Oracle Efficiency and Stability in Additive Partially Linear Triangular Systems

November 23, 2018

Eric Penner [1]

Abstract: In this paper I propose a multi-stage mixed sieve and kernel estimator for a partially linear regression model in a triangular system of equations. The model consists of D+1 equations; a single partially linear primary equation having a mixture of endogenous and exogenous regressors, as well as D fully nonparametric secondary equations with exogenous regressors. Regressor endogeneity in the primary equation is handled using the control function approach of Newey et al. (1999). The estimator realizes efficiency gains by imposing an additive structure on the nonparametric component functions of the primary equation and secondary equations of the system (Yu et al. (2011)). As an added benefit, the additive structure circumvents the curse of dimensionality associated with nonparametric estimators. In particular, I show that the estimator of the parametric component β_1 is consistent, $\sqrt{n}$ asymptotically normally distributed, and Oracle efficient having an asymptotic covariance matrix equal to one derived from an identical estimation procedure for a model consisting solely of the primary equation where all regressors are exogeneous. Furthermore, I propose a consistent and easy to compute estimator for the asymptotic covariance matrix of the estimator for β_1. I subsequently plug my estimator for the parametric component into a two stage estimation procedure for the nonparametric component functions of the primary equation developed in Ozabaci et al. (2013) which results in estimates which are consistent, asymptotically normal, and Oracle efficient in the traditional nonparametric sense. Lastly I address the reproducibility of results generated with this estimator by demonstrating the stability of statistical inference conducted on $\hat{\beta}_1$ to the presence of weak instruments and overfitting.

Keywords: Partially Linear Regression; Endogeneity; Sieve Estimation; Kernel Estimation; Structural Model; $\sqrt{n}$ Asymptotic Normality; Nonparametric Modeling, Backfitting, Oracle Efficiency, Stability.

JEL Classifications. C13, C14

---

[1] School of Information, University of California, Berkeley 102 South Hall Berkeley, CA 94720-4600 ,email: epenner@berkeley.edu, website: https://ericpnr.github.io/

1 Introduction

Endogeneity is among the most common problems encountered in empirical economic research. Originally studied in the context of simultaneity by Working (1927) and Wright (1928), endogeneity typically arises whenever observed data is subject to an equilibrium condition (Haavelmo, 1943), when agents can select themselves into a sample (Heckman, 1979), when regressors are measured with error (see e.g. Hausman et al. (1991) and Hausman et al. (1995)), or when agents can choose the value of one or more regressors in order to optimize an observed outcome (see e.g. Olley and Pakes (1996), and Petrin and Train (2010)). In light of this and the behavior typically assumed of economic agents, one would expect endogenous relationships to be present in most economic data sets. As a result, the study of endogeneity remains one of the most active and important areas of econometric research. Furthermore, as businesses seek tools to perform data driven decision making and rapid policy analysis with ever larger data sets the demand for developing fast and flexible methods for treating endogeneity will only increase.

Broadly, there are two approaches to identifying a regression in the presence of endogenous regressors; instrumental variables, and control function. The instrumental variables approach identifies the structural relationship $m(X)$ between regressand Y and a vector of endogenous regressors X as a consequence of the identification of $E(X|W)$ where W is a vector of instruments.

For example, if $m(X) = X'\beta$ then β is identified by $E(Y|W) = E(X|W)'\beta$ and can be consistently estimated by regressing Y on estimates of $E(X|W)$ as is done in two stage least squares (Angrist and Pischke, 2009). In general however, the problem of recovering a nonparametric $m(X)$ from $E(Y|W)$ is ill posed, a condition which leads to estimators that require time consuming numerical optimization routines (Ai and Chen (2003)), the careful choice of regularization parameters (Darroles et al. (2011)), and have convergence rates slower then identical models without endogenous regressors (Chen and Qui, 2016). As a result, in time sensitive applications where endogenous regressors are present in large, often terabyte sized, data sets such estimators are of little practical use. In order to allow fast computation with large data sets there have been attempts to adapt the estimation (or training) of deep neural networks to perform instrumental variables estimation e.g. Hartford et al. (2017). Unfortunately, as is common in the machine learning field, at best, these techniques provide an incomplete framework for inference.

Alternatively, the control function approach identifies $m(X)$ by assuming the existence of a vector of instruments W, which allows for the decomposition of X into two components. The first component $E(X|W)$ embodies the variation in X which is exogenous, having no functional relationship with the error term ε of the primary regression. The second component $V = X - E(X|W)$ is the endogeneous variation in X or the component which has some functional relationship with ε. The idea is that the endogenous relationship between X and ε is embodied by V and once V is included as a regressor in the primary regression $m(X)$ is identified. In contrast to instrumental variables, the control function approach leads to estimators with analytical solutions (see inter alia Newey et al. (1999), Pinkse (2000), Ozabaci et al. (2013)) allowing for fast computation, making them far more practical for time sensitive large data set applications.

The downside of taking the control function approach is that it typically results in multi step semiparametric estimators where the final step of estimation incorporates regressors nonparametrically generated in preliminary steps. The problem of generated regressors in semiparametric estimation has been studied extensively in recently years, see inter alia Escanciano et al. (2016), Hahn and Ridder (2013), and Mammen et al. (2016), where the asymptotic distribution of a feasible estimator, one depending on a vector of generated regressors, is compared to the asymptotic distribution of an infeasible *oracle* estimator where all regressors are observed directly. In the case of a finite dimensional parameter Mammen et al. (2016) show that, in general, the asymptotic covariance matrix of the feasible estimator is unambiguously larger than the infeasible *oracle* estimator and that this covariance matrix may be difficult to estimate in practice. This is the case in Geng et al. (2017) where the asymptotic covariance matrix contains an additional term not present in its corresponding oracle estimator. Although Mammen et al. (2016) show that the presence of this additional term is true in general, they do allow for the existence of an estimator with the *oracle property* (Escanciano et al., 2010) or is *oracle efficient* meaning that the asymptotic covariance matrix of the feasible estimator is identical to its infeasible oracle estimator, meaning that asymptotically there is no penalty to conducting estimation with generated regressors rather than observing them directly.

The model studied in this paper takes the control function approach to treating endogeneity in an additive partially linear regression by decomposing each endogenous regressor with an additive nonparametric regression on instruments W. In this paper I show that the imposition of this additive structure combined with an identification procedure developed in Manzan and Zerom (2005) results in an estimator for the parametric component β_1 which is consistent, $\sqrt{n}$ asymptotically normally distributed, and oracle efficient as its asymptotic covariance matrix is identical to its oracle estimator developed in Manzan and Zerom (2005). As a result this estimator is fast, easy to compute, and provides for easy asymptotic inference while avoiding the main drawback of the control function approach. Furthermore I show that $\hat{\beta}_1$ can be incorporated into a procedure developed in Ozabaci et al. (2013) for the estimation of each additive component of $h(X)$ without affecting the estimators rate of convergence or oracle efficiency.

In particular, this model consists of D+1 equations; a single partially linear primary equation (1) with a mixture of endogenous and exogenous regressors, as well as D fully nonparametric secondary equations (2) having only exogenous regressors. Consider the following triangular system of equations where a set of exogenous regressors enters the primary equation (1) parametrically while allowing endogenous regressors to enter nonparametrically.

$$Y = \beta_0 + Z'\beta_1 + \sum_{d=1}^{D} h_d(X_d) + \varepsilon, \tag{1}$$

$$X = \sum_{a=1}^{q} m_a(W_a) + V, \tag{2}$$

$$E(V|W) = 0, \quad E(\varepsilon|W, X) = E(\varepsilon|V). \tag{3}$$

Y is a scalar random variable, X is a D dimensional vector of endogenous random variables in that $E(\varepsilon|Z, X) = E(\varepsilon|X) \neq 0$. W is a q dimensional random vector, and Z is a $p < q$ dimensional subvector of W. ε and V are random disturbances. $[\beta_0 \;\; \beta_1']' \in \mathbb{R}^{p+1}$ and $h_d(\cdot)$ are unknown parameters of interest. Each $m_a(W_a) : \mathbb{R} \to \mathbb{R}^D$ is a vector of D real valued functions $m_{ad}(W_a)$ where $d \in \{1, 2, \ldots, D\}$.

Specifying a partially linear form for the primary equation has the distinct advantage of allowing one to impose a parametric form for some regressors, when justified, while allowing others to be the arguments of a much broader class of functions. This type of dimension reduction often allows for the separate identification and estimation of parametric and nonparametric components in such a way that the rate of convergence of $\hat{\beta}_1$ is much faster than the nonparametric component. For example, in the case of a non additive unknown function $h(X)$, Robinson (1988) showed that in models consisting solely of equation (1) where $E(\varepsilon|Z, X) = 0$, his estimator for $[\beta_0 \;\; \beta_1']'$ is consistent, $\sqrt{n}$ asymptotically normal, and semiparametrically efficient while only making mild smoothness assumptions regarding $h(X)$. Similar results were established for an additive $h(X)$ in Li (2000). Enabling the separate estimation of β_1 from $h(X)$ is, in essence, an identification problem akin to the presence of endogenous regressors. Here we have a structural relationship $Z'\beta_1$ which we would like to isolate and identify but can't without requiring $E(Y - Z'\beta|Z) = E(h(X) + \varepsilon|Z) = 0$. To remedy this one can take an approach similar to instrumental variables by projecting Y into the space of functions of a vector of instruments (in this case X) in such a way that β_1 can be identified by a residual of that projection e.g by $Y - E(Y|X)$ in Robinson (1988), and by $Y - \sum_d E(Y|X_d)$ in Li (2000). The identification technique used in this paper, developed in Manzan and Zerom (2005), is a variation of the technique of used Li (2000) where Y is scaled by a density ratio before projection.

The primary aim of this paper is to study the effect that imposing an additive structure on all unknown functions has on the ultimate estimation of β_1. In general, the benefits of imposing an additive structure are two fold. Firstly, Yu et al. (2011) show that in the partially linear model of Robinson (1988) there are efficiency gains to be had in the estimation of the parametric component by imposing additivity on unknown function $h(X)$. In this sense both Li (2000) and Manzan and Zerom (2005) can be viewed as extensions of the Robinson (1988) model attempting to achieve these efficiency gains. In fact, like Robinson, Manzan and Zerom (2005) showed that their estimator for $[\beta_0 \quad \beta_1']'$, is $\sqrt{n}$ asymptotically normal and semiparametrically efficient in the case of homoskedastic errors. This paper imposes additivity on $h(X)$ in an effort to realize these efficiency gains in the greater context of endogeneity. Secondly, as pointed out in Linton and Nielsen (1995), and Buja et al. (1989), imposing additivity on an unknown function allows one to develop estimators which avoid the "curse of dimensionality" while sacrificing little in terms of model flexibility. This "curse" refers to the case where the sum of univariate nonparametric estimators converges at a rate equal to the slowest univariate estimator, while the rate of convergence of a multivariate estimator is inversely proportional to the dimension of its argument. As shown in Mammen et al. (2016), the influence that a generated regressor has on the asymptotic covariance matrix of $\hat{\beta}_1$ is additively separable from the component originating from the oracle estimator. As a result, I impose additivity on $m(W)$ to avoid this "curse" so that the influence of these generated regressors is asymptotically negligible i.e. is oracle efficient. The combination of imposing additivity on unknown functions in both (1) and (2) results in an estimator that has an asymptotic covariance matrix equal to that of its infeasible additive partially linear oracle estimator which is semiparametrically efficient[2] in the case of homoskedastic errors.

Recently there has been a great deal of attention in the sciences on the inability of empirical researchers to replicate their own and others research. In some fields this "Reproducibility Crisis" (Baker, 2016) is so severe that as little as 11% (Begley and Ellis, 2012) of published results are able to be replicated. A number of solutions for this crisis have been offered such as pre-registration and a realignment of publishing incentives (Nosek et al., 2012) but many have called for "better statistics". For example, in Yu (2013) the author makes the case that "At a minimum reproducibility manifests itself in the stability of statistical results relative to 'reasonable' perturbations to data" e.g. small changes to sample size and quality should be accompanied by correspondingly small changes in statistical results. A criticism often leveled at nonparametric estimators is that the benefits of more complex and flexible econometric models come at the cost of stronger identification

[2]Note that I have not made any claim to semiparametric efficiency for my estimator as the semiparametric lower bound for my model has yet to be established

conditions that make estimators more sensitive to data perturbations. In the context of the multistep semiparametric estimator presented in this paper concerns tend to focus on the sensitivity of statistical inference on $\hat{\beta}_1$ to changes in the "quality" of generated regressors caused by two related sources; weak instruments, and overfitting.

Weak instruments, are loosely defined as having a low but non zero correlation between endogenous regressors X and instruments W. Originally studied by Morimune (1983), and later by Staiger and Stock (1997) the problem was formulated in the context of two stage least squares IV estimation where the coefficients on the first stage regression(s) are close, or converging, to zero giving a very "flat" regression function that leaves very little identifying variation left over for inclusion in the primary regression function. In the control function approach we have the opposite problem where instead of too little variation we have too much. This is because, as the variation in X_i explained by W_i diminishes, the left over exogenous variation accumulates in $\hat{V}_i$ effectively concealing the endogneous variation needed to identify the primary regression. In more practical terms as the correlation between X_i and W_i diminishes $\hat{V}_i$ converges to X_i ultimately causing a multicolinearity problem (Han, 2017) in the regressor matrix of the typical two step non parametric control function approach like Ozabaci et.al (2015).

