: Comment. American Economic Review 94, 774–780.

Easterly, W., Ritzen, J., Woolcock, M., 2006. Social cohesion, institutions, and growth. Economics and Politics 18 (2), 103–120.

Hahn, J., Ridder, G., 2013. Asymptotic Variance of Semiparametric Estimators With Generated Regressors. Econometrica 81 (1), 315–340.

Hall, P., Horowitz, J. L., 2005. Nonparametric methods for inference in the presence of instrumental variables. Annals of Statistics 33 (2904-2929).

Hansen, H., Tarp, F., 2001. Aid and growth regressions. Journal of Development Economics 64, 547–570.

Härdle, W., Liang, H., Gao, J., 2000. Partially linear models. Physica-Verlag, Heidelberg.

Hoeffding, W., 1948. A class of statistics with asymptotically normal distribution. The Annals of Mathematical Statistics 19, 293–325.

Hoeffding, W., 1961. The strong law of large numbers for U-Statistics, Institute of Statistics, Mimeo Series 302, University of North Carolina, Chapel Hill, NC.

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

Li, Q., 2000. Efficient estimation of additive partially linear models. International Economic Review 41, 1073–1092.

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

Li, Q., Wooldridge, J., 2002. Semiparametric estimation of partially linear models for dependent data with generated regressors. Econometric Theory 18, 625–645.

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

Manzan, S., Zerom, D., 2005. Kernel estimation of a partially linear additive model. Statistics and Probability Letters 72, 313–322.

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

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

Newey, W. K., Powell, J. L., 2003. Instrumental variable estimation of nonparametric models. Econometrica 71, 1565–1578.

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

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

Ozabaci, D., Henderson, D., Su, L., 2014. Additive nonparametric regression in the presence of endogenous regressors. Journal of Business and Economic Statistics 32, 555–575.

Parzen, E., 1962. On estimation of a probability density and mode. Annals of Mathematical Statistics 33, 1065–1076.

Pinkse, J., 2000. Nonparametric two-step regression estimation when regressors and error are dependent. Canadian Journal of Statistics 28, 289–300.

Rajan, R. G., Subramanian, A., 2008. Aid and Growth: What Does the Cross-Country Evidence Really Show? Review of Economics and Statistics 90, 643–665.

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

Speckman, P., 1988. Kernel smoothing in partial linear models. Journal of the Royal Statistical Society B 50, 413–433.

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

Yao, F., Martins-Filho, C., 2015. An asymptotic characterization of finite degree U-statistics with sample size dependent kernels: applications to nonparametric estimators and test statistics. Communications in Statistics - Theory and methods 44, 3251–3265.

Appendix

This appendix presents the proofs of the main theorems and statements and proofs of the supporting lemmas. For a scalar variable x, $f'(x)$ denotes the derivative of $f(x)$ evaluated at x. For $D\times 1$ vectors γ,β, define $\gamma^{\beta}=\prod_{d=1}^{D}\gamma_d^{\beta_d}$, $|\beta|=\sum_{d=1}^{D}\beta_d$, $D_d f(\gamma)=\frac{\partial}{\partial d}f(\gamma)$, $D_{dk}^2 f(\gamma)=\frac{\partial^2}{\partial d\partial k}f(\gamma)$, $D^{\beta}f(\gamma)=\frac{\partial^{|\beta|}}{\partial_1^{\beta_1}\cdots\partial_D^{\beta_D}}f(\gamma)$. $\mathbf{J}f(\gamma)$ and $\mathbf{H}f(\gamma)$ denote the Jacobian and Hessian matrix of $f(\gamma)$. Note that for a scalar function $f(\gamma)$, $\mathbf{J}f(\gamma)$ is the transpose of the gradient vector of $f(\gamma)$. $A\times B$ denotes the Cartesian product of two sets A and B. χ_A denotes the indicator function for the set A. $P(A)$ denotes the probability of event A in the probability space $(\Omega,\mathscr{F},P)$, $\mathrm{E}(\cdot)$ denotes expectation, and $\mathrm{V}(\cdot)$ denotes variance.

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

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

where $\sum_{(n,k)}$ denotes the sum over all subsets $1\le i_1<\cdots<i_k\le n$ of $\{1,\cdots,n\}$. Now let $\phi_{cn}(z_1,\cdots,z_c)=\mathrm{E}(\phi_n(P_1,\cdots,P_c,P_{c+1},\cdots,P_k)|P_1=p_1,\cdots,P_c=p_c)$, $\sigma_{cn}^2=\mathrm{V}(\phi_{cn}(P_1,\cdots,P_c))$ and $\theta_n=\mathrm{E}(\phi_n(P_{i_1},\cdots,P_{i_k})$. In addition, recursively define $h_n^{(1)}(p_1)=\phi_{1n}(p_1)-\theta_n,\cdots,h_n^{(c)}(p_1,\cdots,p_c)=\phi_{cn}(p_1,\cdots,p_c)-\sum_{j=1}^{c-1}\sum_{(c,j)}h_n^{(j)}(p_{i_1},\cdots,p_{i_j})-\theta_n$ for $c=2,\cdots,k$. By Hoeffding's H-decomposition in Hoeffding (1961) we have

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

where $H_n^{(j)}(P_{i_1},\cdots,P_{i_j})=\binom{n}{j}^{-1}\sum_{(n,j)}h_n^{(j)}(P_{i_1},\cdots,P_{i_j})$. The order of U_n can be determined by studying each $H_n^{(j)}$ and θ_n in the finite sum. By Theorem 1 in Yao and Martins-Filho (2015), the order of $H_n^{(j)}$ is determined by n and the leading variance σ_{jn}^2. Throughout the proofs, we will use $\{P_i\}_{i=1}^{n}$ and the above notation to characterize the U-statistics of interest, denoted by U_n .

Proof of Theorems

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

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

$$\hat{f}_U(\hat{U}_i)-\hat{f}_U(U_i)=\frac{1}{nh_2^{D_2}}\sum_{t=1}^{n}\left(\hat{K}_{2ti}-K_{2ti}\right)=\frac{1}{nh_2^{D_2}}\sum_{t=1}^{n}\left(\sum_{|\beta|=1}^{3}\frac{H^{\beta}}{|\beta|!}D^{\beta}K_{2ti}+\sum_{|\beta|=4}\frac{H^{\beta}}{|\beta|!}D^{\beta}K_2\left(\frac{U_t-U_i}{h_2}+\lambda H\right)\right)\equiv\sum_{i=1}^{4}T_i,$$

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

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

Step 1: We rewrite T_1 into two parts:

$$T_1 = \sum_{d=1}^{D_2} \Big(-\frac{1}{nh_2^{D_2+1}} \sum_{t=1}^{n} (\hat{U}_{id} - U_{id}) D_d K_{2ti} + \frac{1}{nh_2^{D_2+1}} \sum_{t=1}^{n} (\hat{U}_{td} - U_{td}) D_d K_{2ti} \Big) \equiv \sum_{d=1}^{D_2} (T_{11} + T_{12}),$$

where $T_{11} \equiv -\big(\hat{U}_{id} - U_{id}\big) C_1(U_i)$ and $C_1(U_i) \equiv (nh_2^{D_2+1})^{-1} \sum_{t=1}^{n} D_d K_{2ti}$. By Lemma 3, it can be shown that $\sup_{U_i \in \mathscr{G}_U} |C_1(U_i) - \mathrm{E}(C_1(U_i))| = O_p\Big(\big(\log n/(nh_2^{D_2+2})\big)^{1/2}\Big) = o_p(1)$, and by integration by parts, $\mathrm{E}(C_1(U_i)) = \int K_2(\gamma) D_d f_U(U_i - h_2\gamma) \mathrm{d}\gamma \leq C$ uniformly. Thus, $\sup_{U \in \mathscr{G}_U} |C_1(U_i)| = O_p(1)$. Note that $\big|\hat{U}_{id} - U_{id}\big| = \big|\hat{\Pi}_d(Z_i) - \Pi_d(Z_i)\big|$, and by the uniform convergence rate of Nadaraya-Watson estimator, we have $\sup_{Z_i \in \mathscr{G}_Z} \big|\hat{U}_{id} - U_{id}\big| = O_p(L_{1n})$. Consequently, $T_{11} = O_p(L_{1n})$ uniformly.

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

$$-(\hat{U}_{td} - U_{td}) = \hat{\Pi}_d(Z_t) - \Pi_d(Z_t) = \frac{1}{nh_1^{D_1} f_Z(Z_t)} \sum_{l=1}^{n} K_{1lt} \Big(U_{ld} + \Pi_d(Z_l) - \Pi_d(Z_t) \Big) + O_p(L_{1n}^2) \tag{A.1}$$

by the uniform order of $\hat{f}_Z(Z_t) - f_Z(Z_t)$ and $\hat{U}_{td} - U_{td}$. Thus, we have

$$\begin{aligned} T_{12} &= \frac{1}{n^2} \sum_{t=1}^{n} \sum_{l=1}^{n} \frac{1}{h_1^{D_1} h_2^{D_2+1} f_Z(Z_t)} K_{1lt} D_d K_{2ti} U_{ld} + \frac{1}{n^2} \sum_{t=1}^{n} \sum_{l=1}^{n} \frac{1}{h_1^{D_1} h_2^{D_2+1} f_Z(Z_t)} K_{1lt} D_d K_{2ti} \Big(\Pi_d(Z_l) - \Pi_d(Z_t) \Big) \\ &\quad + O_p(L_{1n}^2) \frac{1}{nh_2^{D_2+1}} \sum_{t=1}^{n} |D_d K_{2ti}| \equiv T_{121} + T_{122} + O_p(L_{1n}^2/h_2), \\ T_{121} &= \frac{1}{n^2} \sum_{t=1}^{n} \frac{1}{h_1^{D_1} h_2^{D_2+1} f_Z(Z_t)} K_1(0) D_d K_{2ti} U_{td} + \frac{1}{n^2} \sum_{\substack{t=1 \\ t \neq l}}^{n} \sum_{l=1}^{n} \frac{1}{h_1^{D_1} h_2^{D_2+1} f_Z(Z_t)} K_{1lt} D_d K_{2ti} U_{ld} \equiv E_{1n} + E_{2n}. \end{aligned}$$

We can show that $E_{1n} = O_p\Big((nh_1^{D_1} h_2)^{-1}\Big)$ uniformly over $\mathscr{G}_U$ by Lemma 3, and $E_{2n} \leq C|U_n|$, where $U_n = \binom{n}{2}^{-1} \sum_{\substack{t=1 \\ t \neq l}}^{n} \sum_{l=1}^{n} \frac{K_{1lt} D_d K_{2ti}}{h_1^{D_1} h_2^{D_2+1} f_Z(Z_t)} U_{ld} \equiv \binom{n}{2}^{-1} \sum_{\substack{t=1 \\ t \neq l}}^{n} \sum_{l=1}^{n} \psi_{nlt} \equiv \binom{n}{2}^{-1} \sum_{\substack{t=1 \\ t < l}}^{n} \sum_{l=1}^{n} \phi_{nlt} = \theta_n + 2H_n^{(1)} + H_n^{(2)}$ is a U-statistic. $\theta_n = \mathrm{E}(\phi_{nlt}) = 0$ in this case. $H_n^{(1)} = \frac{1}{n} \sum_{l=1}^{n} h_n^{(1)}(U_i, P_l) = \frac{1}{n} \sum_{l=1}^{n} \phi_{1n}(U_i, P_l) = \frac{1}{n} \sum_{l=1}^{n} \mathrm{E}(\phi_{nlt} | U_i, P_l) = \frac{1}{n} \sum_{l=1}^{n} U_{ld} c(U_i, Z_l)$, where $c(U_i, Z_l) \equiv \int K_1(\gamma_1) K_2(\gamma_2) D_d f_{U|Z}(U_i + h_2\gamma_2 | Z_l - h_1\gamma_1) \mathrm{d}\gamma_1 \mathrm{d}\gamma_2$. Given Cramer's condition in A3 and Lemma 2, we have $\sup_{\{Z,U\} \in \mathscr{G}_Z \times \mathscr{G}_U} H_n^{(1)} = O_p\big((\log n/n)^{1/2}\big)$, as $\mathrm{E}(H_n^{(1)}) = 0$. For $H_n^{(2)}$, by Theorem 1 in Yao and Martins-Filho (2015), $H_n^{(2)} = (\sigma_{2n}^2/n^2)^{1/2} O_p(1)$, where $\sigma_{2n}^2 \equiv \mathrm{V}(\phi_{nlt}) = \mathrm{E}(\phi_{nlt}^2) \leq 4\mathrm{E}(\psi_{nlt}^2) = O\Big((h_1^{D_1} h_2^{D_2+2})^{-1}\Big)$. Thus $H_n^{(2)} = (n^2 h_1^{D_1} h_2^{D_2+2})^{-1/2} O_p(1)$ uniformly. In sum, $T_{121} = O_p\Big((nh_1^{D_1} h_2)^{-1} + (\log n/n)^{1/2} + (n^2 h_1^{D_1} h_2^{D_2+2})^{-1/2}\Big) = O_p(L_{1n})$ uniformly by A5.

The order of T_{122} could be analyzed in the same way, given that Π and f_Z are s_1 times partially continuously differentiable, and K_1 is a multivariate kernel of order s_1, we have $T_{122} = O_p\big(h_1^{s_1} + (\log n/n)^{1/2} + (n^2 h_1^{D_1-2} h_2^{D_2+2})^{-1/2}\big) =$

$O_p(L_{1n})$ uniformly by A5. In sum, $\sup_{\{Z,U\}\in\mathscr{G}_Z\times\mathscr{G}_U} T_1 = O_p(L_{1n})$.

Step 2: $T_2 = \sum_{|\beta|=2}\left(nh_2^{D_2}\right)^{-1}\sum_{t=1}^n H^\beta D^\beta K_{2ti}$, when 1 appears in the d^{th} and k^{th} position of β, we have

$$\frac{1}{nh_2^{D_2}}\sum_{t=1}^n H^\beta D^\beta K_{2ti} = \frac{1}{2nh_2^{D_2+2}}\sum_{t=1}^n \left[(\hat{U}_{td}-U_{td})-(\hat{U}_{id}-U_{id})\right]\left[(\hat{U}_{tk}-U_{tk})-(\hat{U}_{ik}-U_{ik})\right] D^2_{dk}K_{2ti}.$$

Since $\sup_{Z\in\mathscr{G}_Z}\left|\hat{U}_{ab}-U_{ab}\right| = O_p(L_{1n})$, for $a=i,j$ and $b=d,k$, we have $T_2 = O_p\left(L_{1n}^2/h_2^2\right)\left(nh_2^{D_2}\right)^{-1}\sum_{t=1}^n\left|D^2_{dk}K_{2ti}\right| \equiv O_p\left(L_{1n}^2/h_2^2\right)C_2(U_i)$. By Lemma 3 and that $\mathrm{E}(C_2(U_i)) = O(1)$ uniformly over $\mathscr{G}_U$, we have $C_2(U_i) = O_p(1)$ uniformly. Thus, $\sup_{\{Z,U\}\in\mathscr{G}_Z\times\mathscr{G}_U} T_2 = O_p\left(L_{1n}^2/h_2^2\right)$.

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

Step 4: T_4 is different from T_2 and T_3 in that $\sup_{U\in\mathscr{G}_U} C_4(U_i) = O_p(1/h_2^{D_2})$, where $C_4(U_i) \equiv \left(nh_2^{D_2}\right)^{-1}\sum_{t=1}^n\left|D^\beta K^*_{2ti}\right|$, for any $|\beta|=4$, and $D^\beta K^*_{2ti} \equiv D^\beta K_2((U_t-U_i)/h_2+\lambda H)$. Thus, $\sup_{\{Z,U\}\in\mathscr{G}_Z\times\mathscr{G}_U} T_4 = O_p\left(L_{1n}^4/h_2^{D_2+4}\right)$. By A5, it can be shown that $T_2,T_3,T_4 = o_p(n^{-1/2})$, and $T_1 = O_p(L_{1n}) = O_p(L_{2n})$, which gives us

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

The uniform order of $\left|\hat{\phi}(W_i,\hat{U}_i)-\phi(W_i,U_i)\right|$ can be derived in the similar way under A5, and consequently, here, we omit the details.

□

Theorem 2 *Proof.* We start with the j^{th} element of $\hat{g}_2(\hat{U}_i)-g_2(U_i)$. Note that

$$\begin{aligned}
\hat{g}_{2j}(\hat{U}_i)-g_{2j}(U_i) &= \frac{1}{nh_2^{D_2}\hat{f}_U(\hat{U}_i)}\sum_{t=1}^n \hat{K}_{2ti}\hat{\eta}_t X_{2t,j} - g_{2j}(U_i)\\
&= \frac{1}{nh_2^{D_2}\hat{f}_U(\hat{U}_i)}\sum_{t=1}^n \hat{K}_{2ti}\underbrace{\left\{(\hat{\eta}_t-\eta_t)X_{2t,j}+v_{g2t,j}+\left((g_{2j}(U_t)-g_{2j}(U_i)\right)\right\}}_{C_{g2ti}}\\
&= \left\{\frac{1}{nh_2^{D_2}f_U(U_i)}\sum_{t=1}^n K_{2ti}C_{g2ti} + \frac{1}{nh_2^{D_2+1}f_U(U_i)}\sum_{t=1}^n \mathbf{J}K_{2ti}\big(\hat{U}_t-U_t-(\hat{U}_i-U_i)\big)C_{g2ti}\right.\\
&\quad \left. + \frac{1}{nh_2^{D_2}f_U(U_i)}\sum_{t=1}^n R_{ti}C_{g2ti}\right\}\big(1+O_p(L_{2n})\big) \qquad \text{(A.2)}\\
&\equiv \left(\sum_{k=1}^3 T_k\right)\big(1+O_p(L_{2n})\big),
\end{aligned}$$

where R_{ti} is the remainder term of a Taylor's expansion of $\hat{K}_{2ti}$ at $(U_t-U_i)/h_2$, and $v_{g2t,j}$ is the j^{th} element of v_{g2t}. We complete the proof by showing in three steps that $T_1 = O_p(L_n)$, $T_2 = O_p\left(L_{1n}/h_2\right)$, and $T_3 = o_p(n^{-1/2})$.

Step 1: Let $T_1 \equiv \sum_{k=1}^{3} T_{1k}$, corresponding to the three components in C_{g2ti} separately. By Theorem 1 and A2, we have

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

By Lemma 3, $T_{11} = O_p(L_n)(nh_2^{D_2})^{-1}\sum_{t=1}^{n}|K_{2ti}\eta_t X_{2t,j}| = O_p(L_n)$ uniformly, since by A3 and A4,

$$\begin{aligned}
\mathrm{E}\left(\frac{1}{nh_2^{D_2}}\sum_{t=1}^{n}\left|K_{2ti}\eta_t X_{2t,j}\right|\right) &= \frac{1}{h_2^{D_2}}\mathrm{E}\Big(\left|K_{2ti}(g_{2j}(U_t)+v_{g2t,j})\right|\Big)\\
&\leq \int |K_2(\gamma)|(|g_{2j}(U_i+h_2\gamma)|+C)f_U(U_i+h_2\gamma)\mathrm{d}\gamma\\
&\leq C\,|g_{2j}(U_i)|+C\int |K_2(\gamma)|(|g_{2j}(U_i+h_2\gamma)|-|g_{2j}(U_i)|)\mathrm{d}\gamma+C\\
&\leq C\,|g_{2j}(U_i)|+C\,h_2\int |K_2(\gamma)|\sum_{d=1}^{D_2}|\gamma_d|\mathrm{d}\gamma+C\\
&\leq C\,|g_{2j}(U_i)|+C, \quad \text{which is bounded uniformly over } \mathscr{G}_U.
\end{aligned}$$

By Lemma 3, we have $\sup_{U\in\mathscr{G}_U}|T_{12}| = O_p\left((\log n/nh_2^{D_2})^{1/2}\right) = O_p(L_{2n})$, given $\mathrm{E}(T_{12})=0$.