Overfitting, is the act of estimating a regression within a space of functions which is too complex, when measured in terms of its VC dimension, relative to the number of observations used (Vapnik, 1998). The practical effect of overfitting is that the estimation procedure allocates a portion of the variation in the regressand originating from error term to the regressors. In the context of instrumental variables estimation, over fitting allows some of the endogenous variation V_i that one is trying to filter out of X_i to be allocated to the estimator of its exogenous variation $\hat{E}(X_i|W_i)$ in effect concealing the variation needed to identify the primary regression. The opposite is true for the control function approach, as the VC dimension of the space in which one is estimating $E(X_i|W_i)$ increases to the sample size n the left over variation needed to identify the primary regression, $\hat{V}_i = X_i - \hat{E}(X_i|W_i)$ converges to zero. Here one can see that overfitting in the control function approach is a problem akin to weak instruments in IV estimation while weak instruments in IV estimation is akin to overfitting in the control function approach. In this paper I will demonstrate the stability of the estimator for β_1 to varying degrees of overfitting and weak instruments.

The remainder of this paper is organized into 5 Sections; Section 2 details the relevant moment conditions, identification, and estimation of my model. Section 3 gives all necessary assumptions, and summarizes the asymptotic results for each step of the estimation procedure. Section 4 presents a Monte Carlo study of this estimation procedure that demonstrates the finite sample properties of my estimator for $[\beta_0 \;\; \beta_1']'$. Section 5 gives a demonstration of the stability of the estimator to weak instruments and overfitting. Section 6 gives a summary and conclusion. Proofs of all primary Theorems and Lemmas are given in the Appendix. Proof of all supporting lemmas are given in the accompanying technical supplement.

2 Moment Conditions, Identification, and Estimation

2.1 Moment Conditions

$E(\varepsilon|X) \neq 0$ implies that ε is not orthogonal to the space of measurable functions of X. The approach taken to deal with this lack of orthogonality is a variation on the control function approach of Newey et al. (1999). In particular, defining $u = \varepsilon - f(V)$, I assume,

$$E[\varepsilon|Z, X, V] = E[\varepsilon|V] = E[f(V) + u|V] = f(V) + E[u|V] = f(V), \qquad \text{and} \qquad E[u|W, X, V] = 0. \tag{4}$$

Furthermore, I assume,

$$h(X) = \sum_{d=1}^{D} h_d(X_d), \qquad f(V) = \sum_{d=1}^{D} f_d(V_d), \quad \text{and} \quad m_d(W) = \sum_{a=1}^{q} m_{da}(W_a). \tag{5}$$

Consequently,

$$Y = \beta_0 + Z'\beta_1 + \sum_{d=1}^{D} h_d(X_d) + \sum_{d=1}^{D} f_d(V_d) + u, \tag{6}$$

$$X_d = \sum_{a=1}^{q} m_{da}(W_a) + V_d, \tag{7}$$

$$E(V_d|W) = 0, \quad E(\varepsilon|W, X, V) = 0. \tag{8}$$

The goal of this paper is to identify and estimate the parameters $[\beta_0 \ \ \beta_1']' \in \mathbb{R}^{p+1}$ in such a way that neither procedure requires any restriction on the functions $h(X)$ and $f(V)$ beyond the standard identification condition that $E(h_d(X_d)) = 0$ and $E(f_d(V_d)) = 0$. The identification detailed in the following section is a variation of the identification procedure developed in Manzan and Zerom (2005).

2.2 Identification

Let $p(\cdot)$ be the marginal/joint densities (provided they exist) of its random variable/vector argument and define,

$$g(X,V)=\prod_{j=1}^{D}p(X_j)\prod_{k=1}^{D}p(V_k),\quad g(X_{-d},V)=\prod_{j\neq d}^{D}p(X_j)\prod_{k=1}^{D}p(V_k),\quad g(X,V_{-d})=\prod_{j=1}^{D}p(X_j)\prod_{k\neq d}^{D}p(V_k).$$

Furthermore, as in Kim et al. (1999) and Geng et al. (2016), define an "instrument function" $\phi\equiv\phi(X,V)\equiv g(X,V)/p(X,V)$ and related functions, $\theta_1^d\equiv\theta_1^d(X,V)\equiv g(X_{-d},V)/p(X,V)$ and $\theta_2^d\equiv\theta_2^d(X,V)\equiv g(X,V_{-d})/p(X,V)$. Now, let A be any subvector of $[\,Y\ \ Z'\,]'$, let $\mu_A=E[\phi A]$, and define

$$H_1^d(A)=E[\phi A|X_d],\qquad H_2^d(A)=E[\phi A|V_d],\tag{9}$$

$$H(A)=\sum_{d=1}^{D}\Big[H_1^d(A)+H_2^d(A)\Big],\qquad H^*(A)=H(A)-(2D-1)\mu_A.\tag{10}$$

Furthermore, for all $c\in\{1,2,\dots,p\}$ define the following differences,

$$\zeta_c\equiv Z_c-H^*(Z_c),\qquad \zeta\equiv Z-H^*(Z),\tag{11}$$

$$\rho_c\equiv Z_c-E[Z_c|X,V],\qquad \rho\equiv Z-E[Z|X,V],\tag{12}$$

$$\eta_{1c}^d\equiv E[Z_c|X,V]-H_1^d(Z_c),\qquad \eta_{2c}^d\equiv E[Z_c|X,V]-H_2^d(Z_c),\tag{13}$$

$$\eta\equiv E[Z|X,V]-H^*(Z).\tag{14}$$

The following lemma gives conditions for the identification of $[\beta_0\ \ \beta_1']'\in\mathbb{R}^{p+1}$.

Lemma 1. *Parameters β_0 and β_1 are identified if either of the following sets of condition are satisfied.*

i) $E(h_d(X_d))=0$, $E(f_d(V_d))=0$ *for all* $d\in\{1,2,\dots,D\}$ *and either*

- *a)* $E(\rho\rho')>0$, *or*
- *b)* $E(\eta\eta')>0$.

ii) $E(h_d(X_d))=0$, $E(f_d(V_d))=0$ *for all* $d\in\{1,2,\dots,D\}$ *and,* $E(\phi\zeta\zeta')>0$.

In addition to the conditions under which β_0 and β_1 are identified, I now state moment conditions implied by identification, which are critical to the proof of $\sqrt{n}$ convergence of the estimators for $[\beta_0, \beta_1']'$.

$$E[\phi\zeta_c|X_d] = 0, \qquad\qquad E[\phi\zeta_c|V_d] = 0, \tag{15}$$

$$E[\phi(Y - H^*(Y))|X_d] = 0, \qquad\qquad E[\phi(Y - H^*(Y))|V_d] = 0. \tag{16}$$

2.3 Estimation

Basis Functions: In steps one and five of the following estimation procedure I utilize normalized B-Spline basis functions in order to perform series estimation. Here I present the functions, and vectors necessary to define these estimators. Let A be any scalar random variable with compact support G_A, θ be any natural number, $\{z_n\}_{n=1}^{\infty}$ be any non decreasing sequence of natural numbers, and $\{e_j\}_{j=1}^{z_n+2\theta}$ be any knot set[3] for G_A. For any $k \in \mathbb{N}$ define,

$$b_{j,1}(A) = \begin{cases} 1 & \text{if } A \in [e_j, e_{j+1}) \\ 0 & \text{if } A \notin [e_j, e_{j+1}), \end{cases} \qquad \text{and} \qquad \omega_{j,k}(A) = \frac{A - e_j}{e_{j+k-1} - e_j}.$$

Now, define the jth normalized B-spline basis function of order θ recursively,

$$b_{j,\theta}(A) = \omega_{j,\theta}(A) b_{j,\theta-1}(A) + (1 - \omega_{j+1,\theta}(A)) b_{j+1,\theta-1}(A), \qquad \text{and} \qquad B_{j,\theta}(A) = \frac{b_{j,\theta}(A)}{||b_{j,\theta}(A)||_2}.$$

Where $\{B_{j,\theta}(A)\}_{j=1}^{z_n+2\theta}$ is a sequence of uniformly bounded, positive, twice integrable, compactly supported B-spline basis functions. Now applying this definition, let $\{l_n\}_{n=1}^{\infty}$ and $\{k_n\}_{n=1}^{\infty}$ be a nondecreasing sequences of natural numbers. Let G_{W_a}, G_{X_d}, and G_{V_d} be the compact supports of W_a, X_d, and V_d respectively. For some $k, a \in \mathbb{N}$ let $\{t_j\}_{j=1}^{l_n+2k}$, $\{t_{X_d j}\}_{j=1}^{k_n+2a}$ and $\{t_{V_d j}\}_{j=1}^{k_n+2a}$ be a knot sets for G_{W_a}, G_{X_d} and G_{V_d} respectively and define the following basis vectors;

$$\mathbf{B}_n^{l_n}(W_{ai})' = [\; B_{1,k}(W_{ai}) \;\; B_{2,k}(W_{ai}) \;\; \cdots \;\; B_{l_n+2k,k}(W_{ai}) \;],$$

$$\mathbf{P}_n^{k_n}(X_{di})' = [\; B_{1,a}(X_{di}) \;\; B_{2,a}(X_{di}) \;\; \cdots \;\; B_{k_n+2a,a}(X_{di}) \;],$$

$$\mathbf{P}_n^{k_n}(V_{di})' = [\; B_{1,a}(V_{di}) \;\; B_{2,a}(V_{di}) \;\; \cdots \;\; B_{k_n+2a,a}(V_{di}) \;].$$

[3] See de Boor (2001, pg.149) for details of knot set construction

Motivated by the moment and identification conditions given in Section 2.2, and the assumption that $\{Y_i, X_i, W_i\}_{i=1}^n$ is an i.i.d sequence (see Assumption A3), I now describe the estimation procedure.

Step One: Obtain a B-Spline series estimate of $m(W_i)$ with dth element

$$\hat{m}_d(W_i) = \mathbf{B}_n(W_i)'(\mathbf{B}_n'\mathbf{B}_n)^{-1}\mathbf{B}_n'\mathbf{X}_{dn}. \tag{17}$$

where, $\mathbf{B}_n(W_i) = \left[\mathbf{B}_n^{l_n}(W_{1i}) \;\; \mathbf{B}_n^{l_n}(W_{2i}) \;\; \cdots \;\; \mathbf{B}_n^{l_n}(W_{qi})\right]'$ is a $(q(l_n + 2k) \times 1)$ vector of B-spline basis functions, $\mathbf{B}_n = [\mathbf{B}_n(W_1) \;\; \mathbf{B}_n(W_2) \;\; \cdots \;\; \mathbf{B}_n(W_n)]'$, and $\mathbf{X}_{dn} = [X_{d1} \;\; X_{d2} \;\; \cdots \;\; X_{dn}]'$. These estimates will then be used to obtain residuals,

$$\hat{V}_{di} = X_{di} - \hat{m}_d(W_i), \qquad \text{and define} \qquad \hat{\mathbf{V}}_{dn} = \left[\begin{array}{cccc} \hat{V}_{d1} & \hat{V}_{d2} & \cdots & \hat{V}_{dn} \end{array}\right]'. \tag{18}$$

Step Two: Obtain Rosenblatt kernel density estimates of $p(X_i, V_i), p(X_{di})$, and $p(V_{di})$ using observed values X_i and estimated values $\hat{V}_i$. Let $\mathbf{1}_D$ be a $(D \times 1)$ vector of ones, and define $H = diag([h_1 \cdot \mathbf{1}_D', h_2 \cdot \mathbf{1}_D']')$,

$$\hat{p}(X_{di}) = \frac{1}{nh_0}\sum_{l=1}^n K_0\left(h_0^{-1}[X_{dl} - X_{di}]\right), \qquad \hat{p}(\hat{V}_{di}) = \frac{1}{nh_0}\sum_{l=1}^n K_0\left(h_0^{-1}[\hat{V}_{dl} - \hat{V}_{di}]\right), \tag{19}$$

$$\hat{p}(X_i, \hat{V}_i) = \frac{1}{nh_1^D h_2^D}\sum_{l=1}^n K_3\big(H^{-1}[(X_l', \hat{V}_l')' - (X_i', \hat{V}_i')']\big), \tag{20}$$

where $K_0(\cdot) : \mathbb{R} \to \mathbb{R}$ is a univariate Kernel with bandwidth h_0. $K_3(\cdot) : \mathbb{R}^{2D} \to \mathbb{R}$ is a multivariate kernel with associated bandwidths h_1 corresponding to the X arguments and h_2 corresponding to the $\hat{V}$ arguments. Next define,

$$\hat{g}(X_i, \hat{V}_i) = \prod_{d=1}^D \hat{p}(X_{di})\hat{p}(\hat{V}_{di}), \quad \hat{g}(X_{-di}, \hat{V}_i) = \prod_{j \neq d}^D \hat{p}(X_{ji}) \prod_{k=1}^D \hat{p}(\hat{V}_{ki}), \quad \hat{g}(X_i, \hat{V}_{-di}) = \prod_{j=1}^D \hat{p}(X_j) \prod_{k \neq d}^D \hat{p}(\hat{V}_k),$$

with which I obtain $\hat{\phi}_i \equiv \hat{g}(X_i, \hat{V}_i)/\hat{p}(X_i, \hat{V}_i)$, $\hat{\theta}_{1i}^d \equiv \hat{g}(X_{-di}, \hat{V}_i)/\hat{p}(X_i, \hat{V}_i)$, and $\hat{\theta}_{2i}^d \equiv \hat{g}(X_i, \hat{V}_{-di})/\hat{p}(X_i, \hat{V}_i)$.