For T_{13}, note that by Taylor's Theorem, $\mathrm{E}(T_{13}) = h_2^{-D_2}f_U^{-1}(U_i)\mathrm{E}\Big(K_{2ti}\big(g_{2j}(U_t)-g_{2j}(U_i)\big)\Big) = f_U^{-1}(U_i)\int K_2(\gamma)\big(g_{2j}(U_i+h_2\gamma)-g_{2j}(U_i)\big)f_U(U_i+h_2\gamma)\mathrm{d}\gamma = O(h_2^{s_2}) = O(L_{2n})$ uniformly over $\mathscr{G}_U$, given that K_2 is of order s_2, $g_{2j}(U_t), f_U(U_t)\in C^{s_2}$ and all the partial derivatives of $g_{2j}(U_t)$ up to order s_2 are uniformly bounded by A4. By Lemma 3, we have $h_2^{-1}\sup_{U\in\mathscr{G}_U}|T_{13}-\mathrm{E}(T_{13})| = O_p\big((\log n/(nh_2^{D_2}))^{1/2}\big) = O_p(L_{2n})$. Thus, $\sup_{U\in\mathscr{G}_U}|T_{13}| = O_p(L_{2n})$, and we have $T_1 = O_p(L_n)$ uniformly.

Step 2: For T_2, similar to T_{11}, by Lemma 3, we have

$$\begin{aligned}
T_2 &= \frac{1}{nh_2^{D_2+1}f_U(U_i)}\sum_{t=1}^{n}\mathbf{J}K_{2ti}\big(\hat{U}_t-U_t-(\hat{U}_i-U_i)\big)C_{g2ti}\\
&= O_p\Big(\frac{L_{1n}}{h_2}\Big)\sum_{d=1}^{D_2}\frac{1}{nh_2^{D_2}f_U(U_i)}\sum_{t=1}^{n}\left|D_dK_{2ti}\Big((\hat{\eta}_t-\eta_t)X_{2t,j}+v_{g2t,j}+\big((g_{2j}(U_t)-g_{2j}(U_i)\big)\Big)\right|\\
&= O_p\Big(\frac{L_{1n}}{h_2}\Big).
\end{aligned}$$

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

$$\begin{aligned}
T_{31} &= \sum_{d=1}^{D_2}\sum_{l=1}^{D_2}\frac{1}{2nh_2^{D_2+2}f_U(U_i)}\sum_{t=1}^{n}D_{dl}^2K_{2ti}\Big(\hat{U}_{td}-U_{td}-(\hat{U}_{id}-U_{id})\Big)\Big(\hat{U}_{tl}-U_{tl}-(\hat{U}_{il}-U_{il})\Big)C_{g2ti}\\
&= O_p\left(\frac{L_{1n}^2}{h_2^2}\right)\sum_{d=1}^{D_2}\sum_{l=1}^{D_2}\frac{1}{nh_2^{D_2}}\sum_{t=1}^{n}\left|D_{dl}^2K_{2ti}C_{g2ti}\right| = O_p\left(\frac{L_{1n}^2}{h_2^2}\right),
\end{aligned}$$

by A3. Similarly, $T_{32} = O_p\big(L_{1n}^3/h_2^3\big)$. By A1, $T_{33} = O_p\big(L_{1n}^4/h_2^{D_2+4}\big)\frac{1}{n}\sum_{t=1}^n \big|C_{g2ti}\big| = O_p\big(L_{1n}^4/h_2^{D_2+4}\big)$. By A5, we can show that $T_3 = O_p\big(L_{1n}^2/h_2^2 + L_{1n}^3/h_2^3 + L_{1n}^4/h_2^{D_2+4}\big) = o_p(n^{-1/2})$ uniformly.

Combining 1-3, we have $\sup_{\{Z_i,U_i\}\in\mathscr{G}_Z\times\mathscr{G}_U} |\hat{g}_2(\hat{U}_i) - g_2(U_i)| = O_p\left(L_n + \frac{L_{1n}}{h_2}\right)$. For $\hat{m}_{2j}(W_i) - m_{2j}(W_i)$, note that

$$\begin{aligned}
\hat{m}_{2j}(W_i) - m_{2j}(W_i) &= \frac{1}{nh_3^{D_3}\hat{f}_W(W_i)}\sum_{t=1}^n K_{3ti}\hat{\eta}_t X_{2t,j} - m_{2j}(W_i) \\
&= \left\{\frac{1}{nh_3^{D_3}f_W(W_i)}\sum_{t=1}^n K_{3ti}\underbrace{\Big\{(\hat{\eta}_t - \eta_t)X_{2t,j} + v_{m2t,j} + \big(m_{2j}(W_t) - m_{2j}(W_i)\big)\Big\}}_{C_{m2ti}}\right\}\big(1 + O_p(L_{3n})\big) \\
&= O_p(L_n), \qquad \text{(A.3)}
\end{aligned}$$

where the order can be found similarly to T_1 in part 1. For $\hat{\mu}_{2j}$, we have

$$\hat{\mu}_{2j} - \mu_{2j} = \frac{1}{n}\sum_{t=1}^n(\hat{\eta}_t - \eta_t)X_{2t,j} + \Big(\frac{1}{n}\sum_{t=1}^n \eta_t X_{2t,j} - \mathrm{E}(\eta_i X_{2i,j})\Big) = O_p(L_n) + O_p(n^{-1/2}) = O_p(L_n).$$

The uniform orders of $\hat{g}_1(\hat{U}_i)$, $\hat{m}_1(W_i)$, $\hat{\mu}_1$, $\hat{g}_3(\hat{U}_i)$, and $\hat{m}_3(W_i)$ can be found similarly by replacing $\hat{\eta}_t X_{2t,j}$ with $\hat{\eta}_t Y_t$ or $\hat{\eta}_t$, respectively. Thus, the details of these proofs are not be provided here.

□

Theorem 3 *Proof.* Note that $m = m_1 - m_2\beta - \beta_0$, $g = g_1 - g_2\beta - \beta_0$, where $m \equiv \big(m(W_1), \cdots, m(W_n)\big)'$, and g, m_1, g_1, m_2, g_2 and their associated estimators are defined similarly in vector forms. Denote $V_Y \equiv \sum_{k=\{m,g,\mu\}} V_{k1}$ and $V_X \equiv \sum_{k=\{m,g,\mu\}} V_{k2}$, where $V_{m1} \equiv \hat{m}_1 - m_1$, $V_{g1} \equiv \hat{g}_1 - g_1$, $V_{\mu 1} \equiv -(\hat{\mu}_1 - \mu_1)$, $V_{m2} \equiv \hat{m}_2 - m_2$, $V_{g2} \equiv \hat{g}_2 - g_2$, and $V_{\mu 2} \equiv -(\hat{\mu}_2 - \mu_2)$. Thus, since $\hat{Y} = Y^* - V_Y$, $\hat{X}_2 = X_2^* - V_X$, and $\hat{Y} - \hat{X}_2\beta = v - \sum_{k=\{m,g,\mu\}}(V_{k1} - V_{k2}\beta)$, we have

$$\hat{\beta} - \beta = \big(\frac{1}{n}\hat{X}_2'\hat{\eta}\hat{X}_2\big)^{-1}\frac{1}{n}\hat{X}_2'\hat{\eta}\big(\hat{Y} - \hat{X}_2\beta\big),$$

where
$$\begin{aligned}
\frac{1}{n}\hat{X}_2'\hat{\eta}\hat{X}_2 &= \frac{1}{n}X_2^{*\prime}\hat{\eta}X_2^* - \frac{1}{n}X_2^{*\prime}\hat{\eta}V_X - \frac{1}{n}V_X'\hat{\eta}X_2^* + \frac{1}{n}V_X'\hat{\eta}V_X \equiv \sum_{k=1}^4 A_k, \\
\frac{1}{n}\hat{X}_2'\hat{\eta}(\hat{Y} - \hat{X}_2\beta) &= \frac{1}{n}\hat{X}_2'\hat{\eta}v - \frac{1}{n}\hat{X}_2'\hat{\eta}(V_{m1} - V_{m2}\beta) - \frac{1}{n}\hat{X}_2'\hat{\eta}(V_{g1} - V_{g2}\beta) - \frac{1}{n}\hat{X}_2'\hat{\eta}(V_{\mu 1} - V_{\mu 2}\beta) \equiv \sum_{k=1}^4 B_k.
\end{aligned}$$

The proof has five steps:

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

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

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

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

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

Step 1: By uniform order of $|\hat{\eta}_i - \eta_i|$, Kolmogorov's LLN and A3, we have

$$A_1 = \frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_i X_{2i}^* X_{2i}^{*\prime} = \frac{1}{n}\sum_{i=1}^{n}\eta_i X_{2i}^* X_{2i}^{*\prime} + O_p(L_n)\frac{1}{n}\sum_{i=1}^{n}|X_{2i}^* X_{2i}^{*\prime}| \xrightarrow{p} \Phi_0,$$

where $\Phi_{0_{(j,k)}} \equiv \mathrm{E}(\eta_t X_{2t,j}^* X_{2t,k}^*) = \mathrm{E}\Big\{\eta_t(X_{2t,j} - m_{2j}(W_t) - g_{2j}(U_t) + \mu_{2j})(X_{2t,k} - m_{2k}(W_t) - g_{2k}(U_t) + \mu_{2k})\Big\} < \infty$, since $\{\eta_i X_{2i}^* X_{2i}^{*\prime}\}_{i=1}^n$ is an IID sequence, and $\mathrm{E}|\eta_i X_{2i,k}^* X_{2i,j}^*| < \infty$ due to (i). η_i is uniformly bounded; (ii). $\mathrm{E}|X_{2i,j}X_{2i,k}| \leq \big(\mathrm{E}(X_{2i,j}^2)\mathrm{E}(X_{2i,k}^2)\big)^{1/2} < \infty$ by Cauchy-Schwarz Inequality; (iii). $\mathrm{E}|X_{2i,j}m_{2k}(W_i)| \leq \big(\mathrm{E}(X_{2i,j}^2)\mathrm{E}(m_{2k}^2(W_i))\big)^{1/2}$; (iv). $\mathrm{E}(m_{2k}^2(W_i)) = \mathrm{E}\big(\mathrm{E}(\eta_i X_{2i,k}|W_i)^2\big) \leq \mathrm{E}\big(\mathrm{E}(\eta_i^2 X_{2i,k}^2|W_i)\big) = \mathrm{E}(\eta_i^2 X_{2i,k}^2) < \infty$. By the non-singularity of Φ_0 in A3, we have $A_1^{-1} \xrightarrow{p} \Phi_0^{-1}$. And for $-A_2 = \frac{1}{n}\sum_{i=1}^n \hat{\eta}_i X_{2i}^* V_{Xi}'$, the $(k,j)^{\text{th}}$ element is $-A_{2_{(k,j)}} = \frac{1}{n}\sum_{i=1}^n \hat{\eta}_i X_{2i,k}^* V_{Xi,j} \leq O_p(L_n + L_{1n}/h_2)\frac{1}{n}\sum_{i=1}^n |X_{2i,k}^*| = o_p(1)$ by Theorem 2. Similarly we have $A_3, A_4 = o_p(1)$. Thus, $\big(\frac{1}{n}\hat{X}_2'\hat{\eta}\hat{X}_2\big)^{-1} \xrightarrow{p} \Phi_0^{-1}$.

Step 2: We rewrite B_1 into four elements:

$$B_1 = \frac{1}{n}\sum_{i=1}^{n}\hat{X}_{2i}\hat{\eta}_i v_i = \frac{1}{n}\sum_{i=1}^{n}X_{2i}^*\eta_i v_i + \frac{1}{n}\sum_{i=1}^{n}V_{Xi}(\hat{\eta}_i - \eta_i)v_i + \frac{1}{n}\sum_{i=1}^{n}X_{2i}^*(\hat{\eta}_i - \eta_i)v_i + \frac{1}{n}\sum_{i=1}^{n}V_{Xi}\eta_i v_i \equiv \sum_{k=1}^{4}B_{1k},$$

and show that $\sqrt{n}B_1 \xrightarrow{d} \mathscr{N}(0,\Phi_1)$ by establishing that $\sqrt{n}B_{11} \xrightarrow{d} \mathscr{N}(0,\Phi_1)$, and $B_{12}, B_{13}, B_{14} = o_p(n^{-1/2})$.

First, by Levy's Central Limit Theorem and the Cramer-Wold device, we have $\sqrt{n}B_{11} \xrightarrow{d} \mathscr{N}(0,\Phi_1)$, since (i). $\{X_{2i}^*\eta_i v_i\}_{i=1}^n$ is IID; (ii). $\mathrm{E}(X_{2i}^*\eta_i v_i) = 0$; (iii). $\mathrm{E}(v_i^2|Z_i,U_i) = \sigma_v^2$; (iv). $\mathrm{V}(X_{2i}^*\eta_i v_i) = \mathrm{E}(X_{2i}^*\eta_i^2 v_i^2 X_{2i}^{*\prime}) = \sigma_v^2\mathrm{E}(\eta_i^2 X_{2i}^* X_{2i}^{*\prime})) \equiv \Phi_1 < \infty$, where $\Phi_{1_{(j,k)}} = \sigma_v^2\,\mathrm{E}(\eta_t^2 X_{2t,j}^* X_{2t,k}^*) = \sigma_v^2\,\mathrm{E}\Big\{\eta_t^2(X_{2t,j} - m_{2j}(W_t) - g_{2j}(U_t) + \mu_{2j})(X_{2t,k} - m_{2k}(W_t) - g_{2k}(U_t) + \mu_{2k})\Big\} < \infty$.

Second, given that $|V_{Xi}|, |\hat{\eta}_i - \eta_i| = O_p(L_n + L_{1n}/h_2)$, we have $B_{12} = O_p(L_n^2 + L_{1n}^2/h_2^2)\frac{1}{n}\sum_{i=1}^n |v_i| = o_p(n^{-1/2})$.

Third, the j^{th} element of B_{13} is $\frac{1}{n}\sum_{i=1}^n G(M_i)(\hat{\eta}_i(W_i,\hat{U}_i) - \eta_i(W_i,U_i))$, where $G(M_i) \equiv X_{2i,j}^* v_i$ and $M_i \equiv (X_i, Z_i, U_i, \varepsilon_i)$. Note that since $\mathrm{E}(v_i|X_i,Z_i,U_i) = 0$, $\mathrm{E}(G(M_i)|X_i,Z_i,U_i) = 0$. In addition, $\mathrm{E}(G^2(M_i)) = \mathrm{E}(X_{2i,j}^{*2}v_i^2) < \infty$ by A3. By A4, $G(M_i)$ is continuous, hence using Lemma 4, $B_{13} = o_p(n^{-1/2})$.

Fourth, for B_{14}, the j^{th} element can be written as

$$B_{14,j} = \frac{1}{n}\sum_{i=1}^{n}V_{Xi,j}\eta_i v_i = \frac{1}{n}\sum_{i=1}^{n}V_{m2i,j}\eta_i v_i + \frac{1}{n}\sum_{i=1}^{n}V_{g2i,j}\eta_i v_i + \frac{1}{n}\sum_{i=1}^{n}V_{\mu 2i,j}\eta_i v_i \equiv \sum_{k=1}^{3}B_{14k}.$$

We show that $B_{14k} = o_p(n^{-1/2})$ for $k = 1,2,3$.

Note that $B_{143} = -\frac{1}{n}\sum_{i=1}^n\big(\hat{\mu}_{2j} - \mu_{2j}\big)\eta_i v_i = -\big(\hat{\mu}_{2j} - \mu_{2j}\big)\frac{1}{n}\sum_{i=1}^n \eta_i v_i = O_p(L_n)O_p(n^{-1/2}) = o_p(n^{-1/2})$.

For B_{141}, given that $\big(nh_3^{D_3}f_W(W_i)\big)^{-1}\sum_{t=1}^n K_{3ti}C_{m2ti} = O_p(L_n)$, and by the decomposition of $\hat{m}_{2j}(W_i) - m_{2j}(W_i)$

in A.3 from the proof of Theorem 2, we have

$$B_{141} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i K_{3ti}}{h_3^{D_3} f_W(W_i)} C_{m2ti,j} + \frac{1}{n}\sum_{i=1}^{n}\left|\eta_i v_i\right| O_p(L_n) O_p(L_{3n}) \equiv \sum_{k=1}^{3} B_{141k} + o_p(n^{-1/2}),$$

where
$$B_{1411} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i K_{3ti}}{h_3^{D_3} f_W(W_i)}(\hat{\eta}_t - \eta_t) X_{2t,j}, \qquad B_{1412} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i K_{3ti}}{h_3^{D_3} f_W(W_i)} v_{m2t,j},$$
$$B_{1413} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i K_{3ti}}{h_3^{D_3} f_W(W_i)}\big(m_{2j}(W_t) - m_{2j}(W_i)\big).$$

1. We show that $B_{141k} = o_p(n^{-1/2})$ for $k = 1,2,3$.

1.1. Let $Q_t \equiv \frac{1}{n}\sum_{i=1}^{n}\big(h_3^{D_3} f_W(W_i)\big)^{-1}\eta_i v_i K_{3ti}$. So $B_{1411} = \frac{1}{n}\sum_{t=1}^{n}(\hat{\eta}_t - \eta_t) X_{2t,j} Q_t$. By Lemma 3, we can show that $Q_t = O_p(L_{3n})$ uniformly over $\mathscr{G}_W$, given A3 and $\mathrm{E}(Q_t) = 0$. Given $\hat{\eta}_t - \eta_t = \eta_t O_p(L_n)$ uniformly, we have $B_{1411} = O_p(L_n) O_p(L_{3n}) \frac{1}{n}\sum_{t=1}^{n}|\eta_t X_{2t,j}| = o_p(n^{-1/2})$ by A5.

1.2. Let $B_{1412} \equiv \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\psi_{nit} \equiv E_{1n} + E_{2n}$, where $\psi_{nit} \equiv \big(h_3^{D_3} f_W(W_i)\big)^{-1}\eta_i v_i K_{3ti} v_{m2t,j}$. Thus, $E_{1n} = \frac{1}{n^2}\sum_{i=1}^{n}\psi_{nii} = o_p(n^{-1/2})$ by Chebyshev's Inequality, since $\mathrm{E}(E_{1n}) = 0$, $\mathrm{V}(E_{1n}) = \frac{1}{n^3}\mathrm{E}(\psi_{nii}^2) = O\big(n^{-3}h_3^{-D_3}\big) = o(n^{-1})$. And $|E_{2n}| \le C|U_n|$, where U_n is a U-statistic of degree 2 such that $U_n = \binom{n}{2}^{-1}\sum_{\substack{i=1\\ i<t}}^{n}\sum_{t=1}^{n}\phi_{nit}$ with $\phi_{nit} \equiv \psi_{nit} + \psi_{nti}$.

- $\theta_n, \sigma_{1n}^2 = 0$, as $\mathrm{E}(v_i|W_i) = \mathrm{E}(v_{m2t,j}|M_t) = 0$;
- $\sigma_{2n}^2 = \mathrm{V}(\phi_{nit}) \le C\mathrm{E}(\psi_{nit}^2) \le C h_3^{-2D_3}\sigma_v^2\sigma_{X1,j}^2\mathrm{E}(K_{3ti}^2) = O(h_3^{-D_3})$;
- $H_n^{(1)} = 0$, $H_n^{(2)} = O_p\big((\sigma_{2n}^2/n^2)^{1/2}\big) = O_p(n^{-1/2}(nh_3^{D_3})^{-1/2}) = o_p(n^{-1/2})$.

We have $B_{1412} = o_p(n^{-1/2})$.

1.3. Given $B_{1413} = \frac{1}{n^2}\sum_{\substack{i=1\\ i\neq t}}^{n}\sum_{t=1}^{n}\psi_{nit}$, where $\psi_{nit} = h_3^{-D_3} f_W^{-1}(W_i)\eta_i v_i K_{3ti}\big(m_{2j}(W_t) - m_{2j}(W_i)\big)$, we have $|B_{1413}| \le C|U_n|$, where $U_n = \binom{n}{2}^{-1}\sum_{\substack{i=1\\ i<t}}^{n}\sum_{t=1}^{n}\phi_{nit}$, with $\phi_{nit} \equiv \psi_{nit} + \psi_{nti}$, is a U-statistic of degree 2.

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

We have $B_{1413} = o_p(n^{-1/2})$.

Combining 1.1-1.3, we have $B_{141} = o_p(n^{-1/2})$.

2. For B_{142}, as it is shown in the proof of Theorem 2, $V_{g2i,j} = \hat{g}_{2j}(\hat{U}_i) - g_{2j}(U_i) \equiv \left(\sum_{k=1}^{3} T_k\right)(1 + O_p(L_{2n}))$, where $T_1 = O_p(L_n)$, $T_2 = O_p(L_{1n}/h_2)$, and $T_3 = o_p(n^{-1/2})$. Thus, by the decomposition of $V_{g2i,j}$ in A.2, we have

$$
\begin{aligned}
B_{142} &= \frac{1}{n}\sum_{i=1}^{n} V_{g2i,j}\eta_i v_i = \sum_{k=1}^{3} B_{142k} + \frac{1}{n}\sum_{i=1}^{n}|\eta_i v_i|\left(o_p(n^{-\frac{1}{2}}) + \big(O_p(L_n) + O_p\big(L_{1n}/h_2\big)\big)O_p(L_{2n})\right) \\
&\equiv \sum_{k=1}^{3} B_{142k} + o_p(n^{-1/2}) \quad \text{by A5},
\end{aligned}
$$

where
$$
B_{1421} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i}{h_2^{D_2} f_U(U_i)} K_{2ti} C_{g2ti}, \qquad B_{1422} = -\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i}{h_2^{D_2+1} f_U(U_i)} \mathbf{J}K_{2ti}(\hat{U}_i - U_i) C_{g2ti},
$$
$$
B_{1423} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i}{h_2^{D_2+1} f_U(U_i)} \mathbf{J}K_{2ti}(\hat{U}_t - U_t) C_{g2ti}, \quad C_{g2ti} = (\hat{\eta}_t - \eta_t)X_{2t,j} + v_{g2t,j} + \big(g_{2j}(U_t) - g_{2j}(U_i)\big).
$$

Similar to B_{141} we just analyzed, we have $B_{1421} = o_p(n^{-1/2})$, with U_i replacing W_i. B_{1422} and B_{1423} are similar in structure, so here we only show that $B_{1422} = o_p(n^{-1/2})$. Given the three components in C_{g2ti}, let $B_{1422} = \sum_{k=1}^{3} B_{1422k}$.

We show that $B_{1422k} = o_p(n^{-1/2})$ for $k = 1, 2, 3$.

2.1.
$$
\begin{aligned}
B_{14221} &= -\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i}{h_2^{D_2+1} f_U(U_i)} \mathbf{J}K_{2ti}(\hat{U}_i - U_i)(\hat{\eta}_t - \eta_t) X_{2t,j} \\
&\leq O_p(L_n) O_p\left(\frac{L_{1n}}{h_2}\right) \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{d=1}^{D_2}\frac{\left|\eta_i v_i \eta_t X_{2t,j} D_d K_{2ti}\right|}{h_2^{D_2} f_U(U_i)} \\
&= O_p(L_n) O_p\left(\frac{L_{1n}}{h_2}\right) = o_p(n^{-1/2}), \quad \text{by A5}.
\end{aligned}
$$

2.2. By A.1 in the proof of Theorem 1, we have

$$
\begin{aligned}
B_{14222} &= -\sum_{d=1}^{D_2}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i v_{g2t,j} D_d K_{2ti}}{h_2^{D_2+1} f_U(U_i)}(\hat{U}_{id} - U_{id}) \\
&= \sum_{d=1}^{D_2}\left\{\frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{\eta_i v_i v_{g2t,j} D_d K_{2ti} K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)}\Big(U_{ld} + \big(\Pi_d(Z_l) - \Pi_d(Z_i)\big)\Big)\right. \\
&\quad \left. + O_p(L_{1n}^2)\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\left|\frac{\eta_i v_i v_{g2t,j} D_d K_{2ti}}{h_2^{D_2+1} f_U(U_i)}\right|\right\} \\
&\equiv \sum_{d=1}^{D_2}(T_{1d} + T_{2d}) + o_p(n^{-1/2}).
\end{aligned}
$$

where the last equality follows by Markov's Inequality and that $O_p(L_{1n}^2/h_2) = o_p(n^{-1/2})$ by A5.

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

2.2.1.
$$
T_{1d} = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{\eta_i v_i v_{g2t,j} D_d K_{2ti} K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{ld} \equiv \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\psi_{nitl}.
$$

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

- $\theta_n, \mathrm{E}(\phi_{nitl}|P_i), \mathrm{E}(\phi_{nitl}|P_i,P_t) = 0, \quad \text{as} \quad \mathrm{E}(v_i|W_i,U_i), \mathrm{E}(v_{g2t,j}|U_t), \mathrm{E}(U_{ld}|Z_l) = 0;$
- $\sigma_{1n}^2, \sigma_{2n}^2 = 0, \quad \sigma_{3n}^2 = \mathrm{V}(\phi_{nitl}) \le C\mathrm{E}(\psi_{nitl}^2) = O\big((h_1^{D_1} h_2^{D_2+2})^{-1}\big);$
- $H_n^{(1)}, H_n^{(2)} = 0, H_n^{(3)} = O_p\Big(\big(\sigma_{3n}^2/n^3\big)^{1/2}\Big) = O_p\Big(\big(n^3 h_1^{D_1} h_2^{D_2+2}\big)^{-1/2}\Big) = o_p(n^{-1/2}).$

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

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

$$\text{if} \quad i=t=l, \quad \frac{1}{n^3}\sum_{i=1}^n \psi_{niii} = \frac{1}{n^3}\sum_{i=1}^n \frac{\eta_i v_i v_{g2i,j} D_d K_2(0) K_1(0)}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{id} = O_p\Big(\big(n^2 h_1^{D_1} h_2^{D_2+1}\big)^{-1}\Big) = o_p(n^{-1/2});$$

$$\text{if} \quad i=t\neq l, \quad \frac{1}{n^3}\sum_{\substack{i=1 \\ i\neq l}}^n \sum_{l=1}^n \psi_{niil} = \frac{1}{n^3}\sum_{\substack{i=1 \\ i\neq l}}^n \sum_{l=1}^n \frac{\eta_i v_i v_{g2i,j} D_d K_2(0) K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{ld} = O_p\Big(\big(n h_2^{D_2+1}\big)^{-1}\Big) = o_p(n^{-1/2});$$

$$\text{if} \quad i=l\neq t, \quad \frac{1}{n^3}\sum_{\substack{i=1 \\ i\neq t}}^n \sum_{t=1}^n \psi_{niti} = \frac{1}{n^3}\sum_{\substack{i=1 \\ i\neq t}}^n \sum_{t=1}^n \frac{\eta_i v_i v_{g2t,j} D_d K_{2ti} K_1(0)}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{id} = O_p\Big(\big(n h_1^{D_1} h_2\big)^{-1}\Big) = o_p(n^{-1/2});$$

$$\text{if} \quad i\neq t=l, \quad \frac{1}{n^3}\sum_{\substack{i=1 \\ i\neq t}}^n \sum_{t=1}^n \psi_{nitt} = \frac{1}{n^3}\sum_{\substack{i=1 \\ i\neq t}}^n \sum_{t=1}^n \frac{\eta_i v_i v_{g2t,j} D_d K_{2ti} K_{1ti}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{td} = O_p\Big(\big(n h_2\big)^{-1}\Big) = o_p(n^{-1/2}).$$

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

2.2.2. $$T_{2d} = \frac{1}{n^3}\sum_{i=1}^n\sum_{t=1}^n\sum_{l=1}^n \frac{\eta_i v_i v_{g2t,j} D_d K_{2ti} K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} \big(\Pi_d(Z_l) - \Pi_d(Z_i)\big) \equiv \frac{1}{n^3}\sum_{i=1}^n\sum_{t=1}^n\sum_{l=1}^n \psi_{nitl}.$$

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

- $\theta_n, \mathrm{E}(\phi_{nitl}|P_i), \mathrm{E}(\psi_{nitl}|P_i,P_l), \mathrm{E}(\psi_{nitl}|P_t,P_l) = 0$, as $\mathrm{E}(v_i|W_i,U_i)$, $\mathrm{E}(v_{g2t,j}|U_t) = 0$,
- $\mathrm{E}(\psi_{nitl}|P_i,P_t) = \big(h_2^{D_2+1} f_U(U_i) f_Z(Z_i)\big)^{-1} \eta_i v_i v_{g2t,j} D_d K_{2ti} \mathrm{E}\Big(h_1^{-D_1} K_{1li}\big(\Pi_d(Z_l) - \Pi_d(Z_i)\big)\big|Z_i\Big);$
- $\sigma_{1n}^2 = 0$, $\sigma_{2n}^2 \le \mathrm{E}\left(\mathrm{E}^2(\psi_{nitl}|P_i,P_t)\right) = O\big(h_1^{2s_1}/h_2^{D_2+2}\big)$, $\sigma_{3n}^2 = \mathrm{V}(\phi_{nitl}) \le C\mathrm{E}(\psi_{nitl}^2) = O\big((h_1^{D_1-2} h_2^{D_2+2})^{-1}\big);$
- $H_n^{(1)} = 0$, $H_n^{(2)} = O_p\Big(\big(\sigma_{2n}^2/n^2\big)^{1/2}\Big) = O_p\Big(h_1^{s_1}\big(n^2 h_2^{D_2+2}\big)^{-1/2}\Big) = o_p(n^{-1/2})$,
  $H_n^{(3)} = O_p\big((\sigma_{3n}^2/n^3)^{1/2}\Big) = O_p\Big(\big(n^3 h_1^{D_1-2} h_2^{D_2+2}\big)^{-1/2}\Big) = o_p(n^{-1/2}).$

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

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

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

$$\text{if} \quad i=t\neq l, \ \frac{1}{n^3}\sum_{\substack{i=1 \\ i\neq l}}^n \sum_{l=1}^n \psi_{niil} = \frac{1}{n^3}\sum_{\substack{i=1 \\ i\neq l}}^n \sum_{l=1}^n \frac{\eta_i v_i v_{g2i,j} D_d K_2(0) K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} \big(\Pi_d(Z_l) - \Pi_d(Z_i)\big) = O_p\Big(h_1\big(n h_2^{D_2+1}\big)^{-1}\Big);$$

$$\text{if} \quad i\neq t=l, \ \frac{1}{n^3}\sum_{\substack{i=1 \\ i\neq t}}^n \sum_{t=1}^n \psi_{nitt} = \frac{1}{n^3}\sum_{\substack{i=1 \\ i\neq t}}^n \sum_{t=1}^n \frac{\eta_i v_i v_{g2t,j} D_d K_{2ti} K_{1ti}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} \big(\Pi_d(Z_t) - \Pi_d(Z_i)\big) = O_p\Big(h_1\big(n h_2\big)^{-1}\Big).$$

We have $B_{14222} = o_p(n^{-1/2})$.

2.3. Similar to part 2.2, we have

$$
\begin{aligned}
B_{14223} &= -\sum_{d=1}^{D_2}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i D_d K_{2ti}\big(g_{2j}(U_t)-g_{2j}(U_i)\big)}{h_2^{D_2+1} f_U(U_i)}(\hat{U}_{id}-U_{id}) \\
&= \sum_{d=1}^{D_2}\left\{\frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{\eta_i v_i D_d K_{2ti}K_{1li}\big(g_{2j}(U_t)-g_{2j}(U_i)\big)}{h_1^{D_1}h_2^{D_2+1} f_U(U_i) f_Z(Z_i)}\Big(U_{ld}+\big(\Pi_d(Z_l)-\Pi_d(Z_i)\big)\Big)\right. \\
&\quad \left. +O_p(L_{1n}^2)\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\left|\frac{\eta_i v_i D_d K_{2ti}\big(g_{2j}(U_t)-g_{2j}(U_i)\big)}{h_2^{D_2+1} f_U(U_i)}\right|\right\} \equiv \sum_{d=1}^{D_2}(T_{1d}+T_{2d})+o_p(n^{-1/2}).
\end{aligned}
$$

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

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

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

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

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

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

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

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

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

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

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

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

- $\theta_n, \mathrm{E}(\psi_{nitl}|P_t), \mathrm{E}(\psi_{nitl}|P_l), \mathrm{E}(\psi_{nitl}|P_t,P_l) = 0$, as $\mathrm{E}(v_i|Z_i,U_i,W_i) = 0$;
- $\mathrm{E}(\psi_{nitl}|P_i) = \big(f_U(U_i) f_Z(Z_i)\big)^{-1}\eta_i v_i \mathrm{E}\Big(h_1^{-D_1}h_2^{-D_2-1}D_d K_{2ti}K_{1li}\big(g_{2j}(U_t)-g_{2j}(U_i)\big)\big(\Pi_d(Z_l)-\Pi_d(Z_i)\big)\Big|P_i\Big)$,
  $\mathrm{E}(\psi_{nitl}|P_i,P_t) = \big(h_2^{D_2+1} f_U(U_i) f_Z(Z_i)\big)^{-1}\eta_i v_i D_d K_{2ti}\big(g_{2j}(U_t)-g_{2j}(U_i)\big)\mathrm{E}\Big(h_1^{-D_1}K_{1li}\big(\Pi_d(Z_l)-\Pi_d(Z_i)\big)\Big|Z_i\Big)$,
  $\mathrm{E}(\psi_{nitl}|P_i,P_l) = \big(h_1^{D_1} f_U(U_i) f_Z(Z_i)\big)^{-1}\eta_i v_i K_{1li}\big(\Pi_d(Z_l)-\Pi_d(Z_i)\big)\mathrm{E}\Big(h_2^{-D_2-1}D_d K_{2ti}\big(g_{2j}(U_t)-g_{2j}(U_i)\big)\Big|U_i\Big)$;

- $\sigma_{1n}^2 \le \mathrm{E}\big(\mathrm{E}^2(\psi_{nitl}|P_i)\big) = O(h_1^{2s_1}) = o(1)$, $\sigma_{2n}^2 \le C\mathrm{E}\left(\mathrm{E}^2(\psi_{nitl}|P_i,P_t)+\mathrm{E}^2(\psi_{nitl}|P_i,P_l)\right) = O\big(h_1^{2s_1}/h_2^{D_2} + 1/h_1^{D_1-2}\big)$, $\sigma_{3n}^2 = \mathrm{V}(\phi_{nitl}) \le C\mathrm{E}(\psi_{nitl}^2) = O_p\big((h_1^{D_1-2}h_2^{D_2})^{-1}\big)$;
- $H_n^{(1)} = O_p\big((\sigma_{1n}^2/n)^{1/2}\big) = o_p(n^{-1/2})$, $H_n^{(2)} = O_p\big((\sigma_{2n}^2/n^2)^{1/2}\big) = O_p\Big(h_1^{s_1}\big(n^2h_2^{D_2}\big)^{-1/2} + \big(n^2h_1^{D_1-2}\big)^{-1/2}\Big)$ $= o_p(n^{-1/2})$, $H_n^{(3)} = O_p\big((\sigma_{3n}^2/n^3)^{1/2}\big) = O_p\Big(\big(n^3h_1^{D_1-2}h_2^{D_2}\big)^{-1/2}\Big) = o_p(n^{-1/2})$.

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

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

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

if $\quad i\neq t=l,$

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

We have $B_{14223} = o_p(n^{-1/2})$. By 2.1-2.3, we have $B_{142} = o_p(n^{-1/2})$.

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

Step 3: We first show that $B_4 = o_p(n^{-1/2})$. Note that

$$-B_4 = \frac{1}{n}\hat{X}_2'\hat{\eta}(V_{\mu 1}-V_{\mu 2}\beta) = \frac{1}{n}\hat{X}_2'\eta V_{\mu 1} - \frac{1}{n}\hat{X}_2'\eta V_{\mu 2}\beta + \frac{1}{n}\hat{X}_2'(\hat{\eta}-\eta)(V_{\mu 1}-V_{\mu 2}\beta) \equiv \sum_{k=1}^3 B_{4k}.$$

By Theorems 1 and 2, we have $|\hat{\eta}_i-\eta_i|$, $V_{\mu 2i}$, $V_{\mu 1i} = O_p(L_n)$. Given that $V_{\mu 1i}$ is the same across i, we have $B_{41} = V_{\mu 1i}\big(\frac{1}{n}\sum_{i=1}^n X_{2i}^*\eta_i - \frac{1}{n}\sum_{i=1}^n V_{Xi}\eta_i\big) = O_p(L_n)\big(O_p(n^{-1/2})+O_p(L_n)\big) = o_p(n^{-1/2})$ by A5. $B_{42} = o_p(n^{-1/2})$ follows similarly, and $B_{43} = O_p(L_n^2) = o_p(n^{-1/2})$ by A5.

Then, we show that $B_4 = o_p(n^{-1/2})$. Note that

$$-B_2 = \frac{1}{n}\hat{X}_2'\eta V_{m1} - \frac{1}{n}\hat{X}_2'\eta V_{m2}\beta + \frac{1}{n}\hat{X}_2'(\hat{\eta}-\eta)(V_{m1}-V_{m2}\beta) \equiv \sum_{k=1}^3 B_{2k}.$$

$B_{23} = O_p(L_n^2) = o_p(n^{-1/2})$ by A5. B_{22} is of the same structure as B_{21}, thus we only show that $B_{21} = o_p(n^{-1/2})$.

Note that $B_{21} = \frac{1}{n}\sum_{i=1}^n X_{2i}^*\eta_i V_{m1i} - \frac{1}{n}\sum_{i=1}^n V_{Xi}\eta_i V_{m1i} \equiv B_{21}' + o_p(n^{-1/2})$ by Theorem 2. By the decomposition of V_{m1i}, similar to V_{m2i} given in A.3 from the proof of Theorem 2, we have the j^{th} element of B_{21}' as

$$B_{21}' = \frac{1}{n^2}\sum_{i=1}^n\sum_{t=1}^n \frac{\eta_i X_{2i,j}^* K_{3ti}}{h_3^{D_3}f_W(W_i)}C_{m1ti} + O_p(L_{3n})O_p(L_n)\frac{1}{n}\sum_{i=1}^n\big|\eta_i X_{2i,j}^*\big| \equiv \sum_{k=1}^3 B_{21k} + o_p(n^{-\frac{1}{2}}),$$

where

$$B_{211} = \frac{1}{n^2}\sum_{i=1}^n\sum_{t=1}^n \frac{\eta_i X_{2i,j}^* K_{3ti}}{h_3^{D_3}f_W(W_i)}(\hat{\eta}_t-\eta_t)Y_t, \qquad B_{212} = \frac{1}{n^2}\sum_{i=1}^n\sum_{t=1}^n \frac{\eta_i X_{2i,j}^* K_{3ti}}{h_3^{D_3}f_W(W_i)}v_{m1t},$$

$$B_{213} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i X^*_{2i,j}K_{3ti}}{h_3^{D_3} f_W(W_i)}\big(m_1(W_t)-m_1(W_i)\big).$$

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

3.1. Let $Q_t \equiv \frac{1}{n}\sum_{i=1}^n \big(h_3^{D_3} f_W(W_i)\big)^{-1}\eta_i X^*_{2i,j}K_{3ti}$. So $B_{211} = \frac{1}{n}\sum_{t=1}^n(\hat{\eta}_t-\eta_t)Y_t Q_t$. By Lemma 3, we can show that $Q_t = O_p(L_{3n})$ uniformly over $\mathscr{G}_W$, given A3 and $\mathrm{E}(Q_t)=0$. Given $\hat{\eta}_t-\eta_t = \eta_t O_p(L_n)$ uniformly, we have $B_{211} = O_p(L_n)O_p(L_{3n})\frac{1}{n}\sum_{t=1}^n|\eta_t Y_t| = o_p(n^{-1/2})$ by A5.

3.2. $B_{212} = \frac{1}{n^2}\sum_{i=1}^n\sum_{t=1}^n h_3^{-D_3} f_W^{-1}(W_i)\eta_i X^*_{2i,j}K_{3ti}v_{m1t} \equiv \frac{1}{n^2}\sum_{i=1}^n\sum_{t=1}^n\psi_{nit} \equiv E_{1n}+E_{2n}$, where $E_{1n} = \frac{1}{n^2}\sum_{i=1}^n\psi_{nii} = \frac{1}{n^2}\sum_{i=1}^n h_3^{-D_3} f_W^{-1}(W_i)\eta_i X^*_{2i,j}K_3(0)v_{m1i} = O_p\big((nh_3^{D_3})^{-1}\big) = o_p(n^{-1/2})$, and $|E_{2n}| \le C|U_n|$ with $U_n = \binom{n}{2}^{-1}\sum_{\substack{i=1\\ i\neq t}}^n\sum_{t=1}^n\psi_{nit} = \theta_n + 2H_n^{(1)} + H_n^{(2)}$, a U-statistic of degree 2.