Step Three : Obtain Nadaraya-Watson estimates of conditional expectation's $H_1^d(Z_{ci}), H_2^d(Z_{ci})$, $H_1^d(Y_i)$, and $H_2^d(Y_i)$, i.e.

$$\hat{H}_1^d(Z_{ci}) = [(n-1)b_1]^{-1}\sum_{l \neq i}^n K_1[b_1^{-1}(X_{dl} - X_{di})]\hat{\theta}_{1l}^d Z_{cl}, \tag{21}$$

$$\hat{H}_2^d(Z_{ci}) = [(n-1)b_2]^{-1} \sum_{l \neq i}^{n} K_2[b_2^{-1}(\hat{V}_{dl} - \hat{V}_{di})]\hat{\theta}_{2l}^d Z_{cl}, \tag{22}$$

$$\hat{H}_1^d(Y_i) = [(n-1)b_1]^{-1} \sum_{l \neq i}^{n} K_1[b_1^{-1}(X_{dl} - X_{di})]\hat{\theta}_{1l}^d Y_l, \tag{23}$$

$$\hat{H}_2^d(Y_i) = [(n-1)b_2]^{-1} \sum_{l \neq i}^{n} K_2[b_2^{-1}(\hat{V}_{dl} - \hat{V}_{di})]\hat{\theta}_{2l}^d Y_l, \tag{24}$$

where $K_1(\cdot), K_2(\cdot) : \mathbb{R} \to \mathbb{R}$ are univariate Kernels with bandwidths b_1 and b_2 respectively. Additionally, $\mu(Z_c)$, and μ_Y are estimated with $\hat{\mu}_{Z_c} = n^{-1}\sum_{i=1}^{n} \hat{\phi}_i Z_{ci}$ and $\hat{\mu}_Y = n^{-1}\sum_{i=1}^{n} \hat{\phi}_i Y_i$ so that,

$$\hat{H}(Z_{ci}) = \sum_{d=1}^{D}[\hat{H}_1^d(Z_{ci}) + \hat{H}_2^d(Z_{ci})], \qquad \hat{H}(Y_i) = \sum_{d=1}^{D}[\hat{H}_1^d(Y_i) + \hat{H}_2^d(Y_i)], \tag{25}$$

$$\hat{H}^*(Z_{ci}) = \hat{H}(Z_{ci}) - (2D-1)\hat{\mu}_{Z_c}, \qquad \hat{H}^*(Y_i) = \hat{H}(Y_i) - (2D-1)\hat{\mu}_Y. \tag{26}$$

Let $\hat{\mathbf{H}}_n^*(Y) = [\ \hat{H}^*(Y_1) \quad \hat{H}^*(Y_2) \quad \cdots \quad \hat{H}^*(Y_n)\ ]'$, $\hat{\mathbf{H}}_n^*(Z_c) = [\ \hat{H}^*(Z_{c1}) \quad \hat{H}^*(Z_{c2}) \quad \cdots \quad \hat{H}^*(Z_{cn})\ ]'$ and $\hat{\mathbf{H}}_n^*(Z) = [\ \hat{\mathbf{H}}_n^*(Z_1) \quad \hat{\mathbf{H}}_n^*(Z_2) \quad \cdots \quad \hat{\mathbf{H}}_n^*(Z_p)\ ]$ where $\mathbf{Y}_n = [\ Y_1 \quad Y_2 \quad \cdots \quad Y_n\ ]'$, and $\mathbf{Z}_n = [Z_1 \quad Z_2 \quad \cdots \quad Z_n\ ]'$ and let,

$$\hat{\zeta}_{ci} = Z_{ci} - \hat{H}^*(Z_{ci}), \qquad \hat{\zeta}_i = Z_i - \hat{H}^*(Z_i), \qquad \hat{\boldsymbol{\zeta}}_n = \mathbf{Z}_n - \hat{\mathbf{H}}_n^*(Z). \tag{27}$$

Step Four: Estimation of $(\ \beta_0 \quad \beta_1'\ )' \in \mathbb{R}^{p+1}$.
I regress $\hat{\boldsymbol{\phi}}_n^{1/2}(\mathbf{Y}_n - \hat{\mathbf{H}}_n^*(Y))$ on $\hat{\boldsymbol{\phi}}_n^{1/2}\hat{\boldsymbol{\zeta}}_n = \hat{\boldsymbol{\phi}}_n^{1/2}(\mathbf{Z}_n - \hat{\mathbf{H}}_n^*(Z))$ and estimate β_1 as,

$$\hat{\beta}_1 = [\hat{\boldsymbol{\zeta}}_n'\hat{\boldsymbol{\phi}}_n\hat{\boldsymbol{\zeta}}_n]^{-1}\hat{\boldsymbol{\zeta}}_n'\hat{\boldsymbol{\phi}}_n(\mathbf{Y}_n - \hat{\mathbf{H}}_n^*(Y)). \tag{28}$$

where $\hat{\boldsymbol{\phi}}_n \equiv \hat{\boldsymbol{\phi}}_n(X, \hat{V}) = diag(\{\hat{\phi}(X_i, \hat{V}_i)\}_{i=1}^n)$. Lastly, β_0 is estimated by,

$$\hat{\beta}_0 = n^{-1}\sum_{i=1}^{n} Y_i - n^{-1}\sum_{i=1}^{n} Z_i'\hat{\beta}_1. \tag{29}$$

Step Five: Use estimates $[\hat{\beta}_0 \ \ \hat{\beta}_1']'$ to obtain a B-Spline estimate of $h(X) + f(V)$;

$$\hat{h}(X) + \hat{f}(V) = \mathbf{P}_n(X,V)'\hat{\Gamma}_n \quad \text{where} \quad \hat{\Gamma}_n = (n^{-1}\hat{\mathbf{P}}_n'\hat{\mathbf{P}}_n)^{-1}n^{-1}\hat{\mathbf{P}}_n'(\mathbf{Y}_n - \hat{\beta}_0\mathbf{1}_n - \mathbf{Z}_n\hat{\beta}_1)$$

and where $\mathbf{P}_n(X_i, V_i) = [\ \mathbf{P}_n^{k_n}(X_{1i})' \ \ \mathbf{P}_n^{k_n}(X_{2i})' \ \ \cdots \ \ \mathbf{P}_n^{k_n}(X_{Di})' \ \ \mathbf{P}_n^{k_n}(V_{1i})' \ \ \mathbf{P}_n^{k_n}(V_{2i})' \ \ \cdots \ \ \mathbf{P}_n^{k_n}(V_{Di})'\ ]'$

is a $(2D(k_n+2a)\times 1)$ dimensional vector, and $\mathbf{P}_n' = [\mathbf{P}_n(X_1,V_1) \;\; \mathbf{P}_n(X_1,V_1) \;\; \cdots \;\; \mathbf{P}_n(X_n,V_n)\;]'$ is an $(n\times 2D(k_n+2a))$ dimensional vector.

Step Six: Obtain 1-Step Local Linear Kernel Backfitting Estimates of each $h_d(X_d)$;

$$[\;\tilde{h}_d(x_d)\;\; \tilde{h}_d^{(1)}(x_d)\;]' = \left(\mathbf{X}_{dn}(x_d)'\mathbf{K}_{4n}(x_d)\mathbf{X}_{dn}(x_d)\right)^{-1}\mathbf{X}_{dn}(x_d)'\mathbf{K}_{4n}(x_d)\tilde{\mathbf{Y}}_{dn}$$

where, $X_{di}(x_d) = [\;1 \quad (X_{di}-x_d)\;]'$, and $\mathbf{X}_{dn}(x_d) = [\;X_{d1}(x_d) \quad X_{d2}(x_d) \quad \cdots \quad X_{dn}(x_d)\;]'$, and for bandwidth $h_4 = o(1)$, $\mathbf{K}_{4n}(x_d) = h_4^{-1}\,\mathrm{diag}\left(\{K_4[h_4^{-1}(X_{di}-x_d)]\}_{i=1}^n\right)$ are matrices of order (2×1), $(n\times 2)$, and $(n\times n)$ respectively. Furthermore , $\hat{\mathbf{h}}_{-dn} = [\;\sum_{j\neq d}^D \hat{h}_j(X_{j1}) \quad \sum_{j\neq d}^D \hat{h}_j(X_{j2}) \quad \cdots \quad \sum_{j\neq d}^D \hat{h}_j(X_{jn})\;]'$, $\hat{\mathbf{f}}_n = [\;\sum_{j=1}^D \hat{f}_j(V_{j1}) \quad \sum_{j=1}^D \hat{f}_j(V_{j2}) \quad \cdots \quad \sum_{j=1}^D \hat{f}_j(V_{jn})\;]'$, and $\tilde{\mathbf{Y}}_{dn} = \mathbf{Y}_n - \hat{\beta}_0\mathbf{1}_n - \mathbf{Z}_n\hat{\beta}_1 - \hat{\mathbf{h}}_{-dn} - \hat{\mathbf{f}}_n$ are $(n\times 1)$ dimensional vectors.

3 Asymptotic Results

In this section, I study the asymptotic properties of $[\hat{\beta}_0\;\hat{\beta}_1']'$ and $\tilde{h}_d(X_d)$. To do so, I first define vector and matrix norms that I will use throughout this paper. For $m,n\in\mathbb{N}$, let A be a matrix of order $(m\times n)$, r be a random vector of order $(m\times 1)$ whose elements have finite second moments, and c be a non stochastic vector of order $(m\times 1)$. Define $||A|| = \sqrt{trace(AA')}$, the Frobenius matrix norm, define the matrix spectral norm $||A||_{sp} = \max_{1\leq i\leq m}\lambda_i(AA')^{1/2}$ where $\{\lambda_i(AA')\}_{i=1}^m$ the collection of all eigenvalues of AA'. Define the inner product $||r||_2^2 = E(r'r)$, and lastly define $||c||_E = (c'c)^{1/2}$ the typical Euclidean norm in finite dimensional space $\mathbb{R}^m$.

3.1 Assumptions

Assumption A1: *(i)* $l_n^3/n = o(1)$, and $k\geq 3$, *(ii)* Define $Q_{BB} = E\left[\mathbf{B}_n(W_i)\mathbf{B}_n(W_i)'\right]$, $Q_{nBB} = n^{-1}\sum_{i=1}^n \mathbf{B}_n(W_i)\mathbf{B}_n(W_i)'$, $Q_{BBV}^d = E\left[\mathbf{B}_n(W_i)\mathbf{B}_n(W_i)'V_{di}^2\right]$, $Q_{nBBV}^d = n^{-1}\sum_{i=1}^n \mathbf{B}_n(W_i)\mathbf{B}_n(W_i)'V_{di}^2$, $Q_{a,BB} = E\big[\mathbb{B}_n(W_{ai})\mathbb{B}_n(W_{ai})'E(e_1|Z_i,X_i,V_i)\big]$. Assume that for n sufficiently large, there exist constants $c_{Bl}, c_{Bh}, c_{BV}\in\mathbb{R}^+$ such that

$$0 < c_{Bl} \leq \lambda_{min}(Q_{BB}) \leq \lambda_{max}(Q_{BB}) \leq c_{Bh} < \infty \qquad \text{and} \qquad 0 < \lambda_{max}(Q_{BBV}^d) \leq c_{BV} < \infty$$

and $\lambda_{max}(Q_{a,BB}) \leq c_{a,Bh} < \infty$ uniformly in n.

(iii) $m_{da}(\cdot)$ and $\{h_d(\cdot), f_d(\cdot)\}$ are k and a smooth functions respectively, such that $k = r_1+m_1$ and $a = r_2+m_2$

where for $j \in \{1, 2\}$ the m_jth derivative of each function satisfies a Hölder condition with exponent r_j.

Remarks: Under these assumptions, by Theorem (6) of de Boor (2001, pg. 149), for every n, there exists a parameter vectors $\alpha^d_{l_n} \in \mathbb{R}^{l_n+2k}$and function $m^{l_n}_d(W_a) = \mathbf{B}_n(W_a)'\alpha^{l_n}_d$ such that, $\sup_{W_a \in G_{W_a}} |m_d(W_a) - m^{l_n}_d(W_a)| = O\left(l_n^{-k}\right)$. Additionally, it should be clear that for any finite k, $B_{jk}(W_a)$ has shrinking compact support. In particular,

$$B_{j,k}(W_a) \begin{cases} > 0 & \text{if } W_a \in [t_j, t_{j+k}), \\ = 0 & \text{if } W_a \notin [t_j, t_{j+k}). \end{cases}$$

Assumption A2: For all, $j \in 0, 1, 2, 3$ (i) $K_j(\gamma) : \mathbb{R} \to \mathbb{R}$ is 4 times continuously differentiable, (ii) $K_j(\gamma)$ is symmetric about zero, (iii) $K_j(\gamma)$ is a kernel of order ν_j where $\int K_j(\gamma)\gamma^\alpha d\gamma = 0$ for all $\alpha = 1, 2, \dots, \nu_j - 1$ and $\int |K_j(\gamma)||\gamma|^{\nu_j} d\gamma < \infty$, (iv) $\int K_j(\gamma) d\gamma = 1$, (v) $|K^{(\alpha)}_j(\gamma)||\gamma|^{5+a} \to 0$ as $|\gamma| \to \infty$ for some $a > 0$, where $K^{(\alpha)}_j(\gamma)$ is the αth derivative of K_j, (vi) $\int |K^{(1)}_j(\gamma)||\gamma|^{\nu_j} d\gamma$, (vii) for $2 \le m \le \nu_j - 1$, $\int K^{(1)}_j(\gamma)^2\gamma^{2m} d\gamma < \infty$. $(viii)$ Let $\gamma_1 = [\gamma_{11}\ \gamma_{12}\ \cdots\ \gamma_{1D}]' \in \mathbf{R}^D$, $\gamma_2 = [\gamma_{21}\ \gamma_{22}\ \cdots\ \gamma_{2D}]' \in \mathbf{R}^D$ and define,

$$K_3([\gamma'_1, \gamma'_2]') \equiv \prod_{d=1}^{D} K_3(\gamma_{1d}) \prod_{d=1}^{D} K_3(\gamma_{2d}).$$

(ix) $K_4(\gamma)$ is a univariate probability density function that is symmetric, bounded, has compact support G_{K_4}, for some $C \in \mathbb{R}$ and for any $\gamma_1, \gamma_2 \in G_{K_4}$, $|K_4(\gamma_1) - K_4(\gamma_2)| \le C|\gamma_1 - \gamma_2|$.