- θ_n, $\sigma^2_{1n} = 0$, as $\mathrm{E}(\eta_i X^*_{2i,j}|W_i)$, $\mathrm{E}(v_{m1t}|W_t) = 0$;
- $\sigma^2_{2n} = \mathrm{V}(\phi_{nit}) \le C\mathrm{E}(\psi^2_{nit}) = O(h_3^{-D_3})$;
- $H_n^{(1)} = 0$, $H_n^{(2)} = O_p\big((\sigma^2_{2n}/n^2)^{1/2}\big) = O_p((n^2h_3^{D_3})^{-1/2}) = o_p(n^{-1/2})$.

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

3.3. $|B_{213}| \le C|U_n|$, where $U_n = \binom{n}{2}^{-1}\sum_{\substack{i=1\\ i\neq t}}^n\sum_{t=1}^n\psi_{nit}$ is a U-statistic of degree 2, with $\psi_{nit} \equiv h_3^{-D_3} f_W^{-1}(W_i)\eta_i X^*_{2i,j}K_{3ti}$ $\big(m_1(W_t)-m_1(W_i)\big)$.

- $\theta_n, \mathrm{E}(\psi_{nit}|P_t) = 0$, as $\mathrm{E}(\eta_i X^*_{2i,j}|W_i) = 0$;
- $\phi_{1n} = \mathrm{E}(\psi_{nit}|P_i) = f_W^{-1}(W_i)\eta_i X^*_{2i,j}\mathrm{E}\big(h_3^{-D_3}K_{3ti}(m_1(W_t)-m_1(W_i))\big|W_i\big) \le Ch_3^{s_3}\eta_i X^*_{2i,j}$ uniformly over W_i;
- $\sigma^2_{1n} = \mathrm{E}(\phi^2_{1n}) = O(h_3^{2s_3}) = o(1)$, $\quad\sigma^2_{2n} = \mathrm{V}(\phi_{nit}) \le C\mathrm{E}(\psi^2_{nit}) = O\big(h_3^{-D_3+2}\big)$;
- $H_n^{(1)} = O_p\big((\sigma^2_{1n}/n)^{1/2}\big) = o_p(n^{-1/2})$, $H_n^{(2)} = O_p\big((\sigma^2_{2n}/n^2)^{1/2}\big) = O_p(n^{-1/2}(nh_3^{D_3-2})^{-1/2}) = o_p(n^{-1/2})$.

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

By 3.1-3.3, we have $B_{21} = o_p(n^{-1/2})$.

Step 4: For B_3, we have $-B_3 = \frac{1}{n}\hat{X}_2'\eta(V_{g1}-V_{g2}\beta) + o_p(n^{-1/2}) \equiv B_{31}+B_{32}+o_p(n^{-1/2})$. We will focus on B_{31} here, since B_{32} has a similar structure to B_{31} and could be analyzed accordingly. By Theorem 2, we have $B_{31} = \frac{1}{n}\sum_{i=1}^n X^*_{2i}\eta_i V_{g1i} - \frac{1}{n}\sum_{i=1}^n V_{Xi}\eta_i V_{g1i} \equiv B'_{31} + o_p(n^{-1/2})$. Similar to A.2 given in the proof of Theorem 2, by Taylor's Theorem, we have

$$V_{g1i} = \hat{g}_1(\hat{U}_i) - g_1(U_i) = \Bigg\{\frac{1}{nh_2^{D_2}f_U(U_i)}\sum_{t=1}^n K_{2ti}C_{g1ti} + \frac{1}{nh_2^{D_2+1}f_U(U_i)}\sum_{t=1}^n \mathbf{J}K_{2ti}\Big(\hat{U}_t - U_t - \big(\hat{U}_i - U_i\big)\Big)C_{g1ti}$$
$$+ \frac{1}{nh_2^{D_2}f_U(U_i)}\sum_{t=1}^n R_{ti}C_{g1ti}\Bigg\}\Big(1+O_p(L_{2n})\Big),$$

where $C_{g1ti} \equiv (\hat{\eta}_t - \eta_t)Y_t + v_{g1t} + \big(g_1(U_t) - g_1(U_i)\big)$, and R_{ti} is the remainder term of a Taylor's expansion of $\hat{K}_{2ti}$ at $(U_t - U_i)/h_2$.

Similar to the T_3 term in the proof of Theorem 2, we have $(nh_2^{D_2} f_U(U_i))^{-1} \sum_{t=1}^n R_{ti} C_{g1ti} = o_p(n^{-1/2})$ uniformly. Thus, we have the j^{th} element of B'_{31} as

$$\begin{aligned} B'_{31,j} &= \frac{1}{n}\sum_{i=1}^n X^*_{2i,j}\eta_i V_{g1i} = \sum_{k=1}^3 B_{31k} + \frac{1}{n}\sum_{i=1}^n |\eta_i X^*_{2i,j}| \Big(o_p(n^{-\frac{1}{2}}) + \big(O_p(L_n) + O_p(L_{1n}/h_2)\big) O_p(L_{2n}) \Big) \\ &\equiv \sum_{k=1}^3 B_{31k} + o_p(n^{-1/2}) \quad \text{by A5}, \end{aligned}$$

where

$$\begin{aligned} B_{311} &= \frac{1}{n^2}\sum_{i=1}^n\sum_{t=1}^n \frac{\eta_i X^*_{2i,j}}{h_2^{D_2} f_U(U_i)} K_{2ti} C_{g1ti}, \qquad B_{312} = -\frac{1}{n^2}\sum_{i=1}^n\sum_{t=1}^n \frac{\eta_i X^*_{2i,j}}{h_2^{D_2+1} f_U(U_i)} \mathbf{J}K_{2ti}\big(\hat{U}_i - U_i\big) C_{g1ti}, \\ B_{313} &= \frac{1}{n^2}\sum_{i=1}^n\sum_{t=1}^n \frac{\eta_i X^*_{2i,j}}{h_2^{D_2+1} f_U(U_i)} \mathbf{J}K_{2ti}\big(\hat{U}_t - U_t\big) C_{g1ti}. \end{aligned}$$

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

$$a_{1ni,j} = \sum_{d=1}^{D_2} \frac{U_{id}}{2h_1^{D_1} h_2^{D_2}} \mathrm{E}\left(\frac{\eta_l X^*_{2l,j} D_d K_{2tl} K_{1il}}{f_U(U_l) f_Z(Z_l)} \mathbf{J} g_1(U_l) \Big(\frac{U_t - U_l}{h_2}\Big) \middle| Z_i \right).$$

B_{311} is of similar structure as B_{141} with U_i replacing W_i, $\eta_i X^*_{2i,j}$ replacing $\eta_i v_i$, C_{g1ti} replacing $C_{m2ti,j}$, and $\mathrm{E}(\eta_i X^*_{2i,j}|U_i) = 0$ replacing $E(\eta_i v_i|W_i) = 0$. By the same arguments in $1.1 - 1.3$, we have $B_{311} = o_p(n^{-1/2})$. Given the three components in C_{g1ti}, let $-B_{312} \equiv \sum_{k=1}^3 B_{312k}$, with

$$\begin{aligned} B_{3121} &= \frac{1}{n^2}\sum_{i=1}^n\sum_{t=1}^n \frac{\eta_i X^*_{2i,j}}{h_2^{D_2+1} f_U(U_i)} \mathbf{J}K_{2ti}\big(\hat{U}_i - U_i\big)(\hat{\eta}_t - \eta_t)Y_t, \qquad B_{3122} = \frac{1}{n^2}\sum_{i=1}^n\sum_{t=1}^n \frac{\eta_i X^*_{2i,j}}{h_2^{D_2+1} f_U(U_i)} \mathbf{J}K_{2ti}\big(\hat{U}_i - U_i\big) v_{g1t}, \\ B_{3123} &= \frac{1}{n^2}\sum_{i=1}^n\sum_{t=1}^n \frac{\eta_i X^*_{2i,j}}{h_2^{D_2+1} f_U(U_i)} \mathbf{J}K_{2ti}\big(\hat{U}_i - U_i\big)\big(g_1(U_t) - g_1(U_i)\big). \end{aligned}$$

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

4.1. Given $\hat{\eta}_t - \eta_t = O_p(L_n)$ and $\hat{U}_i - U_i = O_p(L_{1n})$ uniformly, by Markov's Inequality and A5, we have

$$B_{3121} = O_p(L_n) O_p\left(\frac{L_{1n}}{h_2}\right) \frac{1}{n^2}\sum_{i=1}^n\sum_{t=1}^n\sum_{d=1}^{D_2} \frac{\left|\eta_i X^*_{2i,j} \eta_t Y_t D_d K_{2ti}\right|}{h_2^{D_2} f_U(U_i)} = o_p(n^{-1/2}).$$

4.2. By A.1 in the proof of Theorem 2, we have

$$\begin{aligned} B_{3122} &= \sum_{d=1}^{D_2} \frac{1}{n^2}\sum_{i=1}^n\sum_{t=1}^n \frac{\eta_i X^*_{2i,j} v_{g1t} D_d K_{2ti}}{h_2^{D_2+1} f_U(U_i)} (\hat{U}_{id} - U_{id}) \\ &= -\sum_{d=1}^{D_2} \left\{ \frac{1}{n^3}\sum_{i=1}^n\sum_{t=1}^n\sum_{l=1}^n \frac{\eta_i X^*_{2i,j} v_{g1t} D_d K_{2ti} K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} \Big(U_{ld} + \big(\Pi_d(Z_l) - \Pi_d(Z_i)\big) \Big) \right. \end{aligned}$$

$$+ O_p(L_{1n}^2)\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\left|\frac{\eta_i X_{2i,j}^* v_{g1t} D_d K_{2ti}}{h_2^{D_2+1} f_U(U_i)}\right|\Bigg\} \equiv -\sum_{d=1}^{D_2}(T_{1d}+T_{2d}) + o_p(n^{-1/2}).$$

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

4.2.1. $T_{1d} = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{\eta_i X_{2i,j}^* v_{g1t} D_d K_{2ti} K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{ld} \equiv \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\psi_{nitl}.$

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

- θ_n, σ_{1n}^2, $\mathrm{E}(\psi_{nitl}|P_i,P_t)$, $\mathrm{E}(\psi_{nitl}|P_i,P_l) = 0$, as $\mathrm{E}(v_{g1t}|U_t)$, $\mathrm{E}(U_{ld}|Z_l) = 0$;
- $\phi_{2n} = \mathrm{E}(\psi_{nitl}|P_t,P_l) = \frac{v_{g1t}U_{ld}}{h_1^{D_1}h_2^{D_2+1}}\mathrm{E}\left(\frac{\eta_i X_{2i,j}^* D_d K_{2ti} K_{1li}}{f_U(U_i)f_Z(Z_i)}\Big|Z_l,U_t\right) \leq \frac{C|v_{g1t}U_{ld}|}{h_2}$;
- $\sigma_{2n}^2 \leq \mathrm{E}(\phi_{2n}^2) = O(h_2^{-2})$, $\sigma_{3n}^2 = \mathrm{V}(\phi_{nitl}) \leq C\mathrm{E}(\psi_{nitl}^2) = O_p\big((h_1^{D_1}h_2^{D_2+2})^{-1}\big)$;
- $H_n^{(1)} = 0$, $H_n^{(2)} = O_p\big((\sigma_{2n}^2/n^2)^{1/2}\big) = O_p\big((nh_2^2)^{-1}\big) = o_p(n^{-1/2})$,
  $H_n^{(3)} = O_p\big((\sigma_{3n}^2/n^3)^{1/2}\big) = O_p\Big(\big(n^3 h_1^{D_1}h_2^{D_2+2}\big)^{-1/2}\Big) = o_p(n^{-1/2})$.

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

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

if $i=t=l$, $\quad \frac{1}{n^3}\sum_{i=1}^{n}\psi_{niii} = \frac{1}{n^3}\sum_{i=1}^{n}\frac{\eta_i X_{2i,j}^* v_{g1t} D_d K_2(0) K_1(0)}{h_1^{D_1}h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{id} = O_p\Big(\big(n^2 h_1^{D_1}h_2^{D_2+1}\big)^{-1}\Big) = o_p(n^{-1/2})$;

if $i=t\neq l$, $\quad \frac{1}{n^3}\sum_{\substack{i=1\\ i\neq l}}^{n}\sum_{l=1}^{n}\psi_{niil} = \frac{1}{n^3}\sum_{\substack{i=1\\ i\neq l}}^{n}\sum_{l=1}^{n}\frac{\eta_i X_{2i,j}^* v_{g1t} D_d K_2(0) K_{1li}}{h_1^{D_1}h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{ld} = O_p\Big(\big(nh_2^{D_2+1}\big)^{-1}\Big) = o_p(n^{-1/2})$;

if $i=l\neq t$, $\quad \frac{1}{n^3}\sum_{\substack{i=1\\ i\neq t}}^{n}\sum_{t=1}^{n}\psi_{niti} = \frac{1}{n^3}\sum_{\substack{i=1\\ i\neq t}}^{n}\sum_{t=1}^{n}\frac{\eta_i X_{2i,j}^* v_{g1t} D_d K_{2ti} K_1(0)}{h_1^{D_1}h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{id} = O_p\Big(\big(nh_1^{D_1}h_2\big)^{-1}\Big) = o_p(n^{-1/2})$;

if $i\neq t=l$, $\quad \frac{1}{n^3}\sum_{\substack{i=1\\ i\neq t}}^{n}\sum_{t=1}^{n}\psi_{nitt} = \frac{1}{n^3}\sum_{\substack{i=1\\ i\neq t}}^{n}\sum_{t=1}^{n}\frac{\eta_i X_{2i,j}^* v_{g1t} D_d K_{2ti} K_{1ti}}{h_1^{D_1}h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{td} = O_p\Big(\big(nh_2\big)^{-1}\Big) = o_p(n^{-1/2})$.

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

4.2.2. $T_{2d} = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{\eta_i X_{2i,j}^* v_{g1t} D_d K_{2ti} K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)}\big(\Pi_d(Z_l) - \Pi_d(Z_i)\big) \equiv \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\psi_{nitl}.$

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

- $\theta_n, \mathrm{E}(\psi_{nitl}|P_i), \mathrm{E}(\psi_{nitl}|P_l), \mathrm{E}(\psi_{nitl}|P_i,P_l) = 0$, as $\mathrm{E}(v_{g1t}|U_t) = 0$;
- $\mathrm{E}(\psi_{nitl}|P_t) = \frac{v_{g1t}}{h_1^{D_1}h_2^{D_2+1}}\mathrm{E}\left(\frac{\eta_i X_{2i,j}^* D_d K_{2ti} K_{1li}}{f_U(U_i)f_Z(Z_i)}\big(\Pi_d(Z_l) - \Pi_d(Z_i)\big)\Big|U_t\right) \leq \frac{Ch_1^{s_1}|v_{g1t}|}{h_2}$,
  $\mathrm{E}(\psi_{nitl}|P_i,P_t) = \frac{\eta_i X_{2i,j}^* v_{g1t} D_d K_{2ti}}{h_1^{D_1}h_2^{D_2+1} f_U(U_i) f_Z(Z_i)}\mathrm{E}\Big(K_{1li}\big(\Pi_d(Z_l) - \Pi_d(Z_i)\big)\Big|Z_i\Big) \leq \frac{Ch_1^{s_1}\left|\eta_i X_{2i,j}^* v_{g1t} D_d K_{2ti}\right|}{h_2^{D_2+1} f_U(U_i) f_Z(Z_i)}$,
  $\mathrm{E}(\psi_{nitl}|P_t,P_l) = \frac{v_{g1t}}{h_1^{D_1}h_2^{D_2+1}}\mathrm{E}\Big(\frac{\eta_i X_{2i,j}^* D_d K_{2ti} K_{1li}}{f_U(U_i)f_Z(Z_i)}\big(\Pi_d(Z_l) - \Pi_d(Z_i)\big)\Big|U_t, Z_l\Big) \leq \frac{Ch_1|v_{g1t}|}{h_2}$;
- $\sigma_{1n}^2 \leq \mathrm{E}(\phi_{1n}^2) = O\big(h_1^{2s_1}/h_2^2\big)$, $\quad \sigma_{2n}^2 \leq C\mathrm{E}\left(\mathrm{E}^2(\psi_{nitl}|P_i,P_t) + \mathrm{E}^2(\psi_{nitl}|P_t,P_l)\right) = O\big(h_1^{2s_1}/h_2^{D_2+2} + h_1^2/h_2^2\big)$,
  $\sigma_{3n}^2 = \mathrm{V}(\phi_{nitl}) \leq C\mathrm{E}(\psi_{nitl}^2) = O_p\big((h_1^{D_1-2}h_2^{D_2+2})^{-1}\big)$;
- $H_n^{(1)} = O_p\big((\sigma_{1n}^2/n^2)^{1/2}\big) = O_p\big(n^{-1/2}h_1^{s_1}h_2^{-1}\big) = o_p(n^{-1/2})$,

$$H_n^{(2)} = O_p\big((\sigma_{2n}^2/n^2)^{1/2}\big) = O_p\big(n^{-1/2}(h_1^{s_1}(nh_2^{D_2+2})^{-1/2} + h_1(nh_2^2)^{-1/2})\big) = o_p(n^{-1/2}),$$
$$H_n^{(3)} = O_p\big((\sigma_{3n}^2/n^3)^{1/2}\big) = O_p\big((n^3h_1^{D_1-2}h_2^{D_2+2})^{-1/2}\big) = o_p(n^{-1/2}).$$

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

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

$$\text{if}\quad i=t=l,\quad i=l\neq t,\qquad \psi_{nitl}=0;$$
$$\text{if}\quad i=t\neq l,\quad \frac{1}{n^3}\sum_{\substack{i=1\\ i\neq l}}^{n}\sum_{l=1}^{n}\psi_{niil} = \frac{1}{n^3}\sum_{\substack{i=1\\ i\neq l}}^{n}\sum_{l=1}^{n}\frac{\eta_i X_{2i,j}^* v_{g1t} D_d K_2(0) K_{1li}}{h_1^{D_1}h_2^{D_2+1} f_U(U_i) f_Z(Z_i)}\big(\Pi_d(Z_l)-\Pi_d(Z_i)\big) = O_p\Big(h_1\big(nh_2^{D_2+1}\big)^{-1}\Big);$$
$$\text{if}\quad i\neq t=l,\quad \frac{1}{n^3}\sum_{\substack{i=1\\ i\neq t}}^{n}\sum_{t=1}^{n}\psi_{nitt} = \frac{1}{n^3}\sum_{\substack{i=1\\ i\neq t}}^{n}\sum_{t=1}^{n}\frac{\eta_i X_{2i,j}^* v_{g1t} D_d K_{2ti} K_{1ti}}{h_1^{D_1}h_2^{D_2+1} f_U(U_i) f_Z(Z_i)}\big(\Pi_d(Z_t)-\Pi_d(Z_i)\big) = O_p\Big(h_1\big(nh_2\big)^{-1}\Big).$$

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

4.3.
$$\begin{aligned} B_{3123} &= \sum_{d=1}^{D_2}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i X_{2i,j}^*\big(g_1(U_t)-g_1(U_i)\big)D_d K_{2ti}}{h_2^{D_2+1}f_U(U_i)}(\hat U_{id}-U_{id}) \\ &= -\sum_{d=1}^{D_2}\Bigg\{\frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{\eta_i X_{2i,j}^*\big(g_1(U_t)-g_1(U_i)\big)D_d K_{2ti}K_{1li}}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}\Big(U_{ld}+\big(\Pi_d(Z_l)-\Pi_d(Z_i)\big)\Big) \\ &\quad + O_p(L_{1n}^2)\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\left|\frac{\eta_i X_{2i,j}^*\big(g_1(U_t)-g_1(U_i)\big)D_d K_{2ti}}{h_2^{D_2+1}f_U(U_i)}\right|\Bigg\} \equiv -\sum_{d=1}^{D_2}(W_{1d}+W_{2d})+o_p(n^{-1/2}). \end{aligned}$$