Assumption A3: (i) $\{Y_i, X_i, W_i\}^n_{i=1}$ is an i.i.d sequence of random vectors distributed as (Y, X, W) with finite first and second moments. (ii) Densities; $p(W)$, $p(X_d)$, $p(V_d)$, $p(X, V)$ and $p(X, V, W)$ exist and are bounded away from zero and infinity on their convex and compact supports. (iii) The parameter vector $[\ \beta_0\ \ \beta'_1\ ]'$ is an element of a bounded (w.r.t. the Euclidean norm) subset of $\mathbb{R}^{p+1}$.

Remarks: (ii) implies that ϕ_i, θ^d_{2i}, and θ^d_{2i} are uniformly bounded away from zero and infinity.

Assumption A4: Let $\alpha, b \in \mathbb{N}$ and define $\mathcal{F}_b$ as a class of real valued $f : \mathbb{R}^\alpha \to \mathbb{R}$ functions such that if $f \in \mathcal{F}_b$ then: f is everywhere $b \in \mathbb{N}$ times continuously (provided $\alpha > 1$) differentiable. f and all of its derivatives are uniformly bounded. $(i.)$ Let, $p(W), p(X_d), p(V_d) \in \mathcal{F}_{\nu_0}$, and $p(X, V), p(X, V, W) \in \mathcal{F}_{\nu_3}$. $(ii.)$ $H^d_1(Z_c)$, $H^d_2(Z_c)$, $H^d_1(Y)$, $H^d_2(Y)$, $E[Z_c|X, V]$, $E[Y|X, V] \in \mathcal{F}_{\nu_{12}}$, where $\nu_{12} = \max(\nu_1, \nu_2)$.

Remarks: (i) is necessary for a sufficiently fast uniform rate of convergence of the Rosenblatt kernel density estimators generated in step 2. (ii) is necessary for various Taylor expansions of those functions in the proofs of the Theorems and Lemmas.

Assumption A5: (i) $E(h_d(X_d)) = 0$, $E(f_d(V_d)) = 0$, and $\Sigma_1 \equiv E(\phi\zeta\zeta')$ is a positive definite matrix, (ii) $E(\zeta_c^2) < \infty$, $E(u^2|Z, X, V) = \sigma_u^2 < \infty$, (iii) The functions; $E(\rho_c^2|X_d)$, $E(\rho_c^2|V_d)$, $E([\eta_{1c}^d]^2|V_d)$, $E([\eta_{2c}^d]^2|X_d)$, $E(\phi^2\zeta_c^2|X_d)$, $E(\phi^2\zeta_c^2|V_d)$, and $E(\phi^2\zeta_c^2|W)$ are uniformly bounded on their compact supports.
Remarks: (i) Ensures identification of $[\beta_o \;\; \beta_1']'$. (ii) Provides upper bounds for conditional moments used throughout the proofs of Theorems 1-3 , along with providing for the uniform convergence of several terms in accordance with the results of Lemma 5.

Assumption A6: (i) $k_n^3/n = o(1)$, $a \in \mathbb{N}$, (ii) Define $Q_{PP} = E\big[\mathbf{P}_n(X_i, V_i)\mathbf{P}_n(X_i, V_i)'\big]$, $Q_{nPP} = n^{-1}\sum_{i=1}^n \mathbf{P}_n(X_i, V_i)\mathbf{P}_n(X_i, V_i)'$, and $\hat{Q}_{nPP} = n^{-1}\sum_{i=1}^n \mathbf{P}_n(X_i, \hat{V}_i)\mathbf{P}_n(X_i, \hat{V}_i)'$. Assume that for n sufficiently large, there exists constants, c_{PL}, c_{PU} such that

$$0 < c_{Pl} \leq \lambda_{min}(Q_{PP}) \leq \lambda_{max}(Q_{PP}) \leq c_{Ph} < \infty$$

Remarks: Assumption A6 is analogous to Assumption A1 but this time for the estimation of $h(X)$ and $f(V)$ rather that $m(W)$. As a result, under these assumptions, by Theorem (6) of de Boor (2001, pg.149) , for every n, there exists a parameter vectors $\gamma_{1d}^{k_n} \in \mathbb{R}^{k_n+2a}$ and $\gamma_{2d}^{k_n} \in \mathbb{R}^{k_n+2a}$ along with functions $h_d^{l_n}(X_d) = \mathbf{B}_n(X_d)'\gamma_{1d}^{k_n}$, and $f_d^{l_n}(V_d) = \mathbf{B}_n(V_d)'\gamma_{2d}^{k_n}$, such that, $\sup_{X_d \in G_{X_d}} |h_d(X_d) - h_d^{l_n}(X_d)| = O\left(k_n^{-a}\right)$, and $\sup_{V_d \in G_{V_d}} |f_d(V_d) - f_d^{l_n}(V_d)| = O\left(k_n^{-a}\right)$. Lastly, for notational convenience denote

$$\Gamma_n = [\; \gamma_{11}^{l_n'} \;\; \gamma_{12}^{l_n'} \;\; \cdots \;\; \gamma_{1D}^{l_n'} \;\; \gamma_{21}^{l_n'} \;\; \gamma_{22}^{l_n'} \;\; \cdots \;\; \gamma_{2D}^{l_n'} \;]'$$

Assumption A7: Let $a, b, c, e, f, C \in \mathbb{R}^+$, $\nu_0, \nu_1, \nu_2, \nu_3 \in \mathbb{N}$ and $k \geq 3$ as defined in Assumption A1. If,

$$l_n = Cn^a, \qquad b_1 = Cn^{-e}, \qquad b_2 = Cn^{-b}, \qquad h_0 = Cn^{-f}, \qquad h_1 = Cn^{-c}, \qquad h_2 = Cn^{-c}.$$

Then constants $\{a, b, c, e, f, \nu_0, \nu_1, \nu_2\}$ satisfy the following: (i) $1/4 > a > 1/2k$, (ii) $b < 1/4 - a$, (iii) $c < \min\Big(1/4(D+1), (2k-1)/(4k(D+1)), (3k-3)/16k\Big)$, (iv) $e < 1/2$, (v) $f < 1/2$, (vi) For some arbitrarily small $\varepsilon > 0$

$$\nu_0 = \inf\left\{x \in \mathbb{N} : x > 1/4f\right\},$$
$$\nu_1 = \inf\left\{x \in \mathbb{N} : x > 1/4e\right\},$$
$$\nu_2 = \inf\left\{x \in \mathbb{N} : x > 1/4b\right\},$$
$$\nu_3 = \inf\left\{x \in \mathbb{N} : x > \left(4\left[\min\left(\frac{1}{4(D+1)}, \frac{2k-1}{4k(D+1)}, \frac{3k-3}{16k}\right) - \varepsilon\right]\right)^{-1}\right\}.$$

(vii) $l_n = k_n$ and $\tau_n \to c_1 \in [0, \infty)$, where $\tau_n = \left[k_n^{3/2}\left([l_n/n]^{1/2} + l_n^{-k}\right)\right]^2$ *(viii)* $nh_4^3 \log(n) \to 0$, $nh_4 k_n^{-2a} \to 0$, $n^{-1}h_4 k_n^5 \to 0$.

Remarks : Assumptions (i) - $(viii)$, are sufficient conditions for the rates of convergence derived in Lemma 3. Note that the conditions on h_0, b_1, ν_0, and ν_1 are set to ensure that uniform rates of convergence of kernel estimators used in steps two through four are at least $o_p(n^{-1/4})$. These constitute minimum requirements. Thus, if possible, some constants may be changed to improve a upon a minimum rate of $o_p(n^{-1/4})$.

3.2 Theorems

For a collection of random variables $\{M_j\}_{j=1}^J$ let G_M be the cartesian product of their support. By Theorem 1 in Newey (1997), under the Assumption A1 of this paper,

$$\sup_{W_a \in G_{W_a}} |\hat{m}_{da}(W_a) - m_{da}(W_a)| = O_p\left(\frac{l_n}{\sqrt{n}} + l_n^{1/2-k}\right) \equiv O_p\left(L_n\right).$$

Consequently, since $\hat{V}_{di} - V_{di} = X_{di} - \hat{m}_d(W_i) - X_{di} + m_d(W_i) = \sum_{a=1}^q \left(\hat{m}_d(W_{ai}) - m_d(W_{ai})\right)$ I have,

$$\sup_{W \in G_W} |\hat{V}_d - V_d| \leq \sum_{a=1}^q \sup_{W_a \in G_{W_a}} |\hat{m}_{da}(W_a) - m_{da}(W_a)| = O_p(L_n). \tag{30}$$

For notational simplicity set,

$$M_{1n} \equiv \left[\frac{\log(n)}{nh_0}\right]^{1/2} + h_0^{\nu_0}, \qquad M_{2n} \equiv \left[\frac{\log(n)}{nh_1^D h_2^D}\right]^{1/2} + h_1^{\nu_3} + h_2^{\nu_3}, \qquad M_n \equiv M_{1n} + M_{2n}, \tag{31}$$
$$N_{1n} \equiv \left[\frac{\log(n)}{nb_1}\right]^{1/2} + b_1^{\nu_1}, \qquad N_{2n} \equiv \left[\frac{\log(n)}{nb_2}\right]^{1/2} + b_2^{\nu_2}, \qquad \mathcal{L}_{0n} \equiv L_n + M_n, \tag{32}$$

$$\mathcal{L}_{1n} \equiv L_n + M_n + N_{1n}, \qquad \mathcal{L}_{2n} \equiv L_n + M_n + N_{2n} + b_2^{-1}\left[\sqrt{\frac{l_n}{n}} + l_n^{-k}\right], \qquad \mathcal{L}_n \equiv \mathcal{L}_{1n} + \mathcal{L}_{2n}. \tag{33}$$

The following Theorem shows that the uniform rate of convergence of $\hat{p}(\hat{V}_{di})$ to $p(V_{di})$, is equal to minimum of the uniform rates of convergence of either $\hat{m}_d(W_{ai})$ to $m_d(W_{ai})$, or $\hat{p}(V_{di})$ to $p(V_{di})$. Likewise the uniform rate of convergence of $\hat{p}(X_i, \hat{V}_i)$ to $p(X_i, V_i)$ is equal to minimum of the uniform rates of convergence of either $\hat{m}_d(W_{ai})$ to $m_d(W_{ai})$, $O_p(L_n)$ or $\hat{p}(X_i, V_i)$ to $p(X_i, V_i)$, $O_p(M_{2n})$. Thus, clearly if both M_{1n} and M_{2n} dominate L_n the rate of convergence derived in Theorem 1 is the same as if $m(W)$ is a known function. Furthermore the same holds true for the uniform rate of convergence of density ratios $\hat{\phi}_i$, $\hat{\theta}^d_{1i}$, and $\hat{\theta}^d_{2i}$.

Theorem 1. *Under Assumptions A1 - A5,*

$$\sup_{X_{di} \in G_{X_d}} |\hat{p}(X_{di}) - p(X_{di})| = O_p(M_{1n}), \qquad \sup_{V_{di} \in G_{V_d}} |\hat{p}(\hat{V}_{di}) - p(V_{di})| = O_p(L_n + M_{1n}),$$

$$\sup_{X_i, V_i \in G_{X,V}} |\hat{p}(X_i, \hat{V}_i) - p(X_i, V_i)| = O_p(L_n + M_{2n}), \qquad \sup_{X_i, V_i \in G_{X,V}} |\hat{\phi}(X_i, \hat{V}_i) - \phi(X_i, V_i)| = O_p(\mathcal{L}_{0n}),$$

$$\sup_{X_i, V_i \in G_{X,V}} |\hat{\theta}(X_{-di}, \hat{V}_i) - \theta(X_{-di}, V_i)| = O_p(\mathcal{L}_{0n}), \qquad \sup_{X_i, V_i \in G_{X,V}} |\hat{\theta}(X_i, \hat{V}_{-di}) - \theta(X_i, V_{-di})| = O_p(\mathcal{L}_{0n}).$$

Theorem 2. *Under assumptions A1 - A7, $\hat{\mu}_Y - \mu_Y = O_p(\mathcal{L}_{0n})$, $\hat{\mu}(Z_c) - \mu(Z_c) = O_p(\mathcal{L}_{0n})$ and*

$$\sup_{X_{di} \in G_{X_d}} |\hat{H}^d_1(Y_i) - H^d_1(Y_i)| = O_p(\mathcal{L}_{1n}), \qquad \sup_{V_{di} \in G_{V_d}} |\hat{H}^d_2(Y_i) - H^d_2(Y_i)| = O_p(\mathcal{L}_{2n}),$$

$$\sup_{X_{di} \in G_{X_d}} |\hat{H}^d_1(Z_{ci}) - H^d_1(Z_{ci})| = O_p(\mathcal{L}_{1n}), \qquad \sup_{V_{di} \in G_{V_d}} |\hat{H}^d_2(Z_{ci}) - H^d_2(Z_{ci})| = O_p(\mathcal{L}_{2n}),$$

$$\sup_{X_i, V_i \in G_{X,V}} |\hat{H}^*(Y_i) - H^*(Y_i)| = O_p(\mathcal{L}_n), \qquad \sup_{X_i, V_i \in G_{X,V}} |\hat{H}^*(Z_{ci}) - H^*(Z_{ci})| = O_p(\mathcal{L}_n).$$

Remarks : The uniform rates of convergence of these estimators are the sum of the uniform rates of convergence of $\hat{m}_d(W_{ai})$, $\hat{\theta}(X_{-di}, \hat{V}_i)$ or $\hat{\theta}(X_i, \hat{V}_{-di})$, and a typical NW estimator, in each case except for those terms where the conditional expectation is taken with respect to (w.r.t) V. In this case, there is an additional term $b_2^{-1}\left([l_n/n]^{1/2} + l_n^{-k}\right)$ which results from having to take a Taylor expansion of the kernel $K_2(\cdot)$ evaluated at $V_{dl} - V_{di}$. This additional term does slow down the rate of convergence. However, all that is required is $\mathcal{L}_{0n} = o(n^{-1/4})$ which, as shown in Lemma 3, is accomplished under Assumption 7 . In the following I present the two main theorems of this paper 3 and 5,

Theorem 3. *Under assumptions A1 - A7, we have* $\hat{\beta}_0 - \beta_0 = O_p(n^{-1/2})$ *and*

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} N(0, \Sigma_0^{-1}\Sigma_1\Sigma_0^{-1}),$$

where matrices Σ_0, Σ_1 *have typical elements* $\Sigma_0(a,c) = E\left[(Z_{ai} - H^*(Z_{ai}))\phi(X_i, V_i)(Z_{ci} - H^*(Z_{ci}))\right]$*, and* $\Sigma_1(a,c) = E\left[(Z_{ai} - H^*(Z_{ai}))\phi(X_i, V_i)^2(Z_{ci} - H^*(Z_{ci}))\right]\sigma_u^2$ *where* $a, c \in \{1, 2, \ldots, p\}$

Remarks: Note that $\hat{\beta}_1$ is $\sqrt{n}$ asymptotically normal with a covariance matrix equal to an identical estimation procedure where $m(W)$ is a known function. This clearly indicates that asymptotically there is no penalty, in terms of variance, to estimating $m(W)$ rather than observing it. Furthermore, the asymptotic variance of $\hat{\beta}_1$ is the same as the asymptotic variance derived by Manzan and Zerom (2005) which, as they point out in the case of homoskedastic errors, meets the efficiency bound derived by Chamberlain (1992) for their model, where $m(X)$ is observed. However I can make no claim to efficiency for my model since such an efficiency bound does not currently exist in the statistical literature. Furthermore, there exists an easy to implement estimator for both Σ_0, Σ_1 using the method of moments framework.

$$\hat{\Sigma}_0(a,c) = n^{-1}\sum_{i=1}^{n}(Z_{ai} - \hat{H}^*(Z_{ai}))\phi(X_i, \hat{V}_i)(Z_{ci} - \hat{H}^*(Z_{ci}))$$
$$\hat{\Sigma}_1(a,c) = n^{-1}\sum_{i=1}^{n}(Z_{ai} - \hat{H}^*(Z_{ai}))\phi(X_i, \hat{V}_i)^2(Z_{ci} - \hat{H}^*(Z_{ci}))\hat{\sigma}_u^2$$

where $\hat{\sigma}_u^2$ is the average square of the residuals generated after all backfitting steps have been conducted.

Theorem 4. *Under assumptions A1 - A7,*

i.) $||\hat{\Gamma}_n - \Gamma_n||_E = O_p\left(\sqrt{\frac{l_n}{n}} + \sqrt{\frac{k_n}{n}} + l_n^{-k} + k_n^{-a}\right)$

ii.) $\sup_{XV \in G_{XV}} |h(X) + f(V) - \hat{h}(X) - \hat{f}(V)| = O_p\left(k_n^{1/2}\left[\sqrt{\frac{l_n}{n}} + \sqrt{\frac{k_n}{n}} + l_n^{-k} + k_n^{-a}\right]\right)$

Remarks: The results of this theorem are the consequence of plugging $\hat{\beta}_1$ and $\hat{\beta}_0$ into a B-spline series for $h(X) + f(V)$ then showing that in theorem 3.1 of Ozabaci et al. (2013) the resulting additional components are $O_p(n^{-1/2})$.

Theorem 5. *Under assumptions A1 - A7,*

$$\sqrt{nh_4}H\Big([\ \tilde{h}_d(x_d)\ \ \tilde{h}_d^{(1)}(x_d)\ ]' - [\ h_d(x_d)\ \ h_d^{(1)}(x_d)\ ]' - [\ 1/2h_d^{(2)}(x_d)h_4^2\int u^2K_4(u)du\ \ 0\ ]'\Big) \xrightarrow{d} N(0, \Omega(x_d))$$

where

$$\Omega(x_d) = diag(\{E(u_i^2|X_{di} = x_d)p(x_d)^{-1}, E(u_i^2|X_{di} = x_d) \int u^2 K_4(u)^2 du \big[p(x_d) \int u^2 K_4(u) du\big]^{-1}\})$$

Remarks: Similarly, the results of this theorem are the consequence of plugging $\hat{\beta}_1$ and $\hat{\beta}_0$ into a 1 step kernel back fitting estimator for $h_d(x_d)$ then showing that in theorem 3.2 of Ozabaci et al. (2013) the resulting additional components are $O_p(n^{-1/2})$.

4 Monte Carlo Study

In this section I investigate the finite sample properties of the estimators developed in this paper. First I demonstrate the finite sample performance of $\hat{\beta}_1$ and compare it to the finite sample performance of both $\hat{\beta}_{1,G}$, and $\hat{\beta}_{1,M}$ the estimator presented in Geng et al. (2016), and the estimator of Manzan and Zerom (2005) where $m(W)$ is taken to be a known vector of functions, respectively. Secondly, I demonstrate the finite sample properties of the nonparametric estimators for $h_1(X_1)$ and $h_2(X_2)$. I illustrate the differences between these estimators in terms of two data generating processes DGP $\in \{1, 2\}$, three values for a parameter $\theta_e \in \{0.3, 0.6, 0.9\}$ indicating an increasing strength of the endogenous relationship between X and ε, and two sample sizes $n \in \{200, 400\}$. Consider the following DGPs whose leading non nonlinear terms are the same, at least in part, as the functions used in, Ai and Chen (2003), Su and Ullah (2008), Martins-filho and Yao (2012), and Geng et al. (2016),

$$\begin{aligned}
\text{DGP} = 1: \quad & Y_i = \beta_0 + Z_{1i}\beta_1 + \log(\big|X_{1i} - 1\big| + 1)\text{sgn}(X_{1i} - 1) - \left[\frac{X_{2i}}{3}\right]^2 + \varepsilon_i, \\
\text{DGP} = 2: \quad & Y_i = \beta_0 + Z_{1i}\beta_1 + \frac{\exp(X_{1i})}{1 + 3\exp(X_{1i})} + \left[\frac{X_{2i}}{3}\right]^2 + \varepsilon_i,
\end{aligned}$$

where I set $\beta_0 = 1$ and $\beta_1 = 10$. Both DGPs have identical secondary equations,

$$X_{1i} = 1/2Z_{1i}^2 - 1/2Z_{2i}^2 - 2Z_{3i} + V_{1i}, \quad \text{and} \quad X_{2i} = 3/2Z_{1i} + 1/2Z_{2i}^2 + 1/3Z_{3i}^3 + V_{2i},$$

where $\{W_i\}_{i=1}^n = \{Z_{1i}, Z_{2i}, Z_{3i}\}_{i=1}^n$ is an i.i.d sequence with distribution $W \sim N(0, I_3)$ and $\{\varepsilon_i, V_{1i}, V_{2i}\}_{i=1}^n$ is an i.i.d sequence with distribution,

$$\begin{pmatrix} \varepsilon_i \\ V_{1i} \\ V_{2i} \end{pmatrix} \sim N\left(0, \begin{pmatrix} 1 & \theta_e & \theta_e \\ \theta_e & 1 & \theta_e^2 \\ \theta_e & \theta_e^2 & 1 \end{pmatrix}\right).$$

To implement estimators $\hat{\beta}_1$, $\hat{\beta}_{1,M}$ and $\hat{\beta}_{1,G}$, I use 3rd order B splines for my estimator $\hat{m}(W)$ with knot sequences chosen according to assumption 1, 6th order gaussian kernels and rule of thumb bandwidths for Nadaraya Watson estimator $\hat{m}_G(W)$, I use 2nd order gaussian kernels and rule of thumb bandwidths for all required density estimators in step 2, and I use 6th order gaussian kernels and cross validated bandwidths for all Nadaraya Watson type estimators in step 3. Data for this study was generated using 500 iterations for each of the 12 unique combinations of $(\text{DGP}, \theta_e, n)$.

Table 1: Finite Sample Performance of $\hat{\beta}_1$

	$\theta_e = 0.3$				$\theta_e = 0.6$				$\theta_e = 0.9$			
	B	S	R	D	B	S	R	D	B	S	R	D
DGP=1	n=200											
$\hat{\beta}_1$	-0.1018	0.1463	0.0317	0.0165	-0.2327	0.1790	0.0860	0.0612	-0.2570	0.1690	0.0945	0.0682
$\hat{\beta}_{1,G}$	-0.2630	0.2411	0.1270	0.0892	-0.4229	0.2590	0.2455	0.1827	-0.4768	0.2567	0.2929	0.2301
$\hat{\beta}_{1,M}$	-0.0851	0.1348	0.0253	0.0120	-0.2103	0.1480	0.0660	0.0463	-0.2196	0.1436	0.0687	0.0519
	n=400											
$\hat{\beta}_1$	-0.0849	0.1019	0.0175	0.0085	-0.1942	0.1076	0.0492	0.0311	-0.2422	0.1071	0.0701	0.0564
$\hat{\beta}_{1,G}$	-0.2016	0.1694	0.0692	0.0516	-0.3135	0.1567	0.1227	0.1023	-0.3481	0.1754	0.1518	0.1210
$\hat{\beta}_{1,M}$	-0.0759	0.0898	0.0138	0.0076	-0.1875	0.0889	0.0430	0.0340	-0.2260	0.0939	0.0598	0.0566
DGP=2	n=200											
$\hat{\beta}_1$	-0.1460	0.1737	0.0513	0.0296	-0.2361	0.1867	0.0904	0.0597	-0.2845	0.1674	0.1088	0.0733
$\hat{\beta}_{1,G}$	-0.1969	0.2011	0.0790	0.0460	-0.3296	0.2237	0.1584	0.1216	-0.3977	0.2529	0.2218	0.1921
$\hat{\beta}_{1,M}$	-0.1186	0.1887	0.0495	0.0251	-0.1966	0.1979	0.0776	0.0423	-0.2352	0.1754	0.0859	0.0589
	n=400											
$\hat{\beta}_1$	-0.1369	0.0964	0.0280	0.0190	-0.2532	0.1031	0.0747	0.0606	-0.2690	0.1103	0.0844	0.0749
$\hat{\beta}_{1,G}$	-0.1487	0.1682	0.0503	0.0297	-0.2998	0.1898	0.1257	0.1020	-0.3046	0.2163	0.1393	0.0999
$\hat{\beta}_{1,M}$	-0.0851	0.1348	0.0253	0.0120	-0.2103	0.1480	0.0660	0.0463	-0.2196	0.1436	0.0687	0.0519