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

4.3.1.
$$W_{1d} = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{\eta_i X_{2i,j}^*\big(g_1(U_t)-g_1(U_i)\big)D_d K_{2ti}K_{1li}}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}U_{ld} \equiv \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\psi_{nitl}$$

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

- $\theta_n, \mathrm{E}(\phi_{nitl}|P_i), \mathrm{E}(\psi_{nitl}|P_t), \mathrm{E}(\psi_{nitl}|P_i,P_t) = 0,\quad \text{as}\quad \mathrm{E}(U_{ld}|Z_l)=0;$
- $\mathrm{E}(\psi_{nitl}|P_l) = \frac{U_{ld}}{h_1^{D_1}h_2^{D_2+1}}\mathrm{E}\Big(\frac{\eta_i X_{2i,j}^* D_d K_{2ti}K_{1li}}{f_U(U_i)f_Z(Z_i)}\big(g_1(U_t)-g_1(U_i)\big)\Big|Z_l\Big) \le C|U_{ld}|,$

  $\mathrm{E}(\psi_{nitl}|P_i,P_l) = \frac{\eta_i X_{2i,j}^* K_{1li}U_{ld}}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}\mathrm{E}\Big(D_d K_{2ti}\big(g_1(U_t)-g_1(U_i)\big)\Big|U_i\Big) \le \frac{C\big|\eta_i X_{2i,j}^* K_{1li}U_{ld}\big|}{h_1^{D_1}f_U(U_i)f_Z(Z_i)},$

  $\mathrm{E}(\psi_{nitl}|P_t,P_l) = \frac{U_{ld}}{h_1^{D_1}h_2^{D_2+1}}\mathrm{E}\Big(\frac{\eta_i X_{2i,j}^* D_d K_{2ti}K_{1li}}{f_U(U_i)f_Z(Z_i)}\big(g_1(U_t)-g_1(U_i)\big)\Big|U_t,Z_l\Big) \le C|U_{ld}|;$
- $\sigma_{1n}^2 \le \mathrm{E}(\phi_{1n}^2) = O(1),\quad \sigma_{2n}^2 \le C\mathrm{E}\big(\mathrm{E}^2(\psi_{nitl}|P_i,P_l)+\mathrm{E}^2(\psi_{nitl}|P_t,P_l)\big) = O\big(h_1^{-D_1}\big),$

  $\sigma_{3n}^2 = \mathrm{V}(\phi_{nitl}) \le C\mathrm{E}(\psi_{nitl}^2) = O_p\big((h_1^{D_1}h_2^{D_2})^{-1}\big);$
- $H_n^{(1)} = O_p(n^{-1/2}),\quad H_n^{(2)} = O_p\big((\sigma_{2n}^2/n^2)^{1/2}\big) = O_p\Big(\big(n^2h_1^{D_1}\big)^{-1/2}\Big) = o_p(n^{-1/2}),$

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

We have $U_n = 3H_n^{(1)} + o_p(n^{-1/2})$, where $H_n^{(1)} = \frac{1}{n}\sum_{l=1}^{n}\mathrm{E}(\psi_{nitl}|P_l)$. In this case, we need to investigate the structure of $H_n^{(1)}$ further. Note that $g_1(U_t)-g_1(U_i) = \mathbf{J}g_1(U_i)(U_t-U_i)+\frac{1}{2}(U_t-U_i)'\mathbf{H}g_1(U_{ti})(U_t-U_i)$, where $U_{ti} = \lambda U_i + (1-\lambda)U_t$, for $\lambda\in(0,1)$. Plugging this into $\mathrm{E}(\psi_{nitl}|P_l)$, we have

$$\begin{aligned}
3H_n^{(1)} &= \frac{3}{n}\sum_{l=1}^{n}\mathrm{E}(\psi_{nitl}|P_l) \equiv \frac{1}{n}\sum_{l=1}^{n}a_{1nl,jd} + \frac{1}{n}\sum_{l=1}^{n}b_{1nl,jd},\\
\text{where}\quad a_{1nl,jd} &= \frac{3U_{ld}}{h_1^{D_1}h_2^{D_2+1}}\mathrm{E}\left(\frac{\eta_i X_{2i,j}^* D_d K_{2ti}K_{1li}}{f_U(U_i)f_Z(Z_i)}\mathbf{J}g_1(U_i)(U_t-U_i)\Big|Z_l\right),\\
b_{1nl,jd} &= \frac{3U_{ld}}{h_1^{D_1}h_2^{D_2+1}}\mathrm{E}\left(\frac{\eta_i X_{2i,j}^* D_d K_{2ti}K_{1li}}{f_U(U_i)f_Z(Z_i)}\frac{1}{2}(U_t-U_i)'\mathbf{H}g_1(U_{ti})(U_t-U_i)\Big|Z_l\right).
\end{aligned}$$

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

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

$$\begin{aligned}
&\text{if}\quad i=t=l,\quad i=t\neq l,\quad \psi_{nitl}=0;\\
&\text{if}\quad i=l\neq t,\quad \frac{1}{n^3}\sum_{\substack{i=1\\ i\neq t}}^{n}\sum_{t=1}^{n}\psi_{niti} = \frac{1}{n^3}\sum_{\substack{i=1\\ i\neq t}}^{n}\sum_{t=1}^{n}\frac{\eta_i X_{2i,j}^* D_d K_{2ti}K_1(0)}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}\big(g_1(U_t)-g_1(U_i)\big)U_{id} = O_p\Big(\big(nh_1^{D_1}\big)^{-1}\Big);\\
&\text{if}\quad i\neq t=l,\quad \frac{1}{n^3}\sum_{\substack{i=1\\ i\neq t}}^{n}\sum_{t=1}^{n}\psi_{nitt} = \frac{1}{n^3}\sum_{\substack{i=1\\ i\neq t}}^{n}\sum_{t=1}^{n}\frac{\eta_i X_{2i,j}^* D_d K_{2ti}K_{1ti}}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}\big(g_1(U_t)-g_1(U_i)\big)U_{td} = O_p(n^{-1}).
\end{aligned}$$

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

$$\begin{aligned}
\sum_{d=1}^{D_2}W_{1d} &= \frac{6}{n^3}\binom{n}{3}\frac{1}{n}\sum_{i=1}^{n}\sum_{d=1}^{D_2}\frac{U_{id}}{2h_1^{D_1}h_2^{D_2}}\mathrm{E}\left(\frac{\eta_l X_{2l,j}^* D_d K_{2tl}K_{1il}}{f_U(U_l)f_Z(Z_l)}\mathbf{J}g_1(U_l)\Big(\frac{U_t-U_l}{h_2}\Big)\Big|Z_i\right) + o_p(n^{-1/2})\\
&\equiv \frac{6}{n^3}\binom{n}{3}\frac{1}{n}\sum_{i=1}^{n}a_{1ni,j} + o_p(n^{-1/2})\\
&= \frac{1}{n}\sum_{i=1}^{n}a_{1ni,j} + \left(\frac{6}{n^3}\binom{n}{3} - 1\right)\frac{1}{n}\sum_{i=1}^{n}a_{1ni,j} + o_p(n^{-1/2})\\
&= \frac{1}{n}\sum_{i=1}^{n}a_{1ni,j} + o_p(n^{-1/2}),
\end{aligned}$$

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

In sum, we have $\sum_{d=1}^{D_2}W_{1d} = \frac{1}{n}\sum_{i=1}^{n}a_{1ni,j} + o_p(n^{-1/2})$.

4.3.2. $W_{2d} = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{\eta_i X_{2i,j}^* D_d K_{2ti}K_{1li}}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}\big(g_1(U_t)-g_1(U_i)\big)\big(\Pi_d(Z_l)-\Pi_d(Z_i)\big) \equiv \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\psi_{nitl}$.

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

- $\theta_n = O(h_1^{s_1}) = o_p(n^{-1/2})$;
- $\mathrm{E}(\psi_{nitl}|P_i) = \frac{\eta_i X_{2i,j}^*}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}\mathrm{E}\big(D_d K_{2ti}K_{1li}(g_1(U_t)-g_1(U_i))(\Pi_d(Z_l)-\Pi_d(Z_i))\big|Z_i,U_i\big) \le \frac{Ch_1^{s_1}\left|\eta_i X_{2i,j}^*\right|}{f_U(U_i)f_Z(Z_i)}$,

$\mathrm{E}(\psi_{nitl}|P_t), \mathrm{E}(\psi_{nitl}|P_l), \mathrm{E}(\psi_{nitl}|P_t, P_l) = O(h_1)$,

$\mathrm{E}(\psi_{nitl}|P_i, P_t) = \frac{\eta_i X^*_{2i,j} D_d K_{2ti}(g_1(U_t)-g_1(U_i))}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} \mathrm{E}\big(K_{1li}\big(\Pi_d(Z_l) - \Pi_d(Z_i)\big)\big|Z_i\big) \leq \frac{Ch_1^{s_1}\left|\eta_i X^*_{2i,j} D_d K_{2ti}(g_1(U_t)-g_1(U_i))\right|}{h_2^{D_2+1} f_U(U_i) f_Z(Z_i)}$,

$\mathrm{E}(\psi_{nitl}|P_i, P_l) = \frac{\eta_i X^*_{2i,j} K_{1li}(\Pi_d(Z_l)-\Pi_d(Z_i))}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} \mathrm{E}\big(D_d K_{2ti}\big(g_1(U_t) - g_1(U_i)\big)\big|U_i\big) \leq \frac{C\left|\eta_i X^*_{2i,j} K_{1li}(\Pi_d(Z_l)-\Pi_d(Z_i))\right|}{h_1^{D_1} f_U(U_i) f_Z(Z_i)}$;

- $\sigma_{1n}^2 \leq C\mathrm{E}\big(\mathrm{E}^2(\psi_{nitl}|P_i) + \mathrm{E}^2(\psi_{nitl}|P_l) + \mathrm{E}^2(\psi_{nitl}|P_t)\big) = O(h_1^2)$,

  $\sigma_{2n}^2 = O\big(h_1^2 + h_1^{2s_1} h_2^{-D_2} + h_1^{2-D_1}\big) = O\big(h_1^{2s_1} h_2^{-D_2} + h_1^{2-D_1}\big)$, $\sigma_{3n}^2 = O\big((h_1^{D_1-2} h_2^{D_2})^{-1}\big)$;

- $H_n^{(1)} = O_p\big((\sigma_{1n}^2/n)^{1/2}\big) = O\left(h_1 n^{-1/2}\right) = o_p(n^{-1/2})$,

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

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

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

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

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

$$\text{if} \quad i \neq t = l, \quad \frac{1}{n^3}\sum_{\substack{i=1\\i\neq t}}^n \sum_{t=1}^n \psi_{nitt} = \frac{1}{n^3}\sum_{\substack{i=1\\i\neq t}}^n \sum_{t=1}^n \frac{\eta_i X^*_{2i,j} D_d K_{2ti} K_{1ti}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} \big(g_1(U_t) - g_1(U_i)\big)\big(\Pi_d(Z_l) - \Pi_d(Z_i)\big)$$

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

We have $B_{312} = -\frac{1}{n}\sum_{i=1}^n a_{1ni,j} + o_p(n^{-1/2})$. For B_{313}, the analysis is exactly similar to B_{312}, but note that for the term having order $O_p(n^{-1/2})$ in B_{3123}, the corresponding term in B_{3133}, denoted as W'_{1d}, is

$$W'_{1d} = \frac{1}{n^3}\sum_{i=1}^n\sum_{t=1}^n\sum_{l=1}^n \frac{\eta_i X^*_{2i,j}\big(g_1(U_t) - g_1(U_i)\big) D_d K_{2ti} K_{1lt}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_t)} U_{ld}.$$

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

As to B_{32}, the analysis is similar to B_{31} given above. For the component with order $O_p(n^{-1/2})$, we can actually combine that in B_{31} and the one in B_{32} together to have a more intuitive result. Note that

$$V_{g1i} - V_{g2i}\beta = \left\{\frac{1}{nh_2^{D_2} f_U(U_i)}\sum_{t=1}^n \hat{K}_{2ti}\Big[(\hat{\eta} - \eta_t)(Y_t - X_{2t}\beta) + (v_{g1t} - v_{g2t}\beta) + \Big(\big(g_1(U_t) - g_1(U_i)\big) - \big(g_2(U_t) - g_2(U_i)\big)\beta\Big)\Big]\right\}\Big(1 + O_p(L_{2n})\Big),$$

and the component of order $O_p(n^{-1/2})$ involves the third term in brackets, which is $\big(g_1(U_t) - g_2(U_i)\beta - \beta_0\big) - \big(g_1(U_i)) - g_2(U_i)\beta - \beta_0\big) = g(U_t) - g(U_i)$. Thus using $\big(g(U_t) - g(U_i)\big)$ instead of $\big(g_1(U_t) - g_1(U_i)\big)$ in W_{1d}, we have

$$B_3 = \frac{1}{n}\sum_{i=1}^n a_{ni} + o_p(n^{-1/2}),$$

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

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

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

$$\begin{aligned}
a_{ni,j} &= \sum_{d=1}^{D_2}\frac{U_{id}}{h_1^{D_1}h_2^{D_2}}\mathrm{E}\left(\frac{\eta_l X_{2l,j}^{*}D_dK_{2tl}K_{1il}}{f_U(U_l)f_Z(Z_l)}\mathbf{J}g(U_l)\Big(\frac{U_t-U_l}{h_2}\Big)\middle|Z_i\right)\\
&= \int\frac{1}{h_1^{D_1}h_2^{D_2}}\Big(\mathit{\Pi}_{2j}(Z_l)+U_{2l,j}-m_{2j}(W_l)-g_{2j}(U_l)+\mu_{2j}\Big)\sum_{d=1}^{D_2}U_{id}D_dK_{2tl}K_{1il}\mathbf{J}g(U_l)\left(\frac{U_t-U_l}{h_2}\right)\\
&\quad\times\frac{\eta_l(W_l,U_l)}{f_U(U_l)f_Z(Z_l)}f_U(U_t)f_{ZUM}(Z_l,U_l,W_l)\mathrm{d}U_t\mathrm{d}Z_l\mathrm{d}U_l\mathrm{d}W_l\\
&= \int\Big(\mathit{\Pi}_{2j}(Z_i-h_1\gamma)+U_{2t,j}-h_2\psi_{2j}-m_{2j}(W_l)-g_{2j}(U_t-h_2\psi)+\mu_{2j}\Big)\sum_{d=1}^{D_2}U_{id}D_dK_2(\psi)K_1(\gamma)\\
&\quad\times\mathbf{J}h(U_t-h_2\psi)\psi\frac{\eta_l(W_l,U_t-h_2\psi)}{f_U(U_t-h_2\psi)f_Z(Z_i-h_1\gamma)}f_U(U_t)f_{ZUM}(Z_i-h_1\gamma,U_t-h_2\psi,W_l)\mathrm{d}\gamma\mathrm{d}\psi\mathrm{d}U_t\mathrm{d}W_l\\
&\to \int\Big(\mathit{\Pi}_{2j}(Z_i)+U_{2t,j}-m_{2j}(W_l)-g_{2j}(U_t)+\mu_{2j}\Big)\sum_{d=1}^{D_2}U_{id}\left(-D_dg(U_t)\right)\eta_l(W_l,U_t)f_{UM|Z}(U_t,W_l|Z_i)\mathrm{d}U_t\mathrm{d}W_l\\
&= -\sum_{d=1}^{D_2}\mathrm{E}\left(\Big(\mathit{\Pi}_{2j}(Z_i)+U_{2t,j}-m_{2j}(W_t)-g_{2j}(U_t)+\mu_{2j}\Big)D_dg(U_t)\eta_t\middle|Z_i\right)U_{id}\\
&= -\sum_{d=1}^{D_2}\mathrm{E}\left(\Big(\mathit{\Pi}_{2j}(Z_i)-\mathit{\Pi}_{2j}(Z_t)\Big)D_dg(U_t)\eta_t\middle|Z_i\right)U_{id}.
\end{aligned}$$

The convergence follows by A3, and that $\int D_dK_2(\psi)\psi d\psi=(0,\cdots,-1,\cdots,0)'$, where -1 appears on the d^{th} position of the vector. The last equation follows by $\mathrm{E}(\eta_t X_{2t}^{*}|U_t)=0$. Hence, the $(j,k)^{\text{th}}$ element of $\mathrm{E}(a_{ni}a_{ni}')$ converges to

$$\Phi_{2_{(j,k)}}\equiv\mathrm{E}\left[\sum_{d=1}^{D_2}\sum_{\delta=1}^{D_2}\mathrm{E}\left(\Big(\mathit{\Pi}_{2j}(Z_i)-\mathit{\Pi}_{2j}(Z_t)\Big)D_dg(U_t)\eta_t\middle|Z_i\right)\mathrm{E}\left(\Big(\mathit{\Pi}_{2k}(Z_i)-\mathit{\Pi}_{2k}(Z_t)\Big)D_\delta g(U_t)\eta_t\middle|Z_i\right)U_{id}U_{i\delta}\right].$$

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

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

where $\quad\mathrm{E}|a_{ni,j}|^{2+\delta}\to\int\left|\sum_{d=1}^{D_2}\mathrm{E}\Big(\big(\mathit{\Pi}_{2j}(Z_i)+U_{2t,j}-m_{2j}(W_t)-g_{2j}(U_t)+\mu_{2j}\big)D_dg(U_t)\eta_t\Big|Z_i\Big)\right|^{2+\delta}$

$$\times |U_{id}|^{2+\delta} f_{ZU}(Z_i,U_i)\mathrm{d}Z_i\mathrm{d}U_i$$

$$\leq C\sum_{d=1}^{D_2}\int \Big|\mathrm{E}\Big(\big(\Pi_{2j}(Z_i)+U_{2t,j}-m_{2j}(W_t)-g_{2j}(U_t)+\mu_{2j}\big)\Big|Z_i\Big)\Big|^{2+\delta}|U_{id}|^{2+\delta} f_{ZU}(Z_i,U_i)\mathrm{d}Z_i\mathrm{d}U_i$$

$$<\infty \quad \text{since} \quad \mathrm{E}\big(|U_{id}|^{2+\delta}|Z_i\big)<C<\infty \quad \text{and} \quad \mathrm{E}|X_{2i,j}|^{2+\delta}<\infty.$$

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

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

□

Theorem 4 *Proof.* By Equation (6) and (11), we have

$$\hat{m}(w)-m(w)=\big(\hat{m}_1(w)-m_1(w)\big)-\big(\hat{m}_2(w)-m_2(w)\big)'\beta-\big(\hat{m}_3(w)-1\big)\beta_0$$
$$-\big(\hat{m}_2(w)-m_2(w)\big)'\big(\hat{\beta}-\beta\big)-\big(\hat{m}_3(w)-1\big)\big(\hat{\beta}_0-\beta_0\big)-m_2(w)'\big(\hat{\beta}-\beta\big)-\big(\hat{\beta}_0-\beta_0\big).$$

Since, by Theorems 2 and 3, $\hat{\beta}_0-\beta_0=O_p(n^{-1/2})$, $\hat{\beta}-\beta=O_p(n^{-1/2})$, and $\hat{m}_2(w)-m_2(w)=o_p(1)$, the last four terms in $\hat{m}(w)-m(w)$ when multiplied by $(nh_3^{D_3})^{1/2}$ are $o_p(1)$. Thus,

$$\sqrt{nh_3^{D_3}}\Big(\hat{m}(w)-m(w)\Big)=\sqrt{nh_3^{D_3}}\Big((\hat{m}_1(w)-m_1(w))-\big(\hat{m}_2(w)-m_2(w)\big)'\beta-\big(\hat{m}_3(w)-1\big)\beta_0\Big)+o_p(1).$$

We first investigate $\sqrt{nh_3^{D_3}}\big(\hat{m}_1(w)-m_1(w)\big)$, and then the asymptotic distribution of $\hat{m}(w)$ follows immediately due to the similar structure of $\hat{m}(w)$ and $\hat{m}_1(w)$. Given the expressions for $\hat{m}_1(w)$ and $\hat{f}_W(w)$, and the uniform order of $\hat{f}_W(w)$, letting $K_{3t,w}\equiv K_3\left(\frac{W_t-w}{h_3}\right)$, we have

$$\hat{m}_1(w)-m_1(w)=\left\{\frac{1}{nh_3^{D_3}f_W(w)}\sum_{t=1}^n K_{3t,w}\Big(\big(m_1(W_t)-m_1(w)\big)+\big(\eta_tY_t-m_1(W_t)\big)+(\hat{\eta}_t-\eta_t)Y_t\Big)\right\}\Big(1+O_p(L_{3n})\Big)$$
$$\equiv\left\{\sum_{k=1}^3 T_k\right\}\Big(1+O_p(L_{3n})\Big).$$

The proof has four steps:

(1) We show that $T_1=b_{m1,1}(w)$, where $b_{m1,1}(w)\equiv h_3^{s_3}\frac{\mu_{k_3,s_3}}{f_W(w)}\sum_{k=1}^{s_3}\frac{1}{k!(s_3-k)!}\sum_{j=1}^{D_3}D_j^k m_1(w)D_j^{s_3-k}f_W(w)+o_p(h_3^{s_3})$.