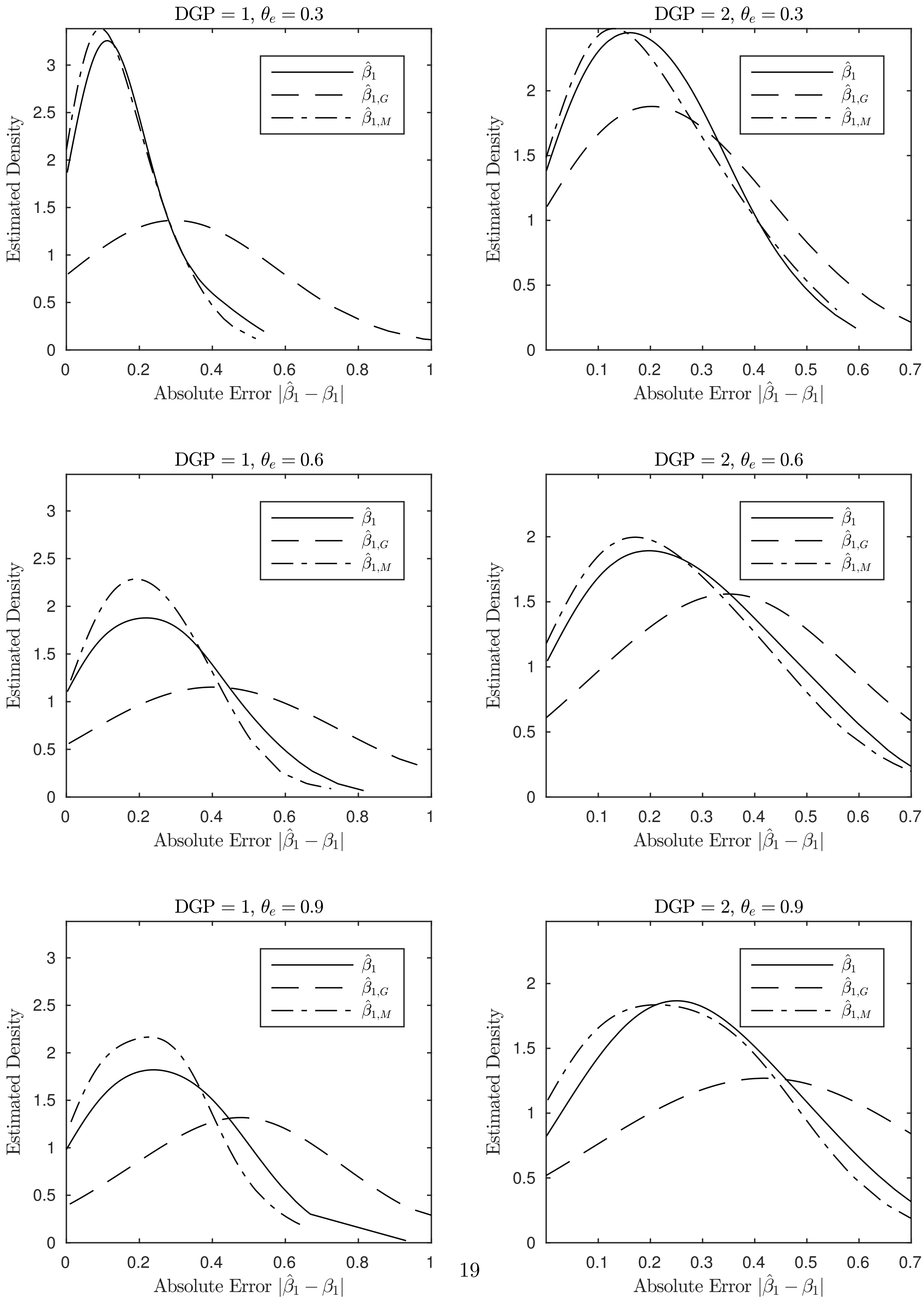


Figure 1: Estimated Densities for Absolute Error of estimators of β_1, $n = 200$

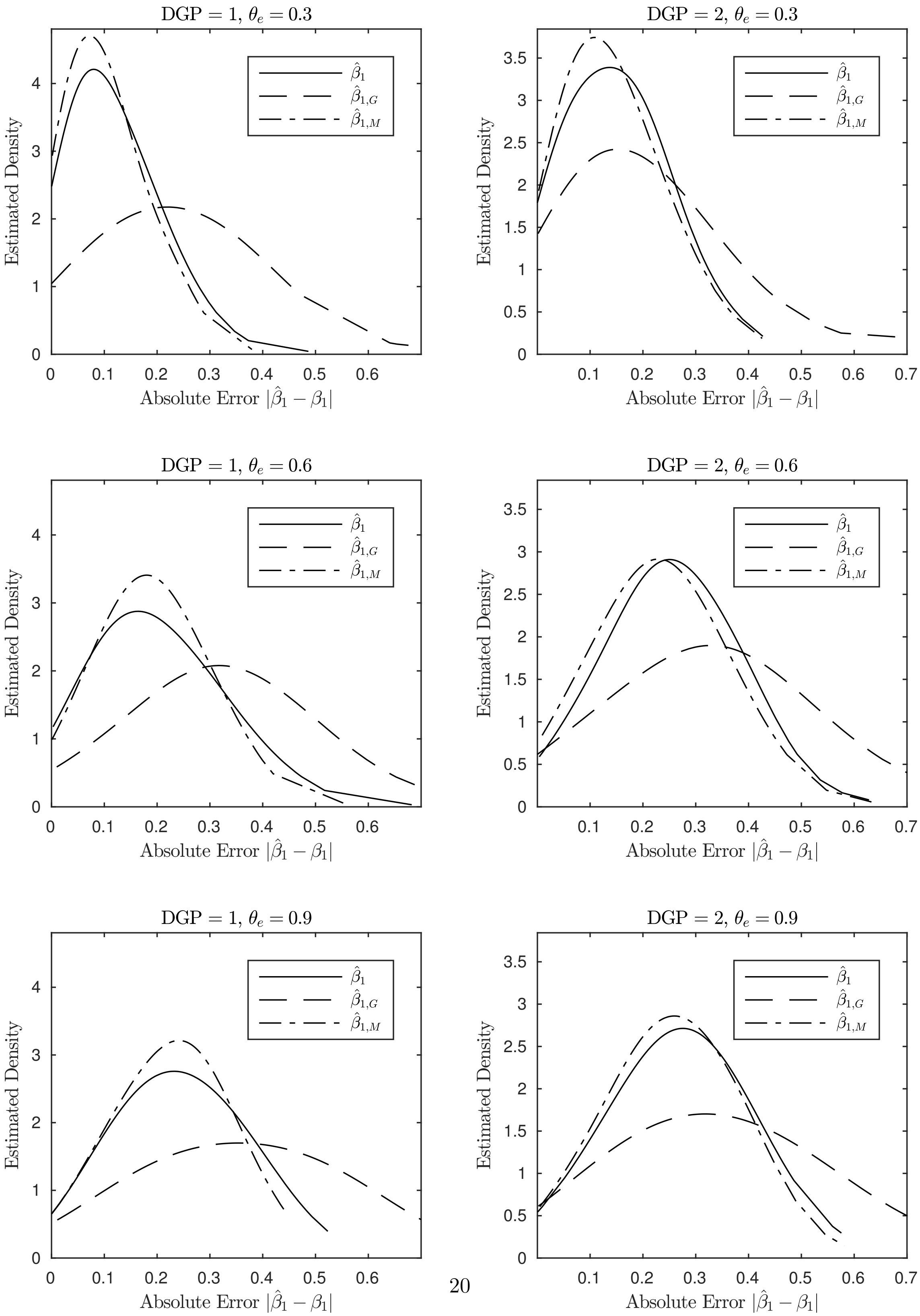


Figure 2: Estimated Densities for Absolute Error of estimators of β_1, $n = 400$

The results of this Monte Carlo exercise pertaining to the estimation of $\hat{\beta}_1$ are summarized in Table 1, which lists the bias (B), the standard deviation (S), the root mean squared error (R), and the median squared error (D), for each estimator. Note, that the Table and Figures summarize only those estimated values within the 2% - 98% sample quantile range. In addition to the results summarized in Table 1, Rosenblatt kernel density estimates for the absolute error of all three estimators are shown in Figure 1 for a sample size of $n = 200$, and in Figure 2 for a sample size of $n = 400$.

The additive model presented in the paper is a restriction on the model developed in Geng et al. (2016) who assume no additivity. As a result one should expect that the estimator in this paper will have better finite sample performance for two reasons. Firstly, since both primary equations in this study are additive functions, the estimator in this paper should perform better than Geng et al. (2016) regardless of the degree of endogeneity or sample size as it is has a sharper specification. The structure of this Monte Carlo exercise is designed to demonstrate changes in the finite sample properties of each estimator due to variations in data generating process, sample size, and degree of endogeneity. As for variations in sample size, in Table 1 it is clear that increasing the sample size has the expected effect on the estimation of β_1 since, in all cases, the bias, standard deviation, and mean squared error of each estimator decline, supporting the consistency results derived for each. As the level of endogeneity increases one should expect that the two disadvantages inherent in Geng et al. (2016) discussed above should be compounded since comparatively poorer estimates for the secondary equation $\hat{m}_G(W)$, due to a more general specification for $m(W)$, are used in a second stage estimator $\hat{\beta}_{1,G}$ which is derived using another more general specification. In Table 1 and Figures 1 and 2, one can see that this is the case as the negative effect that increasing levels of endogeneity have on all estimates is disproportionately large on $\hat{\beta}_{1,G}$.

Recall that the sole difference between estimators $\hat{\beta}_1$ and $\hat{\beta}_{1,M}$ is whether $m(W)$ is a known or estimated. In Theorem 3, I have shown that the asymptotic variance of $\hat{\beta}_1$ is identical to $\hat{\beta}_{1,M}$, so the purpose of comparing $\hat{\beta}_1$ and $\hat{\beta}_{1,M}$ here is to determine to what extent this oracle property is evident in finite samples. The findings presented in Table 1, and both Figures reinforce the oracle efficiency result derived in Theorem 3 as one can clearly see that the density of the absolute error of $\hat{\beta}_1$, and $\hat{\beta}_{1,M}$ are practically coincident in 9 of the 12 panels and the tabulated results are likewise very similar. This is in some ways surprising as one would not expect the asymptotic oracle efficiency of $\hat{\beta}_1$ to so clearly manifest itself in finite samples. These results show that not only is there asymptotically no penalty to estimating $m(W)$ but there is also very little penalty in estimating $m(W)$ in finite samples.

To implement estimators for the unknown nonlinear functions of X_1 and X_2, in Table 2 they are labeled

$h_1(X_1)$ and $h_2(X_2)$ respectively, I used 3rd order B-Splines for the initial series estimation and 2nd order gaussian kernels and rule of thumb bandwidths for the final back fitting estimation step. The result of the Monte Carlo study for $h_1(X_1)$ and $h_2(X_2)$ are listed in Table 2, which lists the average absolute bias (B), and average root mean squared error (R) for both h and its first derivative. As seen in Table 2 an increase in the sample size from $n = 200$ to $n = 400$ in all cases results in a reduction of the bias and root mean squared error of the estimator for h supporting the results of Theorems 4 and 5. Lastly, as the level of endogeneity θ_e increases Table 2 shows the bias and root mean squared error of h increases mildly, matching the same general trend as in the estimation of β_1.

Table 2: Finite Sample Performance of $\hat{h}_1(X_2)$ and $\hat{h}_2(X_2)$

	$\theta_e = 0.3$				$\theta_e = 0.6$				$\theta_e = 0.9$			
	$h_j(X_j)$		$\frac{dh(X_j)}{dX_j}$		$h_j(X_j)$		$\frac{dh(X_j)}{dx_j}$		$h_j(X_j)$		$\frac{dh(X_j)}{dX_j}$	
	B	R	B	R	B	R	B	R	B	R	B	R
DGP=1	n=200											
$\hat{h}_1(X_1)$	0.1645	0.2016	0.7364	0.8223	0.1783	0.2187	0.7224	0.8048	0.1741	0.2182	0.7137	0.7989
$\hat{h}_2(X_2)$	0.1875	0.2201	0.1570	0.1907	0.2674	0.3141	0.2047	0.2419	0.3100	0.3628	0.2322	0.2702
	n=400											
$\hat{h}_1(X_1)$	0.1307	0.1590	0.7414	0.8234	0.1311	0.1616	0.7301	0.8110	0.1429	0.1777	0.7089	0.7915
$\hat{h}_2(X_2)$	0.1437	0.1721	0.1269	0.1571	0.2330	0.2749	0.1707	0.2019	0.2776	0.3247	0.1923	0.2229
DGP=2	n=200											
$\hat{h}_1(X_1)$	0.1422	0.1784	0.1760	0.2064	0.1511	0.1851	0.1785	0.2083	0.1645	0.2021	0.1859	0.2177
$\hat{h}_2(X_2)$	0.2253	0.2637	0.1909	0.2269	0.3122	0.3641	0.2315	0.2687	0.3430	0.3987	0.2493	0.2872
	n=400											
$\hat{h}_1(X_1)$	0.1073	0.1329	0.1766	0.2038	0.1231	0.1479	0.1748	0.2014	0.1191	0.1436	0.1771	0.2028
$\hat{h}_2(X_2)$	0.1884	0.2239	0.1475	0.1810	0.2928	0.3413	0.2000	0.2324	0.3207	0.3748	0.2163	0.2502

5 Stability

The stability of the statistical inference conducted on an estimator relative to small changes to sample data is a key element of reproducibility. With a stable estimator, small modifications to the sample data will be accompanied by similarly small changes in the value of a test statistic used to conduct a hypothesis test or construct a confidence interval. The implication is that the results of such an estimator are less likely to be driven by idiosyncratic features of a particular sample, present by chance or by design, and more likely to be representative of features common to the entire population. As a result, when evaluating the credibility of an empirical study one should take into account the stability of the estimator used to generate the results i.e. the more stable the estimator the more credible the conclusions. Accordingly it is incumbent upon those who develop new estimators to demonstrate their stability to relevant changes in sample data. Recently, as the sciences continue to come to grips with the reproducibility crisis, there has been a growing interest in estimator stability as a number of papers have been developed studying the sensitivity of established estimators to sample changes such as subset selection and outlier removal. Here, I contribute to this literature by examining the effect that changes in the quality of regressors generated in the initial step have on the statistical properties of $\hat{\beta}_1$. To fit this demonstration within the framework of stability one can think of $\hat{\beta}_1$ as a statistic depending on sample $\{Y_i, Z_i, X_i, \hat{V}_i\}_{i=1}^n$, whose "quality", depends on the procedure and regressors used to generate $\{\hat{V}_i\}_{i=1}^n$. In particular, I demonstrate the stability of $\hat{\beta}_1$, by simulating the effect that weak instruments and overfitting in the first stage regression have on the type 1 error rate of the following hypothesis test $H_0 : \beta_1 = 5$ vs. $H_1 : \beta_1 \neq 5$. In this model the presence of weak instruments alters the estimation of $\hat{V}_i$ by allocating some of the exogenous variation in $\{X_i\}_{i=1}^n$ to $\{\hat{V}_i\}_{i=1}^n$, effectively masking some the variation needed to identify and estimate β_1. Note, this is the opposite effect that weak instruments has on IV estimators. In this study I simulate the presence of weak instruments in the original sense of having low but not zero correlation with X by estimating $\hat{V}$ using a convex combination $\{W_{\alpha i}\}_{i=1}^n$ of the regressors, $\{W_i\}_{i=1}^n$, used to generate $\{X_i\}_{i=1}^n$ and "noise" $\{\tilde{W}_i\}_{i=1}^n$ generated i.i.d from the same distribution as W and independent of X i.e.

$$W_{\alpha,i} = \alpha W_i + (1-\alpha)\tilde{W}_i \quad \text{where} \quad W_{ji}, \tilde{W}_{ji} \sim U[-1,1] \quad \text{for } j \in \{1,2,3\} \quad \text{and} \quad \alpha \in (0,1]$$

so that $cov(X_i, W_{\alpha,i}) = \alpha \cdot cov(X, W_i)$. This convex combination allows one to decrease the covariance between $W_{\alpha,i}$ and X_i while preserving the location of the regressors and incurring a variance reduction of at most 50%, when $\alpha = 0.5$. In this simulation I use $\alpha \in \{0.8, 0.6\}$ to represent mildly weak instruments and $\alpha \in \{0.4, 0.2\}$ to represent severely weak instruments. Alternatively, overfitting causes the endogenous variation in X_i to be only partially represented in estimates $\hat{V}_i$, effectively leaving out a portion of the variation needed for the estimation of β_1. The potential for overfitting depends critically on the relationship between the VC dimension of the space of functions in which a regression is being estimated and the number of observations used. In short the VC dimension of a space of functions is the maximal number of vectors which can be *shattered* by a collection of indicators functions indexed by the space of functions. A treatment of VC theory can be found in Vapnik (1998) but for time being one can think of the VC dimension of a space of functions as the ability of those functions to approximate an arbitrary set of n points in euclidean space. In other words, as the VC dimension increases, the euclidean distance between a set of points and it closest approximation in the space of function decreases. Accordingly, as the VC dimension of the space functions used to approximate $m_j(W_j)$ increases one expects that $||\hat{V}_{ji}||_{E,n} = ||X_{ji} - \hat{m}_j(W_{ji})||_E \to 0$. The VC dimension of a set of linear in parameter functions is equal to its dimension, thus I will simulate overfitting by increasing the number of basis functions used in the estimation of each additive secondary regression function by increasing the number of interior knots used. By Assumption 7(i) the number of knots can, at most, be proportional to epsilon less than $Cn^{1/4}$. Supposing that we use the constant $C = 1.25$ then for $n = 250$ the number of interior knots used should be strictly less that 4.97. As a result, in the following simulation I use 4 internal knots as a baseline for all weak instrument runs. In the overfitting runs I use the following set of internal knots $\in \{1, 10, 15, 20, 25\}$ where 1 internal knot is a four fold decrease in the dimension of the space i.e. under fitting, and 25 knots is a more than six fold increase in the dimension of the spline space.

The basic setup of this simulation is the same as the one conducted in section 4 except for a few changes. First, to highlight the stability of $\hat{\beta}_1$ in a situation where it is more likely to be unstable while also providing enough degrees of freedom to justify a nonparametric estimator, I use a sample size of $n = 250$ for each trial. This choice is not without consequence since during the course of this simulation study I found that the estimator for the standard errors, S_{bf} suggested by theorem 3 resulted in a 25% underestimate. These underestimates inflated type 1 error rates by between 5 and 15%, as a result I used bootstrapped standard errors S_{bs} for the following calculations. Second, since the focus is on the properties of $\hat{\beta}_1$ relative to changes in the secondary equations I use only the primary equation labeled DGP 2 in section 4. Note that in table 1 the bias of $\hat{\beta}_1$ under DGP 2 is larger in both cases. Third, in order to demonstrate results which are robust to the specification of both secondary equations I constructed 3 cases for which I randomly chose a set of secondary equations from a collection of cubic polynomials whose coefficients are unit vectors. Lastly, the parameter of interest here is the estimated type 1 error rate for each case, endogeneity level, and α value or number of internal knots. As a result, for each estimated $\hat{\beta}_1$ I use the student t distribution to calculate a p value corresponding to the hypothesis test stated above. After all trials have been completed I calculate the proportion of estimates for which the null hypothesis would have been rejected at significance levels 0.1, 0.05, and 0.01. Here I present one of these cases, the remaining two can be found in the appendix.

Table 3: Estimated Type 1 Error Rate for $\hat{\beta}_1$ when $\theta_e = 0.3$

α	Knots	M	S_{mc}	$\hat{S}_{bs}$	$\hat{S}_{bf}$	0.1	Δ	0.05	Δ	0.01	Δ
1.00	4.00	-0.034	0.167	0.167	0.125	0.120	-	0.052	-	0.004	-
0.80	4.00	-0.027	0.174	0.167	0.125	0.112	-0.008	0.064	0.012	0.012	0.008
0.60	4.00	-0.029	0.166	0.173	0.127	0.108	-0.012	0.052	0.000	0.012	0.008
0.40	4.00	-0.030	0.162	0.178	0.130	0.092	-0.028	0.032	-0.020	0.008	0.004
0.20	4.00	-0.029	0.181	0.181	0.131	0.088	-0.032	0.052	0.000	0.008	0.004
1.00	1.00	-0.032	0.171	0.166	0.125	0.128	0.008	0.060	0.008	0.012	0.008
1.00	10.00	-0.020	0.167	0.167	0.124	0.124	0.004	0.044	-0.008	0.008	0.004
1.00	15.00	-0.025	0.153	0.171	0.124	0.072	-0.048	0.020	-0.032	0.000	-0.004
1.00	20.00	-0.032	0.156	0.173	0.126	0.068	-0.052	0.040	-0.012	0.004	0.000
1.00	25.00	-0.017	0.148	0.178	0.124	0.056	-0.064	0.016	-0.036	0.000	-0.004

Figure 3: Distribution of Absolute Error in Estimation and p-values

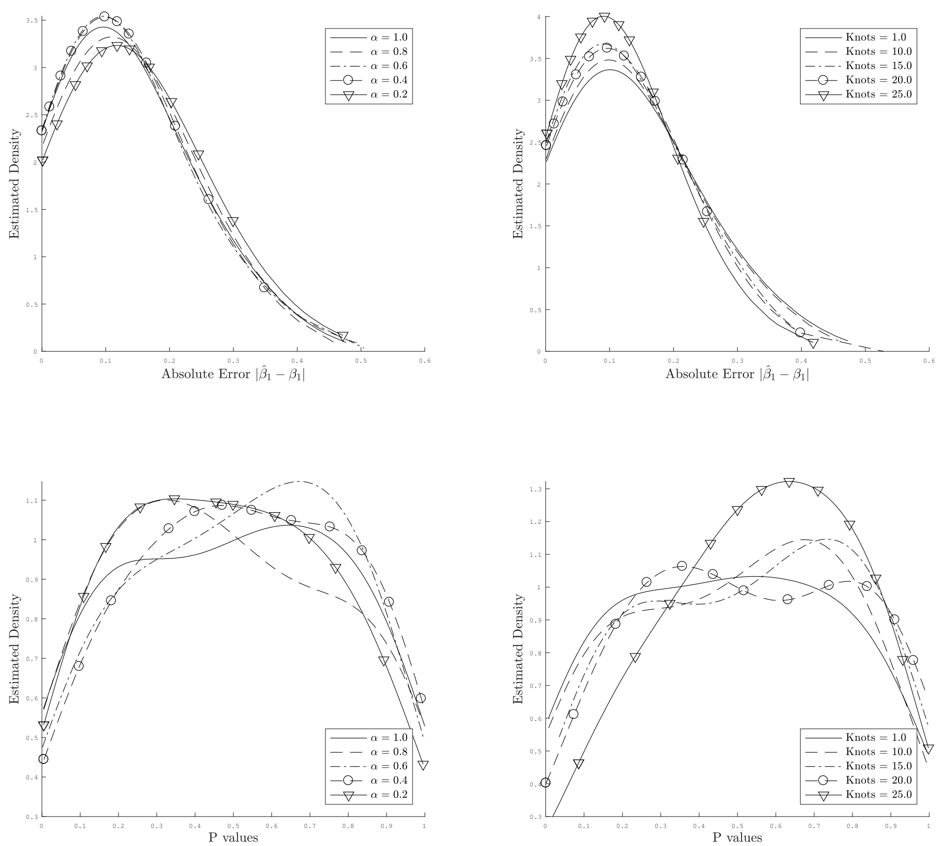


Table 4: Estimated Type 1 Error Rate for $\hat{\beta}_1$ when $\theta_e = 0.6$

α	Knots	M	S_{mc}	$\hat{S}_{bs}$	$\hat{S}_{bf}$	0.1	Δ	0.05	Δ	0.01	Δ
1.00	4.00	-0.050	0.170	0.172	0.125	0.100	-	0.056	-	0.020	-
0.80	4.00	-0.054	0.173	0.173	0.128	0.096	-0.004	0.044	-0.012	0.008	-0.012
0.60	4.00	-0.059	0.173	0.176	0.130	0.104	0.004	0.060	0.004	0.004	-0.016
0.40	4.00	-0.077	0.182	0.184	0.138	0.088	-0.012	0.052	-0.004	0.020	0.000
0.20	4.00	-0.054	0.194	0.189	0.135	0.100	0.000	0.060	0.004	0.020	0.000
1.00	1.00	-0.058	0.170	0.171	0.125	0.112	0.012	0.052	-0.004	0.016	-0.004
1.00	10.00	-0.035	0.180	0.174	0.125	0.112	0.012	0.068	0.012	0.012	-0.008
1.00	15.00	-0.063	0.159	0.175	0.126	0.080	-0.020	0.048	-0.008	0.004	-0.016
1.00	20.00	-0.060	0.165	0.177	0.124	0.100	0.000	0.040	-0.016	0.008	-0.012
1.00	25.00	-0.064	0.151	0.183	0.125	0.064	-0.036	0.028	-0.028	0.004	-0.016

Figure 4: Distribution of Absolute Error in Estimation and p-values

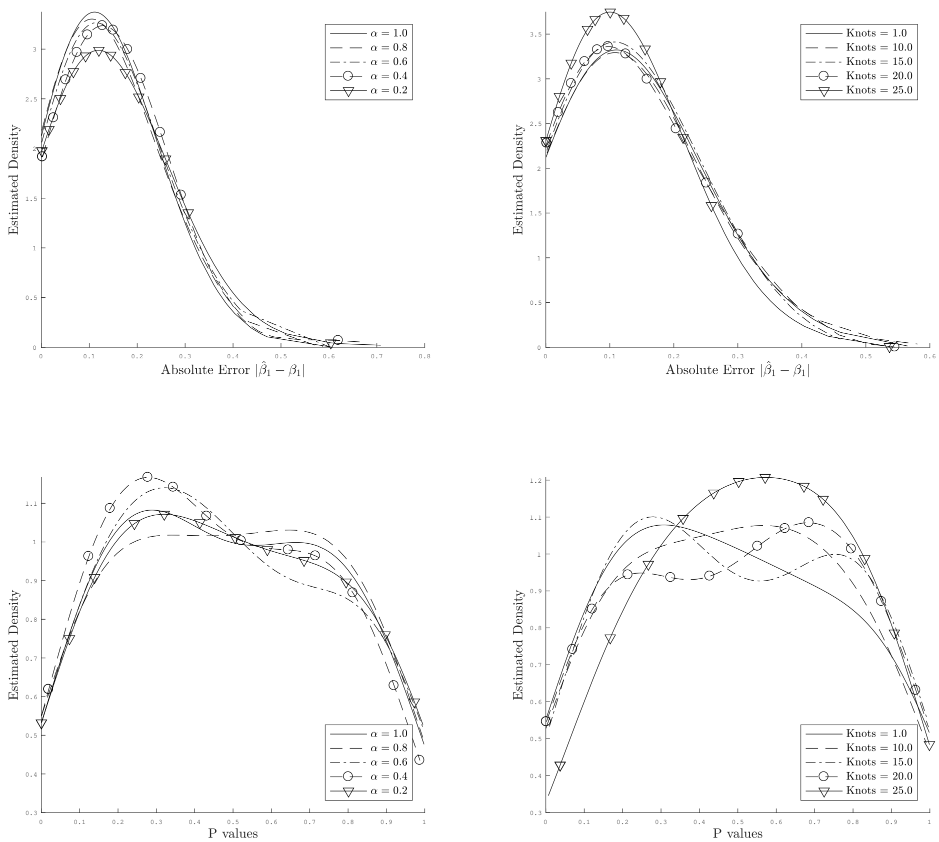
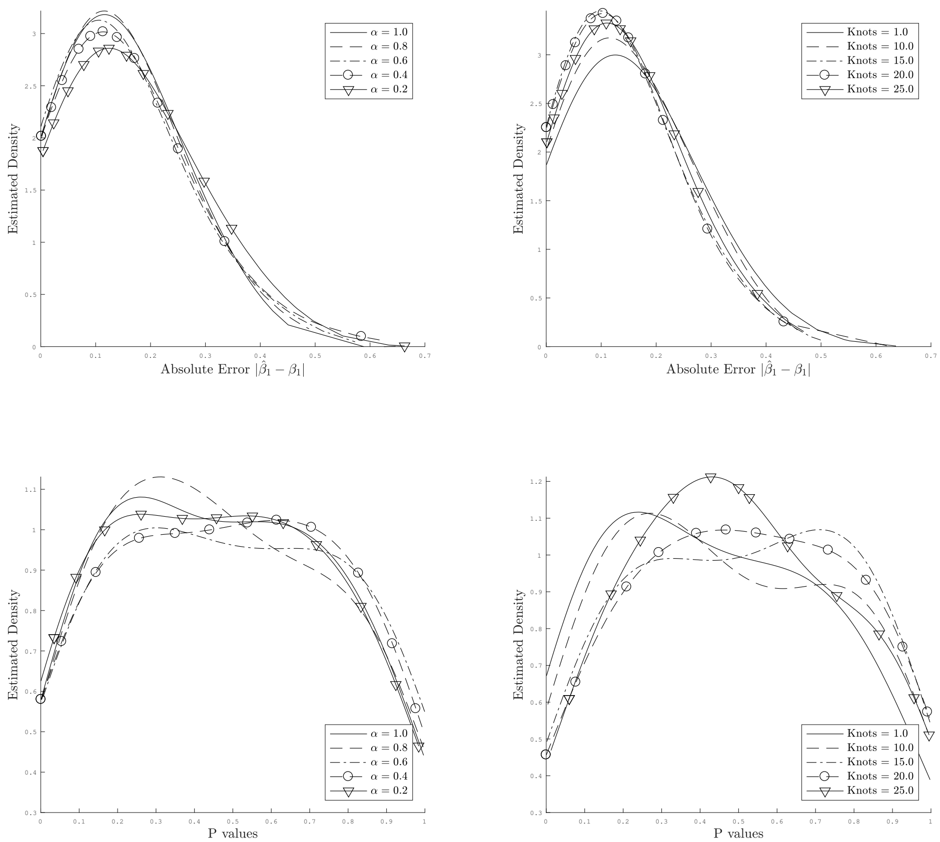