(2) We show that $\sqrt{nh_3^{D_3}}T_2\xrightarrow{d}\mathscr{N}(0,\Phi_{m1,1})$, where $\Phi_{m1,1}\equiv\frac{\sigma^2_{vm1}}{f_W(w)}\int K_3^2(\gamma)\mathrm{d}\gamma$.

(3) We show that $\sqrt{nh_3^{D_3}}\big(T_3-b_{m1,2}(w)\big)\xrightarrow{d}\mathscr{N}(0,\Phi_{m1,2})$, where $b_{m1,2}(w)\equiv h_3^{s_3}\frac{\mu_{k_3,s_3}}{f_W(w)}\frac{1}{s_3!}\sum_{j=1}^{D_3}m_1(w)D_j^{s_3}f_W(w)+$

$o_p(h_3^{s_3})$, $\Phi_{m1,2} \equiv m_1^2(w) f_W(w) \int \left(\int K_3(\gamma_1) K_3(\gamma_1+\gamma_2) \mathrm{d}\gamma_1 \right)^2 \mathrm{d}\gamma_2$.

(4) Combining (1)-(3), we show that $\sqrt{nh_3^{D_3}} \left(\hat{m}_1(w) - m_1(w) - b_{m1}(w) \right) \xrightarrow{d} \mathcal{N}(0, \Phi_{m1,1} + \Phi_{m1,2})$, where $b_{m1}(w) = b_{m1,1}(w) + b_{m1,2}(w)$.

Step 1: By Taylor's Theorem, we have

$$
\begin{aligned}
T_1 &= \frac{1}{nh_3^{D_3} f_W(w)} \sum_{t=1}^{n} K_{3t,w} \left(m_1(W_t) - m_1(w) \right) \\
&= \frac{1}{nh_3^{D_3} f_W(w)} \sum_{t=1}^{n} K_{3t,w} \left(\sum_{|\beta|=1}^{s_3} \frac{1}{|\beta|!} D^\beta m_1(w)(W_t - w)^\beta + \sum_{|\beta|=s_3+1} \frac{1}{(s_3+1)!} D^\beta m_1(\tilde{w})(W_t - w)^\beta \right) \equiv \sum_{k=1}^{s_3+1} T_{3k},
\end{aligned}
$$

where $\tilde{w} \equiv w + \lambda(W_t - w)$, for some $\lambda \in (0,1)$. For each $k = 1, \cdots, s_3$, we rewrite T_{1k} as

$$
T_{1k} = \frac{h_3^k}{k! f_W(w)} \sum_{|\beta|=k} D^\beta m_1(w) t_{k\beta}, \quad \text{where} \quad t_{k\beta} \equiv \frac{1}{nh_3^{D_3}} \sum_{t=1}^{n} K_{3t,w} \left(\frac{W_t - w}{h_3} \right)^\beta.
$$

By Lemma 3, $\sup_{w \in \mathscr{G}_W} |t_{k\beta} - \mathrm{E}(t_{k\beta})| = O_p\left(\left(\log n / nh_3^{D_3} \right)^{1/2} \right)$. If $k = 1$, for any $|\beta| = 1$, by Taylor's Theorem and given that k_3 is of order s_3, we have

$$
\begin{aligned}
\mathrm{E}(t_{1\beta}) &= \int K_3(\gamma) \gamma^\beta f_W(w + h_3\gamma) \mathrm{d}\gamma \\
&= \int K_3(\gamma) \gamma^\beta \left(f_W(w) + \sum_{|\alpha|=1}^{s_3-1} \frac{1}{|\alpha|!} D^\alpha f_W(w)(h_3\gamma)^\alpha + \sum_{|\alpha|=s_3} \frac{1}{s_3!} D^\alpha f_W(\tilde{w})(h_3\gamma)^\alpha \right) \mathrm{d}\gamma \\
&= h_3^{s_3-1} \frac{\mu_{k_3,s_3}}{(s_3-1)!} D^{(s_3-1)\beta} f_W(w) + o\left(h_3^{s_3-1} \right).
\end{aligned}
$$

Thus, given that $h_3 = n^{-1/(2s_3+D_3)}$, we have $h_3(\log n / nh_3^{D_3})^{1/2} = o(h_3^{s_3})$, and

$$
\begin{aligned}
T_{11} &= h_3^{s_3} \frac{\mu_{k_3,s_3}}{f_W(w)} \frac{1}{1!(s_3-1)!} \sum_{|\beta|=1} D^\beta m_1(w) D^{(s_3-1)\beta} f_W(w) + o_p(h_3^{D_3}) \\
&= h_3^{s_3} \frac{\mu_{k_3,s_3}}{f_W(w)} \frac{1}{1!(s_3-1)!} \sum_{j=1}^{D_3} D_j m_1(w) D_j^{s_3-1} f_W(w) + o_p(h_3^{D_3}).
\end{aligned}
$$

Similarly, if $k = 2$, for any β such that $|\beta| = 2$ and 2 is in the j^{th} position of the vector β, 0 elsewhere, we have $\mathrm{E}(t_{2\beta}) = h_3^{s_3-2} \frac{\mu_{k_3,s_3}}{(s_3-2)!} D_j^{s_3-2} f_W(w) + o\left(h_3^{s_3-2} \right)$. And for any remaining β such that $|\beta| = 2$, $\mathrm{E}(t_{2\beta}) = o\left(h_3^{s_3-2} \right)$. Thus, $T_{12} = h_3^{s_3} \frac{\mu_{k_3,s_3}}{f_W(w)} \frac{1}{2!(s_3-2)!} \sum_{j=1}^{D_3} D_j^2 m_1(w) D_j^{s_3-2} f_W(w) + o_p(h_3^{D_3})$. In a similar manner, we have,

$$
T_{1k} = h_3^{s_3} \frac{\mu_{k_3,s_3}}{f_W(w)} \frac{1}{k!(s_3-k)!} \sum_{j=1}^{D_3} D_j^k m_1(w) D_j^{s_3-k} f_W(w) + o_p(h_3^{D_3}), \quad \text{for any } k = 1, \cdots, s_3.
$$

For $k = s_3+1$, we have $T_{1(s_3+1)} = \frac{h_3^{s_3+1}}{nh_3^{D_3} f_W(w)} \sum_{t=1}^n K_{3t,w} \left(\sum_{|\beta|=s_3+1} \frac{1}{(s_3+1)!} D^\beta m_1(\tilde{w}) \left(\frac{W_t - w}{h_3}\right)^\beta \right) = o_p(h_3^{s_3})$, by Markov's Inequality and $\mathrm{E}|T_{1(s_3+1)}| = O(h_3^{s_3+1}) = o(h_3^{s_3})$ since $m_1(w) \in C^{s_3+1}$. Combining all the T_{3k} terms, we have

$$T_1 = b_{m1,1}(w), \quad \text{where } b_{m1,1}(w) \equiv h_3^{s_3} \frac{\mu_{k_3,s_3}}{f_W(w)} \sum_{k=1}^{s_3} \frac{1}{k!(s_3-k)!} \sum_{j=1}^{D_3} D_j^k m_1(w) D_j^{s_3-k} f_W(w) + o_p(h_3^{s_3}).$$

Step 2: Given $\eta_t Y_t = m_1(W_t) + v_{m1t}$, we have $T_2 = \sum_{t=1}^n a_{1tn}$, where $a_{1tn} \equiv (nh_3^{D_3} f_W(w))^{-1} K_{3t,w} v_{m1t}$. Since $\mathrm{E}(v_{m1t}|W_t) = 0$ and $\mathrm{E}(v_{m1t}^2|W_t) = \sigma_{vm1}^2 < \infty$, we have $\mathrm{E}(a_{1tn}) = 0$, and $\mathrm{V}(a_{1tn}) = n^{-2} h_3^{-D_3} f_W^{-2}(w) \sigma_{vm1}^2 \int K_3^2(\gamma) f_W(w + h_3\gamma) \mathrm{d}\gamma$. Let $S_{1n}^2 \equiv \sum_{t=1}^n \mathrm{V}(a_{1tn}) = (nh_3^{D_3})^{-1} f_W^{-2}(w) \sigma_{vm1}^2 \int K_3^2(\gamma) f_W(w+h_3\gamma)\mathrm{d}\gamma$. Then, by Lyapunov's CLT, if $\sum_{t=1}^n \mathrm{E}|a_{1tn}/S_{1n}|^{2+\delta} \to 0$ for some $\delta > 0$ as $n \to \infty$, we have $\sum_{t=1}^n a_{1tn}/S_{1n} \xrightarrow{d} \mathcal{N}(0,1)$, i.e., given $\sqrt{nh_3^{D_3}} S_{1n} \to \Phi_{m1,1}^{1/2}$,

$$\sqrt{nh_3^{D_3}} T_2 \xrightarrow{d} \mathcal{N}(0, \Phi_{m1,1}), \quad \text{where } \Phi_{m1,1} \equiv \frac{\sigma_{vm1}^2}{f_W(w)} \int K_3^2(\gamma)\mathrm{d}\gamma.$$

Given that $nh_3^{D_3} S_{1n}^2 \to \Phi_{m1,1} > 0$ and $\mathrm{E}(|v_{m1t}|^{2+\delta}|W_t) < C$, Lyapunov's condition is satisfied since

$$\sum_{t=1}^n \mathrm{E}\left|\frac{a_{1tn}}{S_{1n}}\right|^{2+\delta} = \frac{\left(nh_3^{D_3}\right)^{\delta/2+1}}{\left(nh_3^{D_3} S_{1n}^2\right)^{\delta/2+1}} \sum_{t=1}^n \mathrm{E}\left(\left| \frac{K_{3t,w} v_{m1t}}{nh_3^{D_3} f_W(w)} \right|^{2+\delta} \right) \le C\left(nh_3^{D_3}\right)^{-\delta/2} \int |K_3(\gamma)|^{2+\delta} \mathrm{d}\gamma \to 0, \text{ as } n \to \infty.$$

Step 3: Denote $\hat{f}_U(\hat{U}_t) = \hat{f}_{\hat{U}_t}$, $\hat{f}_W(W_t) = \hat{f}_{W_t}$, $\hat{\phi}(W_t, \hat{U}_t) = \hat{\phi}_t$, $f_U(U_t) = f_{U_t}$, $f_W(W_t) = f_{W_t}$, $\phi(W_t, U_t) = \phi_t$. According to the uniform order of these density estimators from Theorem 1 and $L_n^2, (L_{1n}/h_2)^2 = o(n^{-1/2})$ by A5, we have

$$\hat{\eta}_t - \eta_t = \frac{1}{\phi_t^2} \left(\phi_t f_{W_t} (\hat{f}_{\hat{U}_t} - f_{U_t}) - f_{U_t} f_{W_t} (\hat{\phi}_t - \phi_t) + \phi_t f_{U_t} (\hat{f}_{W_t} - f_{W_t}) \right) + o_p(n^{-1/2}).$$

Since $T_3 = (nh_3^{D_3} f_W(w))^{-1} \sum_{t=1}^n K_{3t,w}\left((\hat{\eta}_t - \eta_t) Y_t\right)$, and $(nh_3^{D_3} f_W(w))^{-1} \sum_{t=1}^n \left|K_{3t,w} Y_t\right| = O_p(1)$, we have

$$T_3 = \sum_{k=1}^3 T_{3k} + o_p\left(n^{-1/2}\right),$$

where
$$T_{31} = \frac{1}{nh_3^{D_3} f_W(w)} \sum_{t=1}^n \frac{1}{f_{U_t}} (\hat{f}_{\hat{U}_t} - f_{U_t}) K_{3t,w} \eta_t Y_t, \quad T_{32} = -\frac{1}{nh_3^{D_3} f_W(w)} \sum_{t=1}^n \frac{1}{\phi_t} (\hat{\phi}_t - \phi_t) K_{3t,w} \eta_t Y_t,$$
$$T_{33} = \frac{1}{nh_3^{D_3} f_W(w)} \sum_{t=1}^n \frac{1}{f_{W_t}} (\hat{f}_{W_t} - f_{W_t}) K_{3t,w} \eta_t Y_t.$$

From Theorem 1, we have $\left|\hat{f}_{\hat{U}_t} - f_{U_t}\right| = O_p(L_{2n})$ and $|\hat{\phi}_t - \phi_t| = O_p(L_{4n})$ uniformly. Thus, $\sqrt{nh_3^{D_3}} T_{31} = O_p\left(\sqrt{nh_3^{D_3}} L_{2n}\right) = o_p(1)$ by Assumption A5 (iii). Similarly, $\sqrt{nh_3^{D_3}} T_{32} = O_p\left(\sqrt{nh_3^{D_3}} L_{4n}\right) = o_p(1)$. Let $T_{33} =$

$T_{331}+T_{332}$, where

$$T_{331}=\frac{1}{nh_3^{D_3}f_W(w)}\sum_{t=1}^{n}\frac{1}{f_{W_t}}\big(\mathrm{E}(\hat f_{W_t})-f_{W_t}\big)K_{3t,w}\eta_tY_t,\qquad T_{332}=\frac{1}{nh_3^{D_3}f_W(w)}\sum_{t=1}^{n}\frac{1}{f_{W_t}}\big(\hat f_{W_t}-\mathrm{E}(\hat f_{W_t})\big)K_{3t,w}\eta_tY_t.$$

We show that T_{331} contributes to a bias and T_{332} to a normal distribution.

For T_{331}, given that $\mathrm{E}(\hat f_{W_t})-f_{W_t}=h_3^{s_3}\frac{\mu_{k_3,s_3}}{s_3!}\sum_{j=1}^{D_3}D_j^{s_3}f_W(W_t)+o(h_3^{s_3})$ by Taylor's Theorem and the high order of kernel k_3, we have

$$T_{331}=h_3^{s_3}\frac{\mu_{k_3,s_3}}{s_3!}\sum_{j=1}^{D_3}t_j+o(h_3^{s_3}),\qquad\text{where}\quad t_j=\frac{1}{nh_3^{D_3}f_W(w)}\sum_{t=1}^{n}\frac{1}{f_{W_t}}D_j^{s_3}f_W(W_t)K_{3t,w}\eta_tY_t.$$

Since $\eta_tY_t=v_{m1t}+(m_1(W_t)-m_1(w))+m_1(w)$, let $t_j=\sum_{k=1}^{3}t_{jk}$, where

$$t_{j1}=\frac{1}{nh_3^{D_3}f_W(w)}\sum_{t=1}^{n}\frac{1}{f_{W_t}}D_j^{s_3}f_W(W_t)K_{3t,w}v_{m1t},\qquad t_{j2}=\frac{1}{nh_3^{D_3}f_W(w)}\sum_{t=1}^{n}\frac{1}{f_{W_t}}D_j^{s_3}f_W(W_t)K_{3t,w}\big(m_1(W_t)-m_1(w)\big),$$
$$t_{j3}=\frac{m_1(w)}{nh_3^{D_3}f_W(w)}\sum_{t=1}^{n}\frac{1}{f_{W_t}}D_j^{s_3}f_W(W_t)K_{3t,w}.$$

By Markov's Inequality and $\mathrm{E}(t_{j1})=0$, $\mathrm{E}(t_{j1}^2)=O\big((nh_3^{D_3})^{-1}\big)$ due to $\mathrm{E}(v_{m1t}|W_t)=0$ and $\mathrm{E}(v_{m1t}^2|W_t)\le C$, we have $t_{j1}=O_p\big((nh_3^{D_3})^{-1/2}\big)=o_p(1)$. And $t_{j2}=O_p(h_3)=o_p(1)$ since $\mathrm{E}|t_{j2}|\le Ch_3^{-D_3}\mathrm{E}\big|K_{3t,w}\big(m_1(W_t)-m_1(w)\big)\big|=O(h_3)$. For t_{j3}, since $\mathrm{E}(t_{j3})=m_1(w)f_W(w)^{-1}\int D_j^{s_3}f_W(w+h_3\phi)K_3(\phi)\mathrm{d}\phi\to m_1(w)f_W^{-1}(w)D_j^{s_3}f_W(w)$, and $\mathrm{E}(t_{j3}^2)=O\big((nh_3^{D_3})^{-1}\big)=o(1)$, we have $t_{j3}=m_1(w)f_W^{-1}(w)D_j^{s_3}f_W(w)+o_p(1)$. In sum, $T_{331}=h_3^{s_3}\frac{\mu_{k_3,s_3}}{f_W(w)}\frac{1}{s_3!}\sum_{j=1}^{D_3}m_1(w)D_j^{s_3}f_W(w)+o_p(h_3^{s_3})\equiv b_{m1,2}(w)$.

For T_{332}, we show that $(nh_3^{D_3})^{1/2}T_{332}=(nh_3^{D_3})^{1/2}\sum_{t=1}^{n}a_{2tn}+o_p(1)\xrightarrow{d}\mathcal{N}(0,\Phi_{m1,2})$, where

$$a_{2tn}=(nh_3^{2D_3})^{-1}m_1(w)\mathrm{E}\Big(f_{W_i}^{-1}\big(K_{3it}-\mathrm{E}_t(K_{3it})\big)K_{3i,w}\Big|W_t\Big),\quad\Phi_{m1,2}\equiv m_1^2(w)f_W(w)\int\Big(\int K_3(\gamma_1)K_3(\gamma_1+\gamma_2)\mathrm{d}\gamma_1\Big)^2\mathrm{d}\gamma_2.$$

Since $\hat f_{W_t}-\mathrm{E}(\hat f_{W_t})=(nh_3^{D_3})^{-1}\sum_{i=1}^{n}\big(K_{3ti}-\mathrm{E}_i(K_{3ti})\big)$, we have $T_{332}=T_{3321}+T_{3322}$, where

$$T_{3321}=\frac{1}{n^2h_3^{2D_3}f_W(w)}\sum_{t=1}^{n}\frac{1}{f_{W_t}}\big(K_3(0)-\mathrm{E}_i(K_{3ti})\big)K_{3t,w}\eta_tY_t,\quad T_{3321}=\frac{1}{n^2h_3^{2D_3}f_W(w)}\sum_{\substack{t=1\\t\neq i}}^{n}\sum_{i=1}^{n}\frac{1}{f_{W_t}}\big(K_{3ti}-\mathrm{E}_i(K_{3ti})\big)K_{3t,w}\eta_tY_t.$$

Since $\mathrm{E}_i(K_{3ti})=O(h_3^{D_3})$, we have $T_{3321}\le C(nh_3^{D_3})^{-2}\sum_{t=1}^{n}|K_{3t,w}\eta_tY_t|=O_p\big((nh_3^{D_3})^{-1}\big)$, thus $(nh_3^{D_3})^{1/2}T_{3321}=o_p(1)$.

For T_{3322}, we have $T_{3322}=\frac{1}{n^2}\binom{n}{2}U_n=\frac{n-1}{n}\frac{1}{2}U_n$, where $U_n\equiv\binom{n}{2}^{-1}\sum_{\substack{t=1\\t<i}}^{n}\sum_{i=1}^{n}\phi_{nti}=\theta_n+2H_n^{(1)}+H_n^{(2)}$, $\phi_{nti}=\psi_{nti}+\psi_{nit}$, and $\psi_{nti}=(h_3^{2D_3}f_{W_t})^{-1}\big(K_{3ti}-\mathrm{E}_i(K_{3ti})\big)K_{3t,w}\eta_tY_t$. Then $\theta_n=\mathrm{E}(\phi_{nti})=0$, $\sigma_{2n}^2=\mathrm{V}(\phi_{nti})\le C\mathrm{E}(\psi_{nti}^2)=O(h_3^{-2D_3})$, $H_n^{(2)}=O_p\big((\sigma_{2n}^2/n^2)^{1/2}\big)=O_p\big(nh_3^{D_3})^{-1}\big)$, and we have $(nh_3^{D_3})^{1/2}H_n^{(2)}=o_p(1)$. For $H_n^{(1)}=n^{-1}\sum_{t=1}^{n}\mathrm{E}(\psi_{nit}|W_t)$, given

that $\mathrm{E}(\eta_t Y_i|W_i) = m_1(W_i)$, we have $H_n^{(1)} = Q_1 + Q_2$, where

$$Q_1 \equiv \frac{1}{nh_3^{2D_3}}\sum_{t=1}^{n}\mathrm{E}\Big(\frac{1}{f_{W_i}}\big(K_{3it}-\mathrm{E}_t(K_{3it})\big)K_{3i,w}\big(m_1(W_i)-m_1(w)\big)\Big|W_t\Big),\; Q_2 \equiv \frac{m_1(w)}{nh_3^{2D_3}}\sum_{t=1}^{n}\mathrm{E}\Big(\frac{1}{f_{W_i}}\big(K_{3it}-\mathrm{E}_t(K_{3it})\big)K_{3i,w}\Big|W_t\Big).$$

Since $\mathrm{E}(Q_1)=0$, $\mathrm{E}(Q_1^2)=O\big((nh_3^{D_3})^{-1}h_3\big)$, we have $(nh_3^{D_3})^{1/2}Q_1 = O_p(h_3^{1/2}) = o_p(1)$. Since $Q_2 = \sum_{t=1}^n a_{2tn}$, let $Z_{tn} = (nh_3^{2D_3})^{-1}m_1(w)\mathrm{E}_i\big(f_{W_i}^{-1}K_{3it}K_{3i,w}\big)$, and $\mu_n = (nh_3^{2D_3})^{-1}m_1(w)\mathrm{E}\big(f_{W_i}^{-1}K_{3it}K_{3i,w}\big)$, so that $a_{2tn} = Z_{tn}-\mu_n$ and $\mathrm{E}(Z_{tn}) = \mu_n$. Then, we have $Z_{tn} = (nh_3^{D_3})^{-1}m_1(w)\int K_3(\gamma_1)K_3\left(\frac{W_t-w}{h_3}+\gamma_1\right)\mathrm{d}\gamma_1$, $\mu_n = n^{-1}m_1(w)\int K_3(\gamma_1)K_3(\gamma_1+\gamma_2)\mathrm{d}\gamma_1\mathrm{d}\gamma_2 = O(n^{-1})$, and $\mathrm{V}(a_{2tn}) = \mathrm{E}(Z_{tn}^2)-\mu_n^2 = n^{-2}h_3^{-D_3}m_1^2(w)\int\big(\int K_3(\gamma_1)K_3(\gamma_1+\gamma_2)\mathrm{d}\gamma_1\big)^2 f_W(w+h_3\gamma_2)\mathrm{d}\gamma_2-\mu_n^2$. Letting $S_{2n}^2 \equiv \sum_{t=1}^n \mathrm{V}(a_{2tn})$, we have $nh_3^{D_3}S_{2n}^2 = m_1^2(w)\int\big(\int K_3(\gamma_1)K_3(\gamma_1+\gamma_2)\mathrm{d}\gamma_1\big)^2 f_W(w+h_3\gamma_2)\mathrm{d}\gamma_2 - n^2h_3^{D_3}\mu_n^2 \to \Phi_{m1,2}$. Thus, by Lyapunov's CLT, if $\sum_{t=1}^n \mathrm{E}|a_{2tn}/S_{2n}|^{2+\delta}\to 0$ for some $\delta>0$ as $n\to\infty$, we have $\sum_{t=1}^n a_{2tn}/S_{2n} \xrightarrow{d} \mathcal{N}(0,1)$, i.e., combing previous results on other terms in T_3,

$$\sqrt{nh_3^{D_3}}\big(T_3-b_{m1,2}(w)\big) \xrightarrow{d} \mathcal{N}(0,\Phi_{m1,2}), \quad \text{where } \Phi_{m1,2} \equiv m_1^2(w)f_W(w)\int\Big(\int K_3(\gamma_1)K_3(\gamma_1+\gamma_2)\mathrm{d}\gamma_1\Big)^2\mathrm{d}\gamma_2.$$

Given that $nh_3^{D_3}S_{2n}^2 \to \Phi_{m1,2} > 0$, Lyapunov's condition is satisfied since

$$\sum_{t=1}^n \mathrm{E}\left|\frac{a_{2tn}}{S_{2n}}\right|^{2+\delta} \le \frac{C\big(nh_3^{D_3}\big)^{\delta/2+1}}{\big(nh_3^{D_3}S_{2n}^2\big)^{\delta/2+1}}\sum_{t=1}^n \mathrm{E}(|Z_{tn}|^{2+\delta}) \le C\big(nh_3^{D_3}\big)^{-\delta/2}\to 0, \text{ as } n\to\infty.$$

Step 4: Combining results from (1) to (3), we have $\sqrt{nh_3^{D_3}}\big(\hat{m}_1(w)-m_1(w)-b_{m1}(w)\big) = \sqrt{nh_3^{D_3}}\sum_{t=1}^n(a_{1tn}+a_{2tn})$, where $b_{m1}(w) = b_{m1,1}(w)+b_{m1,2}(w) = h_3^{s_3}\frac{\mu_{k_3,s_3}}{f_W(w)}\sum_{k=0}^{s_3}\frac{1}{k!(s_3-k)!}\sum_{j=1}^{D_3}D_j^k m_1(w)D_j^{s_3-k}f_W(w)+o_p(h_3^{s_3})$. Reapplying Lyapunov's CLT, given that $S_n^2 \equiv \mathrm{V}\big(\sum_{t=1}^n(a_{1tn}+a_{2tn})\big) = S_{1n}^2+S_{2n}^2+2\sum_{t=1}^n \mathrm{Cov}(a_{1tn},a_{2tn}) = S_{1n}^2+S_{2n}^2$ as $\mathrm{E}(a_{1tn}a_{2tn})=0$, and $nh_3^{D_3}S_n^2 \to \Phi_{m1,1}+\Phi_{m1,2}$, we have $\sqrt{nh_3^{D_3}}\big(\hat{m}_1(w)-m_1(w)-b_{m1}(w)\big) \xrightarrow{d} \mathcal{N}(0,\Phi_{m1,1}+\Phi_{m1,2})$. Lyapunov's condition can be easily verified using C_r Inequality.

Next, we extend this result for $\hat{m}_1(w)$ to $\hat{m}(w)$. Recall that,

$$\hat{m}_1(w) = \frac{1}{nh_3^{D_3}f_W(w)}\sum_{t=1}^n K_{3t,w}\hat{\eta}_t Y_t, \qquad \hat{m}(w) = \frac{1}{nh_3^{D_3}f_W(w)}\sum_{t=1}^n K_{3t,w}\hat{\eta}_t\big(Y_t-X_{2t}'\beta-\beta_0\big).$$

We see that $\hat{m}(w)$ shares a similar structure as $\hat{m}_1(w)$ except using $\hat{\eta}_t\big(Y_t-X_{2t}'\beta-\beta_0\big)$ instead of $\hat{\eta}_t Y_t$ as the regressand. Given that $\hat{\eta}_t\big(Y_t-X_{2t}'\beta-\beta_0\big) = m(W_t)+v_{mt}$, $\mathrm{E}(v_{mt}^2|W_t)=\sigma_{vm}^2\le C$, and $\mathrm{E}(|v_{mt}|^{2+\delta}|W_t)\le C$, by repeating Step 1-4, we have $\sqrt{nh_3^{D_3}}\big(\hat{m}(w)-m(w)-b_m(w)\big) \xrightarrow{d} \mathcal{N}(0,\Phi_3+\Phi_4)$, where $b_m(w) = h_3^{s_3}\frac{\mu_{k_3,s_3}}{f_W(w)}\sum_{k=0}^{s_3}\frac{1}{k!(s_3-k)!}\sum_{j=1}^{D_3}D_j^k m(w)$ $\times D_j^{s_3-k}f_W(w)+o_p(h_3^{s_3})$, $\Phi_3 = \frac{\sigma_{vm}^2}{f_W(w)}\int K_3^2(\gamma)\mathrm{d}\gamma$, $\Phi_4 = m^2(w)f_W(w)\int(\int K_3(\gamma_1)K_3(\gamma_1+\gamma_2)\mathrm{d}\gamma_1)^2\mathrm{d}\gamma_2$.

□

Lemmas

We start by noting that for any kernel K that satisfies Assumption A1, and for any function $f(x) : \mathbb{R}^D \to \mathbb{R}$ such that $\int |f(\gamma)| \mathrm{d}\gamma < \infty$, we have that if x is a point of continuity of $f(x)$,

$$\int K(\gamma) f(x + h_n\gamma) \mathrm{d}\gamma \to f(x) \int K(\gamma) \mathrm{d}\gamma \quad \text{as} \quad n \to \infty.$$

This result follows directly from Theorem 1A in Parzen (1962).

Lemma 1. *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) $|k^{(1)}(x)||x|^3 \leq C$, for any $x \in \mathbb{R}$ and some $C > 0$. Thus, for any $|\beta| = 0, \cdots, 3$, $K(x)x^\beta$ satisfies a local Lipschitz condition, i.e., for any $x \neq y \in A$, where $A \subset \mathbb{R}^D$ is a bounded convex set, we have $|K(x)x^\beta - K(y)y^\beta| \leq C||x - y||_E$, for some $C > 0$.*

Proof. Note that by a)-c), for any $x \in \mathbb{R}$, we have $|k(x)||x|^i, |k^{(1)}(x)||x|^i \leq C, i = 0, \cdots, 3$.

(a) $|\beta| = 0$. Since by the Mean Value Theorem $K(x) - K(y) = \mathbf{J}K(x^*)(x - y)$, where $x^* = x + \lambda(y - x)$, $\lambda \in (0, 1)$, and $|D_i K(x^*)| = |k^{(1)}(x_i^*)| \prod_{p \neq i}^D |k(x_p^*)| \leq C$, we have $|K(x) - K(y)| \leq C \sum_{i=1}^D |x_i - y_i| \leq CD \left(\sum_{i=1}^D (x_i - y_i)^2\right)^{1/2} \leq C||x - y||_E$ for some $C > 0$ by the Triangle and C_r inequalities.

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

$$\begin{aligned}
|K(x)x_i - K(y)y_i| &= |x_i(K(x) - K(y)) + K(y)(x_i - y_i)| \\
&= |x_i \mathbf{J}K(x^*)(x - y) + K(y)(x_i - y_i)| \text{ by the Mean Value Theorem} \\
&= \left| (x_i D_i K(x^*) + K(y))(x_i - y_i) + \sum_{p \neq i} x_i D_p K(x^*)(x_p - y_p) \right| \\
&\leq C \sum_{i=1}^D |x_i - y_i| \leq C||x - y||_E \text{ by the Triangle and } C_r \text{ inequalities.}
\end{aligned}$$

The Mean value theorem is used in the second equality since $k(x)$ is continuously differentiable on the convex set A. And since set A is bounded, there exists a $C \geq 0$ such that $y_i - x_i = \Delta_i$ and $|\Delta_i| \leq C$. Thus $x_i^* \equiv x_i + \lambda(y_i - x_i) = x_i + \lambda\Delta$, and we have $|x_i k^{(1)}(x_i^*)| = |x_i k^{(1)}(x_i + \lambda\Delta)| \leq C$ by c).

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

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

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

$$
\begin{aligned}
\left|K(x)x_ix_jx_l - K(y)y_iy_jy_l\right| &= \left|x_l(K(x)x_ix_j - K(y)y_iy_j) + K(y)y_iy_j(x_l - y_l)\right| \\
&\le \left|x_ix_jx_lD_iK(x^*) + x_jx_lK(y)\right| |x_i - y_i| + \left|x_ix_jx_lD_jK(x^*) + x_lK(y)y_i\right| |x_j - y_j| \\
&+ \left|x_ix_jx_lD_lK(x^*) + K(y)y_iy_j\right| |x_l - y_l| + \sum_{p \ne i,j,l} \left|x_ix_jx_lD_pK(x^*)\right| |x_p - x_p| \le C\,||x - y||_E.
\end{aligned}
$$

□

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

$$
\sup_{x \in G_x} |S_n(x) - \mathrm{E}(S_n(x))| = O_p\Big(\Big(\frac{\log n}{n}\Big)^{1/2}\Big).
$$

Proof. Since G_x is a compact subset of $\mathbb{R}^K$, there exists $x_0 \in \mathbb{R}^K$ such that $G_x \subset B(x_0, r) = \{x \in \mathbb{R}^K : \|x - x_0\| < r\}$. Thus, for all $x, x' \in G_x$, $\|x - x'\| < 2r$. By the Heine-Borel Theorem, every infinite open cover of G_x contains a finite subcover which we construct as $\{B(x_k, n^{-1/2})\}_{k=1}^{l_n}$ with $x_k \in G_x$ and $l_n < n^{K/2}C$. For $x \in B(x_k, n^{-1/2})$, by condition a), we have

$$
|S_n(x) - S_n(x_k)| \quad \le \quad n^{-1/2}\frac{1}{n}\sum_{i=1}^n B_n(X_i) = O_p(n^{-1/2})
$$

since $\mathrm{E}(B_n(X_i)) < \infty$ and $\{X_i\}_{i=1}^n$ is and IID sequence. Similarly, $|\mathrm{E}(S_n(x)) - \mathrm{E}(S_n(x_k))| = O(n^{-1/2})$ and using the triangle inequality we have, $|S_n(x) - \mathrm{E}(S_n(x))| \le |S_n(x_k) - \mathrm{E}(S_n(x_k))| + O_p(n^{-1/2})$. 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 Δ_ε such that for $n \ge N$

$$
P\left(\left(\frac{n}{\log n}\right)^{1/2} \max_{1 \le k \le l_n} |S_n(x_k) - \mathrm{E}\left(S_n(x_k)\right)| \ge \Delta_\varepsilon\right) \le \varepsilon.
$$

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

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

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

$$
P\left(\left|\sum_{i=1}^n G_n(X_i,x_k) - \sum_{i=1}^n \mathrm{E}\left(G_n(X_i,x_k)\right)\right| \ge n\varepsilon_n\right) \le 2\exp\left(-\frac{n\varepsilon_n^2}{c_n}\right) = 2\exp\left(-\frac{\Delta_\varepsilon^2 \log n}{c_n}\right) = 2n^{-\frac{\Delta_\varepsilon^2}{c_n}}.
$$

Hence, $P\left(\max_{1 \le k \le l_n} |S_n(x_k) - \mathrm{E}\left(S_n(x_k)\right)| \ge \varepsilon_n\right) \le 2l_n n^{-\Delta_\varepsilon^2/c_n} < Cn^{K/2 - \Delta_\varepsilon^2/c_n}$. Since $\varepsilon_n \to 0$ as $n \to \infty$ and $\mathrm{V}(G_n(X_i,$

$x_k)) < \infty$, we can choose Δ_ε sufficiently large such that $K/2 - \Delta_\varepsilon^2/c_n < 0$ and

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

□

Lemma 3. *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) $|k(x)||x|^{7+c} \to 0$ as $|x| \to \infty$, for some $c > 0$; c) $|k^{(1)}(x)||x|^3 \to 0$ as $|x| \to \infty$. In addition, assume that 1) $\{(X_t, \varepsilon_t)'\}_{t=1,2,\cdots}$ is an independent and identically distributed sequence of random vectors; 2) The joint density of X_t and ε_t is given by $f_{X\varepsilon}(x, \varepsilon) = f_X(x) f_{\varepsilon|X}(\varepsilon|x)$; 3) $f_X(x)$ is continuous and uniformly bounded everywhere. Let $w(X_t - x; x) : \mathbb{R}^D \to \mathbb{R}$ and $g(\varepsilon) : \mathbb{R} \to \mathbb{R}$ be measurable functions. Define*

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

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

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

ii) *$w(X_t - x; x)$ satisfies a Lipschitz condition of order 1 in x, i.e., $|w(X_t - x; x) - w(X_t - x^k; x^k)| \leq C||x - x^k||_E$ for some $C > 0$, and $|w(X_t - x, x)| < C$ for all $x \in \mathbb{R}^D$.*

Then, for an arbitrary compact set $\mathscr{G} \subseteq \mathbb{R}^D$, we have $\sup_{x\in\mathscr{G}} |s(x) - \mathrm{E}(s(x))| = O_p\left(\left(\frac{\log n}{nh_n^D}\right)^{1/2}\right)$, provided that $h_n \to 0$, $nh_n^{D+2} \to \infty$ and $\frac{nh_n^D}{\log n} \to \infty$ as $n \to \infty$.

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

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

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

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

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

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

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

By the Borel-Cantelli Lemma $P\left(\liminf_{t\to\infty}\{|g(\varepsilon_t)| \le B_t\}\right) = 1$. Hence, there exists an N such that for all $t > N$ we have $P(|g(\varepsilon_t)| \le B_t) = 1$. Since $B_t \le B_n$ for $t \le n$ we have $P(|g(\varepsilon_t)| \le B_n) = 1$, and therefore $\chi_{\{|g(\varepsilon_t)|>B_n\}} = 0$ with probability 1, which gives $T_1 = 0$ almost surely when n is sufficiently large.

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

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

where the last inequality follows by the assumptions on $K(\cdot)$, and uniform bound on $w(\cdot)$ and $f_X(\cdot)$.

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

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

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

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

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

3. Rewrite T_3 as:

$$\begin{aligned}
T_3 &= \sup_{x\in\mathscr{G}}|s^{\tau}(x) - \mathrm{E}(s^{\tau}(x))| \le \sup_{x\in\mathscr{G}}|s^{\tau}(x) - s^{\tau}(x^k)| + \sup_{x\in\mathscr{G}}|\mathrm{E}(s^{\tau}(x) - s^{\tau}(x^k))| \\
&\quad + \max_{1\le k\le l_n}|s^{\tau}(x^k) - \mathrm{E}(s^{\tau}(x^k))| \equiv T_{31} + T_{32} + T_{33}.
\end{aligned}$$

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

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

$$\leq \quad C\left(\left(\frac{1}{nh_n^D}\right)^{1/2} + h_n\left(\frac{1}{nh_n^D}\right)^{1/2}\right)\frac{1}{n}\sum_{t=1}^{n}|g(\varepsilon_t)|\chi_{\{|g(\varepsilon_t)|\leq B_n\}},$$

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

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

3.3. $T_{33} = \max\limits_{1\leq k\leq l_n}|s^{\tau}(x^k) - \mathrm{E}(s^{\tau}(x^k))|$. Letting $\varepsilon_n = \left(\frac{nh_n^D}{\log n}\right)^{-1/2}\Delta_{\varepsilon}$ with $0 < \Delta_{\varepsilon} < \infty$, we have

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

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

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

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

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

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

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

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

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

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

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

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

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

□

Lemma 4. *Let $\{M_i\}_{i=1}^n$ be a sequence of independent and identically distributed random vectors with the same distribution as $M = (X \quad Z \quad U \quad \varepsilon)$ and $G(M)$ a continuous function of M with $\mathrm{E}(G^2(W)|Z) \leq C < \infty$. Then, if the joint density f_W of M is continuous,*

$$S_n = \frac{1}{n}\sum_{i=1}^n G(M_i)\left(\hat{\eta}(W_i,\hat{U}_i) - \eta(W_i,U_i)\right) = \begin{cases} o_p(n^{-1/2}), & \textit{if} \quad \mathrm{E}(G(M_i)|X_i,Z_i,U_i) = 0 \\ O_p\left(n^{-1/2} + \sum_{i=1}^4 h_i^{s_i}\right), & \textit{if} \quad \mathrm{E}(G(M_i)|X_i,Z_i,U_i) \neq 0 \end{cases}.$$

Proof. First, note that

$$\hat{\eta}(W_i,\hat{U}_i) - \eta(W_i,U_i) = \frac{1}{\hat{\phi}(W_i,\hat{U}_i)}\left[\left(\hat{f}_U(\hat{U}_i) - f_U(U_i)\right)\left(\hat{f}_W(W_i) - f_W(W_i)\right) + f_U(U_i)\left(\hat{f}_W(W_i) - f_W(W_i)\right)\right.$$
$$\left. + f_W(W_i)\left(\hat{f}_U(\hat{U}_i) - f_U(U_i)\right) - \eta(W_i,U_i)\left(\hat{\phi}(W_i,\hat{U}_i) - \phi(W_i,U_i)\right)\right]. \quad \text{(A.4)}$$

Also, for some $\lambda \in (0,1)$ and $d = 1,\cdots,D_2$,

$$\hat{U}_{id} - U_{id} = -\left(\frac{1}{f_Z(Z_i)}\frac{1}{nh_1^{D_1}}\sum_{t=1}^n K_1\left(\frac{Z_t - Z_i}{h_1}\right)(U_{td} + \mathbf{J}\Pi_d(Z_i - \lambda(Z_t - Z_i))(Z_t - Z_i))\right)(1 + O_p(L_{1n})).$$

Recall that $\hat{f}_U(\hat{U}_i) = \frac{1}{nh_2^{D_2}}\sum_{t=1}^n K_{2it} + \frac{1}{nh_2^{D_2+1}}\sum_{t=1}^n \mathbf{J}K_{2it}[\hat{U}_i - U_i - (\hat{U}_t - U_t)] + o_p(n^{-1/2})$, hence we write

$$\hat{f}_U(\hat{U}_i) = \frac{1}{nh_2^{D_2}}\sum_{t=1}^n K_{2it} + T_{1i}^U + T_{2i}^U + o_p(n^{-1/2}),$$

where $T_{1i}^U = -\dfrac{1}{n^2}\sum_{d=1}^{D_2}\sum_{l=1}^{n}\sum_{t=1}^{n}\dfrac{1}{h_1^{D_1}h_2^{D_2+1}f_Z(Z_i)}K_{1li}D_dK_{2it}\Big(U_{ld}+\mathbf{J}\Pi_d\big(Z_i-\lambda(Z_l-Z_i)\big)(Z_l-Z_i)\Big)(1+O_p(L_{1n})),$

$$T_{2i}^U = \frac{1}{n^2}\sum_{d=1}^{D_2}\sum_{l=1}^{n}\sum_{t=1}^{n}\frac{1}{h_1^{D_1}h_2^{D_2+1}f_Z(Z_t)}K_{1lt}D_dK_{2it}\Big(U_{ld}+\mathbf{J}\Pi_d\big(Z_t-\lambda(Z_l-Z_t)\big)(Z_l-Z_t)\Big)(1+O_p(L_{1n})).$$

Similarly, $\hat{\phi}(W_i,\hat{U}_i)=(nh_4^{D_4})^{-1}\sum_{t=1}^n K_{4ti}+T_{1i}^M+T_{2i}^M+o_p(n^{-1/2})$, where,

$$T_{1i}^M = -\frac{1}{n^2}\sum_{d=1}^{D_2}\sum_{l=1}^{n}\sum_{t=1}^{n}\frac{1}{h_1^{D_1}h_4^{D_4+1}f_Z(Z_i)}K_{1li}D_{2d}K_{4ti}\Big(U_{ld}+\mathbf{J}\Pi_d\big(Z_i-\lambda(Z_l-Z_i)\big)(Z_l-Z_i)\Big)(1+O_p(L_{1n})),$$
$$T_{2i}^M = \frac{1}{n^2}\sum_{d=1}^{D_2}\sum_{l=1}^{n}\sum_{t=1}^{n}\frac{1}{h_1^{D_1}h_4^{D_4+1}f_Z(Z_t)}K_{1lt}D_{2d}K_{4ti}\Big(U_{ld}+\mathbf{J}\Pi_d\big(Z_t-\lambda(Z_l-Z_t)\big)(Z_l-Z_t)\Big)(1+O_p(L_{1n})),$$

and $D_{2d}K_4(m,u)$ denotes the partial derivative of $K_4(m,u)$ with respect to u_d, the d^{th} element of u.

By assumption A5 and Theorem 1, we have that $\big(\hat{f}_U(\hat{U}_i)-f_U(U_i)\big)\big(\hat{f}_W(W_i)-f_W(W_i)\big)=o_p(n^{-1/2})$. In addition, $\frac{1}{\hat{\phi}(W_i,\hat{U}_i)}=\frac{1}{\phi(W_i,U_i)}+O_p(L_{4n})$. Thus, we write (A.4) as

$$\begin{aligned}
\hat{\eta}(W_i,\hat{U}_i)-\eta(W_i,U_i) = &\left(\frac{1}{\phi(W_i,U_i)}+O_p(L_{4n})\right)\Bigg[f_U(U_i)\left(\frac{1}{nh_3^{D_3}}\sum_{t=1}^n K_{3ti}-\mathrm{E}_t\left(\frac{1}{nh_3^{D_3}}\sum_{t=1}^n K_{3ti}\right)\right)\\
&+f_U(U_i)\left(\mathrm{E}_t\left(\frac{1}{nh_3^{D_3}}\sum_{t=1}^n K_{3ti}\right)-f_W(W_i)\right)+f_W(W_i)\left(\frac{1}{nh_2^{D_2}}\sum_{t=1}^n K_{2it}-\mathrm{E}_t\left(\frac{1}{nh_2^{D_2}}\sum_{t=1}^n K_{2it}\right)\right)\\
&+f_W(W_i)\left(\mathrm{E}_t\left(\frac{1}{nh_2^{D_2}}\sum_{t=1}^n K_{2it}\right)-f_U(U_i)\right)+f_W(W_i)T_{1i}^U+f_W(W_i)T_{2i}^U\\
&-\eta(W_i,U_i)\left(\frac{1}{nh_4^{D_4}}\sum_{t=1}^n K_{4ti}-\mathrm{E}_t\left(\frac{1}{nh_4^{D_4}}\sum_{t=1}^n K_{4ti}\right)\right)\\
&-\eta(W_i,U_i)\left(\mathrm{E}_t\left(\frac{1}{nh_4^{D_4}}\sum_{t=1}^n K_{4ti}\right)-\phi(W_i,U_i)\right)-\eta(W_i,U_i)T_{1i}^M-\eta(W_i,U_i)T_{2i}^M\\
&+o_p(n^{-1/2})\Bigg],
\end{aligned}$$

where E_t denotes an expectation taken with respect to the random variables indexed by t. Besides the $o_p(n^{-1/2})$ term, there are ten additional terms inside the brackets $[\cdot]$, which we label I_{nip}, with $p=1,\cdots,10$. We will establish the orders of $\frac{1}{n}\sum_{i=1}^n G(M_i)I_{nip}$ for $p=7,\cdots,10$. The the remaining terms are similar, and simpler, in structure. First, we consider

$$\frac{1}{n}\sum_{i=1}^n G(M_i)I_{ni7}=-\frac{1}{n}\sum_{i=1}^n G(M_i)\eta(W_i,U_i)\left(\frac{1}{\phi(W_i,U_i)}+O_p(L_{4n})\right)\left(\frac{1}{nh_4^{D_4}}\sum_{t=1}^n K_{4ti}-\mathrm{E}_t\left(\frac{1}{nh_4^{D_4}}\sum_{t=1}^n K_{4ti}\right)\right),$$

and it suffices to establish the order of

$$I'_{ni7} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{-G(M_i)\eta(W_i,U_i)}{\phi(W_i,U_i)h_4^{D_4}}\left(K_{4ti}-\mathrm{E}_t(K_{4ti})\right) \equiv \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\psi_{nit}$$
$$= \frac{1}{n^2}\sum_{i=1}^{n}\psi_{nii} + \frac{1}{n^2}\sum\sum_{i<t}(\psi_{nit}+\psi_{nti}) = \frac{1}{n^2}\sum_{i=1}^{n}\psi_{nii} + \left(\frac{n-1}{2n}\right)U_n,$$

where U_n is a U-statistic of degree 2. From Yao and Martins-Filho (2015), since $\mathrm{E}(\psi_{nit}+\psi_{nti})=0$,

$$U_n = \frac{2}{n}\sum_{i=1}^{n}\mathrm{E}\big((\psi_{nit}+\psi_{nti})|M_i\big) + O_p\left(n^{-1}(\mathrm{E}((\psi_{nit}+\psi_{nti})^2))^{1/2}\right).$$

Furthermore, since $\mathrm{E}((\psi_{nit}+\psi_{nti})^2) < C\mathrm{E}(\psi_{nit}^2)$ and given that $\mathrm{E}(G^2(M)) < C < \infty$, we have $\mathrm{E}((\psi_{nit}+\psi_{nti})^2) = O(h_4^{-D_4})$, and consequently the last term is $O_p(n^{-1}h_4^{-D_4/2}) = o_p(n^{-1/2})$. If $\mathrm{E}(G(M_i)|X_i,Z_i,U_i)=0$, then $\int(\psi_{nit}+\psi_{nti})f_W(W_t)\mathrm{d}W_t = 0$ and $U_n = o_p(n^{-1/2})$. If $\mathrm{E}(G(M_i)|X_i,Z_i,U_i)\neq 0$, then $\mathrm{E}(\psi_{nit}|M_i)=0$ and $\mathrm{E}\left(\mathrm{E}^2(\psi_{nti}|M_i)\right)=O(1)$. But since $\mathrm{E}(\psi_{nti}|M_i)\neq 0$, we have $U_n = O_p(n^{-1/2})$. By A5, we have $\frac{1}{n^2}\sum_{i=1}^{n}\psi_{nii} = O_p(n^{-1}h_4^{-D_4}) = o_p(n^{-1/2})$. Now, we consider

$$\frac{1}{n}\sum_{i=1}^{n}G(M_i)I_{ni8} = -\frac{1}{n}\sum_{i=1}^{n}G(M_i)\eta(W_i,U_i)\left(\frac{1}{\phi(W_i,U_i)}+O_p(L_{4n})\right)\left(\mathrm{E}_t\left(\frac{1}{nh_4^{D_4}}\sum_{t=1}^{n}K_{4ti}\right)-\phi(W_i,U_i)\right),$$

and it suffices to establish the order of $I'_{ni8} = -\frac{1}{n}\sum_{i=1}^{n}\frac{G(M_i)\eta(W_i,U_i)}{\phi(W_i,U_i)}\left(\mathrm{E}_t\left((nh_4^{D_4})^{-1}\sum_{t=1}^{n}K_{4ti}\right)-\phi(W_i,U_i)\right)$. Note that given assumption A1 and $\phi\in C^s$, by Taylor's Theorem we write

$$\mathrm{E}_t\left(\frac{1}{nh_4^{D_4}}\sum_{t=1}^{n}K_{4ti}\right) = \phi(W_i,U_i) + \sum_{|\beta|=s_4}\frac{1}{s_4!}\int K_4(\gamma)D^{\beta}\phi\big((W_i,U_i)+h_4\lambda\gamma\big)h_4^{s_4}\gamma^{\beta}\mathrm{d}\gamma \equiv \phi(W_i,U_i)+D_{\phi}(W_i,U_i),$$

where $D_{\phi}(W_i,U_i) = O(h_4^{s_4})$. Consequently, if $\mathrm{E}(G(M_i)|X_i,Z_i,U_i)=0$, then $I'_{ni8} = O_p(n^{-1/2}h_4^{s_4}) = o_p(n^{-1/2})$, and if $\mathrm{E}(G(M_i)|X_i,Z_i,U_i)\neq 0$, then $\mathrm{E}(|I'_{ni8}|) = O(h_4^{s_4})$ and $I'_{ni8} = O_p(h_4^{s_4})$.

For the term $\frac{1}{n}\sum_{i=1}^{n}G(M_i)I_{ni9} = -\frac{1}{n}\sum_{i=1}^{n}G(M_i)\eta(W_i,U_i)\left(\frac{1}{\phi(W_i,U_i)}+O_p(L_{4n})\right)T_{1i}^{M}$, we establish the order of

$$I'_{ni9} = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{G(M_i)\eta(W_i,U_i)}{\phi(W_i,U_i)f_Z(Z_i)h_1^{D_1}h_4^{D_4+1}}\sum_{d=1}^{D_2}D_{2d}K_{4ti}\left(U_{ld}+\mathbf{J}\Pi_d\big(Z_i-\lambda(Z_l-Z_i)\big)(Z_l-Z_i)\right)$$
$$= \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}(\psi_{1nitl}+\psi_{2nitl}),$$

where $$\psi_{1nitl} = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{G(M_i)\eta(W_i,U_i)}{\phi(W_i,U_i)f_Z(Z_i)h_1^{D_1}h_4^{D_4+1}}\sum_{d=1}^{D_2}K_{1li}D_{2d}K_{4ti}U_{ld},$$
$$\psi_{2nitl} = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{G(M_i)\eta(W_i,U_i)}{\phi(W_i,U_i)f_Z(Z_i)h_1^{D_1}h_4^{D_4+1}}\sum_{d=1}^{D_2}K_{1li}D_{2d}K_{4ti}\mathbf{J}\Pi_d\big(Z_i-\lambda(Z_l-Z_i)\big)(Z_l-Z_i).$$

For $i \neq t \neq l$, we have for $k = 1,2$ that $\frac{1}{n^3}\sum_{i\neq t\neq l}\psi_{knitl} = (6 - \frac{1}{2n} + \frac{1}{3n^2})U_n^k$, where U_n^k is a U-statistic of degree 3 with a symmetric kernel given by $\phi_{knitl} = \sum_{\mathscr{P}}\psi_{knitl}$, with $\mathscr{P}$ being the permutations of $\{i,t,l\}$.

We first consider U_n^1. Using Theorem 1 in Yao and Martins-Filho (2015) and noting that $\mathrm{E}(U_{ld}|Z_l) = 0$ we have $\theta_n = \mathrm{E}(\phi_{1nitl}) = 0$. Furthermore, using the notation for U-statistics

$$\sigma_{3n}^2 \leq C\mathrm{E}\left(\frac{G^2(W_i)\eta^2(W_i,U_i)}{\phi^2(W_i,U_i)f_Z^2(Z_i)h_1^{2D_1}h_4^{2D_4+2}}\sum_{d=1}^{D_2}K_{1li}^2 D_{2d}^2 K_{4ti}\mathrm{E}(U_{ld}^2|Z_l)\right) = O(h_1^{-D_1}h_4^{-D_4-2}).$$

Hence, $H_n^{(3)} = O_p(n^{-3/2}h_1^{-D_1/2}h_4^{-D_4/2-1}) = o_p(n^{-1/2})$. Similarly, given assumption A5 we have

$$H_n^{(2)} = O_p(n^{-1}h_1^{-D_1/2}h_4^{-D_4/2-1}) = o_p(n^{-1/2}).$$

Now, $\sigma_{1n}^2 \leq C\mathrm{E}\left(\mathrm{E}^2(\psi_{1nitl}|M_l)\right)$ since $\mathrm{E}(\psi_{1nitl}|M_i) = \mathrm{E}(\psi_{1nitl}|M_t) = 0$. If $\mathrm{E}(G(M_i)|X_i,Z_i,U_i) = 0$, then $\mathrm{E}(\psi_{1nitl}|M_l) = 0$ and $\sigma_{1n}^2 = 0$. We have $U_n^1 = o_p(n^{-1/2})$. If $\mathrm{E}(G(M_i)|X_i,Z_i,U_i) \neq 0$, then

$$\mathrm{E}(\psi_{1nitl}|M_l) \to \sum_{d=1}^{D_2}U_{ld}\int\frac{G(X_i,Z_l,U_i,\varepsilon_i)\eta(W_i,U_i)}{\phi(W_i,U_i)f_Z(Z_l)}D_{2d}\phi(W_i,U_i)f_W(X_i,Z_l,U_i,\varepsilon_i)\mathrm{d}X_i\mathrm{d}U_i\mathrm{d}\varepsilon_i.$$

Given that $\mathrm{E}(G^2(W_i)|Z_i) \leq C$, we have $\sigma_{1n}^2 = O(1)$, $H_n^{(1)} = O_p(n^{-1/2})$. Thus, $U_n^1 = O_p(n^{-1/2})$.

We now consider U_n^2. We note that $\sigma_{3n}^2 \leq O(h_1^{-D_1+4}h_4^{-D_4-1})$ and consequently $H_n^{(3)} = o_p(n^{-1/2})$. In a similar manner we obtain $H_n^{(2)} = o_p(n^{-1/2})$. Now,

$$\sigma_{1n}^2 \leq C\left(\mathrm{E}^2(\psi_{2nitl}|M_i) + \mathrm{E}^2(\psi_{2nitl}|M_l) + \mathrm{E}^2(\psi_{2nitl}|M_t)\right) \leq O(h_1^{2s_1}) + O(h_1^2) + O(h_1^{2s_1}h_2^{-2}),$$

where the orders in the last inequality follow from routine integration and the same arguments used to study ψ_{1nimt}. Consequently, we have $H_n^{(1)} = O_p(n^{-1/2}h_1 + n^{-1/2}h_1^{s_1}h_2^{-1}) = o_p(n^{-1/2})$. If $\mathrm{E}(G(M_i)|X_i,Z_i,U_i) = 0$, then $\mathrm{E}(\phi_{2nitl}|M_l) = 0$ and $U_n^2 = o_p(n^{-1/2})$. If $\mathrm{E}(G(M_i)|X_i,Z_i,U_i) \neq 0$, then $\theta_n = 6\mathrm{E}(\mathrm{E}(\psi_{2nitl}|M_i)) = O(h_1^{s_1})$ and $U_n^2 = o_p(n^{-1/2}) + O(h_1^{s_1})$.

For the additional cases, it is straightforward to verify that a) if $i = t = l$, $I'_{ni9} = O_p(n^{-2}h_1^{-D_1}h_4^{-D_4-1}) = o_p(n^{-1/2})$; b) if $i \neq t = l$, $I'_{ni9} = O_p(n^{-1}h_4^{-1}) = o_p(n^{-1/2})$; c) if $i = t \neq l$, $I'_{ni9} = O_p(n^{-1}h_4^{-D_4-1}) = o_p(n^{-1/2})$; d) and if $i = l \neq t$, $I'_{ni9} = O_p(n^{-1}h_1^{-D_1}h_4^{-1}) = o_p(n^{-1/2})$. So, collecting all the orders, we have

$$I_{ni9} = \begin{cases} o_p(n^{-1/2}), & \text{if} \quad \mathrm{E}(G(M_i)|X_i,Z_i,U_i) = 0 \\ O_p(n^{-1/2} + h_1^{s_1}), & \text{if} \quad \mathrm{E}(G(M_i)|X_i,Z_i,U_i) \neq 0 \end{cases}.$$

The term $\frac{1}{n}\sum_{i=1}^n G(M_i)I_{ni10} = -\frac{1}{n}\sum_{i=1}^n G(M_i)\eta(W_i,U_i)\left(\frac{1}{\phi(W_i,U_i)} + O_p(L_{4n})\right)T_{2i}^M$ can be treated precisely as $\frac{1}{n}\sum_{i=1}^n G(M_i)$

$\times I_{ni9}$, and we obtain exactly the same orders, viz., $I_{ni10} = \begin{cases} o_p(n^{-1/2}), & \text{if } \mathrm{E}(G(M_i)|X_i,Z_i,U_i)=0 \\ O_p(n^{-1/2}+h_1^{s_1}), & \text{if } \mathrm{E}(G(M_i)|X_i,Z_i,U_i)\neq 0 \end{cases}$. Combining the orders of terms I_{nip} for $p=1,\cdots,10$, we have

$$S_n = \begin{cases} o_p(n^{-1/2}), & \text{if} \quad \mathrm{E}(G(M_i)|X_i,Z_i,U_i)=0 \\ O_p\left(n^{-1/2}+\sum_{i=1}^{4} h_i^{s_i}\right), & \text{if} \quad \mathrm{E}(G(M_i)|X_i,Z_i,U_i)\neq 0 \end{cases}.$$

□