Table 5: Estimated Type 1 Error Rate for $\hat{\beta}_1$ when $\theta_e = 0.3$

α	Knots	M	S_{mc}	$\hat{S}_{bs}$	$\hat{S}_{bf}$	0.1	Δ	0.05	Δ	0.01	Δ
1.00	4.00	-0.055	0.179	0.177	0.125	0.116	-	0.060	-	0.008	-
0.80	4.00	-0.077	0.173	0.176	0.132	0.116	0.000	0.072	0.012	0.024	0.016
0.60	4.00	-0.068	0.179	0.181	0.133	0.140	0.024	0.076	0.016	0.012	0.004
0.40	4.00	-0.063	0.193	0.187	0.137	0.132	0.016	0.076	0.016	0.024	0.016
0.20	4.00	-0.070	0.199	0.193	0.144	0.148	0.032	0.060	0.000	0.008	0.000
1.00	1.00	-0.057	0.192	0.176	0.125	0.140	0.024	0.076	0.016	0.020	0.012
1.00	10.00	-0.066	0.170	0.177	0.126	0.120	0.004	0.044	-0.016	0.012	0.004
1.00	15.00	-0.067	0.161	0.179	0.126	0.108	-0.008	0.040	-0.020	0.008	0.000
1.00	20.00	-0.046	0.169	0.180	0.129	0.108	-0.008	0.068	0.008	0.004	-0.004
1.00	25.00	-0.075	0.164	0.183	0.126	0.088	-0.028	0.040	-0.020	0.000	-0.008

Figure 5: Distribution of Absolute Error in Estimation and p-values



5.1 Results

The results of this simulation are contained in Tables 3, 4, and 5 which lists the Bias (B) of $\hat{\beta}_1$, the sample standard deviation of the collection of estimates S_{mc}, the average standard deviation obtained by bootstrapping S_{bs}, the average standard deviation of the estimator suggested by Theorem 3, S_{bf}, the sample type 1 rejection rate for significance levels $\{0.1, 0.05, 0.01\}$, and the difference Δ from the baseline case in row one where $\alpha = 1$ and Knots = 4. Across all endogeneity levels, changes to the estimated type 1 error rate are quite small in response to moderate levels of weak instruments having a maximum absolute difference of 2.4% for the 10% threshold and $\theta_e = 0.9$. However, once instrument weakness becomes severe, changes in the type 1 error rate becomes more pronounced with a maximum absolute difference of 3.2% across all endogeneity levels. This suggests that $\hat{\beta}_1$ is stable to moderate levels of weak instruments and somewhat less stable as instrument weakness becomes more severe.

Similarly, in the presence of either underfitting or a 2.5 fold increase in the dimension of the approximating space the type 1 error rate is stable, with a maximum absolute difference of 1.6% across all endogneity levels. In contrast, once the number knots increases past 10 the estimator becomes less stable with a maximum absolute difference of 6.4%. Qualitatively one would expect that in the control function frame work overfitting would have a more pronounced effect on $\hat{\beta}_1$ since, like weak instruments in IV estimation the primary regression is being deprived of identifying variation rather than it being fully present but concealed. One can see that overall $\hat{\beta}_1$ is stable to moderate levels of both weak instruments and overiftting, and at more severe levels $\hat{\beta}_1$ is more sensitive to overfitting. Despite this, from a reproducibility stand point overfitting is the more obvious of the two problems being that one can inspect the graphs of the estimated secondary equations and identify overfitting more easily.

6 Summary and Conclusion

In this paper I have provided an asymptotic characterization of an estimator for the finite dimensional parameter and unknown additive functions of the partially linear primary equation in a triangular system of equations constructed to handle non parametrically defined endogenous regressors using the control function approach. Theorem 3 shows that the estimator for β_1 is consistent, $\sqrt{n}$ asymptotically normal, and Oracle efficient. Theorem 5 shows that the estimator for $h_d(x_d)$ is consistent, point wise asymptotically normal, and Oracle efficient. Additionally, both Theorem 3 and the Monte Carlo exercise presented in the previous section show that in an additive context, the estimator for β_1 presented in this paper practically identical to the Oracle estimator of Manzan and Zerom (2005) and is superior to the estimator developed in Geng et al. (2017). Lastly I showed that statistical inference conducted on estimator $\hat{\beta}_1$ is stable to moderate levels of weak instruments and overfitting.

References

Ai, C., Chen, X., November 2003. Efficient estimation of models with conditional moment restrictions containing unknown functions. Econometrica 71 (6), 1795–1843.

Angrist, J. D., Pischke, J.-S., 2009. Mostly Harmless Econometrics. Princeton University Press, Princeton, NJ.

Baker, M., May 2016. Is there a reproducibilty crisis? Nature 533, 452–454.

Begley, G., Ellis, L., 2012. Raise standards for preclinical cancer research. Nature 483, 531–533.

Buja, A., Hastie, T., Tibrishani, R., June 1989. Linear smoothers and additive models. The Annals of Statistics 17 (2), 453 – 510.

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

Darroles, S., Florens, J., Renault, E., Fan, Y., September 2011. Nonparametric instrumental regression. Econometrica 79 (5), 1541–1565.

de Boor, C., 2001. A Practical Guide to Splines. Vol. 27 of Applied Mathematical Sciences. Springer-Verlag.

Escanciano, J. C., Jacho-Chavez, D., Lewbel, A., May 2010. Identification and estimation of semiparametric two step estimators, working Paper.

Escanciano, J. C., Jacho-Chávez, D., Lewbel, A., 2016. Identification and estimation of semiparametric two step models. Quantitative Economics 7, 561–589.

Geng, X., Martins-filho, C., Yao, F., October 2016. Estimation of a partially linear regression in triangular systems.

Geng, X., Martins-filho, C., Yao, F., August 2017. Estimation of a partially linear regression in triangular systems, working Paper.

Haavelmo, T., January 1943. The statistical implications of a system of simultaneous equations. Econometrica 11 (1), 1–12.

Hahn, J., Ridder, G., January 2013. Asymptotic variance of semiparametric estimators with generated regressors. Econometrica 81 (1), 315–340.

Han, S., September 2017. Nonparametric estimation of triangular simultaneous equations models under weak identification.

Hartford, J., Lewis, G., Leyton-Brown, K., Taddy, M., 2017. Deep iv: A flexible approach to counterfactual prediction. Proceedings of the 34th Annual International Conference on MAchine Learning 70, 1414–1423.

Hausman, J., Newey, W. K., Ichimura, H., Powell, J. L., 1991. Identification and estimation of polynomia errors-in-variables models. Journal of Econometrics 50, 273–295.

Hausman, J., Newey, W. K., Powell, J. L., 1995. Nonlinear errors in variables estimation or some engel curves. Journal of Econometrics 65, 205–233.

Heckman, J., 1979. Sample selection as specification error. Econometrica 47 (1), 153–161.

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

Li, Q., November 2000. Efficient estimation of additive partially linear models. International Economic Review 41 (4), 1073 – 1092.

Li, Q., Racine, J., 2007. Nonparametric Econometrics. Princeton University Press.

Linton, O., Nielsen, J. P., March 1995. A kernel method of estimating structured nonparametric regression based on marginal integration. Biometrika 82 (1), 93–100.

Mammen, E., Rothe, C., Shienle, M., 2016. Semiparametric estimation with generated covariates. Econometric Theory 32, 1140 – 1177.

Manzan, S., Zerom, D., 2005. Kernel estimation of a partially linear model. Statstics and Probability Letters 73, 313 – 322.

Martins-filho, C., Yao, F., 2012. Kernel-based estimation of semiparametric regression in triangular systems. Economics Letters 115, 24–27.

Morimune, K., May 1983. Approximate distributions of k-class estimators when the degree of overidentifiablility is large compared with sample size. Econometrica 51 (3), 821–841.

Newey, W. K., 1997. Convergence rates and asymptotic normality for series estimators. Journal of Econometrics 79, 147–168.

Newey, W. K., Powell, J. L., Vella, F., May 1999. Nonparametric estimation of triangular simultaneous equations models. Econometrica 67 (3), 565–603.

Nosek, B. A., Spies, J. R., Motyl, M., 2012. Scientific utopia: Ii. restructuring incentives and practices to promote truth over publishibility. Perspective on Psychological Sciences 7 (6), 615–613.

Olley, G. S., Pakes, A., November 1996. The dynamics of productivity in the telecommunications equipment industry. Econometrica 64 (6), 1263 – 1297.

Ozabaci, D., Henderson, D., Su, L., March 2013. Additive nonparametric regression in the presence of endogenous regressors, working Paper.

Petrin, A., Train, K., 2010. A control function approach to endogeneity in consumer choice models. Journal of Marketing Research 47, 3–13.

Pinkse, J., Jun 2000. Nonparametric two-step regression estimation when regressors and error are dependent. The Canadian Journal of Statistics 28 (2), 289–300.

Robinson, P., July 1988. Root-n-consistent semiparametric regression. Econometrica 56 (4), 931 – 954.

Staiger, D., Stock, J., May 1997. Instrumental variables regression with weak instruments. Econometrica 65 (3), 557–586.

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

Vapnik, V., 1998. Statistical Learning Theory. Wiley, New York.

Working, E., 1927. What do statistical 'demand curves' show? Quarterly Journal of Economics 41, 212–235.

Wright, P. G., 1928. The Tariff on Animal and Vegetable Oils. Macmillan, New York.

Yu, B., 2013. Stability. Bernoulli 19 (4), 1484–1500.

Yu, K., Mammen, E., Park, B. U., 2011. Semi-parametric regression: Efficiency gains from modeling the nonparemetric part. Bernoulli 17 (2), 736–748.

7 Appendix

7.1 Proofs of Theorems and Lemma 1

Note: The following section contains proofs of Lemma 1 and Theorems 1 through 5, the statement and proof of 11 additional supporting Lemmas are contained within the technical supplement to this paper.

Adopt the following notation $\mathsf{S}_1^d(Y_i) \equiv H_1^d(Y_i) - \hat{H}_1^d(Y_i)$, $\mathsf{S}_2^d(Y_i) \equiv H_2^d(Y_i) - \hat{H}_2^d(Y_i)$, $\mathsf{S}_1^d(Z_i) \equiv H_1^d(Z_i) - \hat{H}_1^d(Z_i)$, $\mathsf{S}_2^d(Y_i) \equiv H_2^d(Z_i) - \hat{H}_2^d(Z_i)$, $K_{1ji}(X_d) \equiv K_1[b_1^{-1}(X_{dj} - X_{di})]$, and $K_{2ji}(V_d) \equiv K_2[b_1^{-1}(V_{dj} - V_{di})]$ furthermore define,

$$\begin{aligned}
\mathsf{C}_{2ji}^d &= (\hat{\theta}_{2j}^d - \theta_{2j}^d)Z_{cj} + \theta_{2j}^d(Z_{cj} - E[Z_{cj}|X_j, V_j]) + \theta_{2j}^d(E[Z_{cj}|X_j, V_j] - H_2^d(Z_{cj})) \\
&\qquad + \theta_{2j}^d(H_2^d(Z_{cj}) - H_2^d(Z_{ci})), \\
\mathsf{C}_{1ji}^d &= (\hat{\theta}_{1j}^d - \theta_{1j}^d)Z_{cj} + \theta_{1j}^d(Z_{cj} - E[Z_{cj}|X_j, V_j]) + \theta_{1j}^d(E[Z_{cj}|X_j, V_j] - H_1^d(Z_{cj})) \\
&\qquad + \theta_{1j}^d(H_1^d(Z_{cj}) - H_1^d(Z_{ci})), \\
\mathsf{C}_{2ji}^{*d} &= (\hat{\theta}_{2j}^d - \theta_{2j}^d)Z_{cj} + \theta_{2j}^d(Z_{cj} - E[Z_{cj}|X_j, V_j]) + \theta_{2j}^d(E[Z_{cj}|X_j, V_j] - H_2^d(Z_{cj})) + \theta_{2j}^d H_2^d(Z_{cj}), \\
\mathsf{C}_{1ji}^{*d} &= (\hat{\theta}_{1j}^d - \theta_{1j}^d)Z_{cj} + \theta_{1j}^d(Z_{cj} - E[Z_{cj}|X_j, V_j]) + \theta_{1j}^d(E[Z_{cj}|X_j, V_j] - H_1^d(Z_{cj})) + \theta_{1j}^d H_1^d(Z_{cj}).
\end{aligned}$$

Furthermore, let $\Delta_{ji}(X, V) \equiv H^{-1}\{(X_j', V_j')' - (X_i', V_i')'\}$ and define,

$$\begin{aligned}
K_{2ji}^{(m)}(V_d) &\equiv K_2^{(m)}[b_2^{-1}(V_{dj} - V_{di})] = \frac{d^m}{d^m\gamma} K_2(\gamma)\Big|_{\gamma = b_2^{-1}(V_{dj} - V_{di})}, \\
K_{2ji}^{(4)}(\tilde{V}_d) &\equiv \frac{d^m}{d^m\gamma} K_2(\gamma)\Big|_{\gamma = \lambda b_2^{-1}(\hat{V}_{dj} - \hat{V}_{di})) + (1-\lambda) b_2^{-1}(V_{dj} - V_{di}))}, \\
D_d K_{3ji}(X, V) &\equiv D_d K_3[\Delta_{ji}(X, V)] = \frac{\partial}{\partial \gamma_{2d}} K_3[(\gamma_1', \gamma_2')']\Big|_{(\gamma_1', \gamma_2')' = \Delta_{ji}(X, V)}, \\
D_{kd} K_{3ji}(X, V) &\equiv D_{kd} K_3[\Delta_{ji}(X, V)] = \frac{\partial^2}{\partial \gamma_{2d} \partial \gamma_{2k}} K_3[(\gamma_1', \gamma_2')']\Big|_{(\gamma_1', \gamma_2')' = \Delta_{ji}(X, V)}, \\
D_{mkd} K_{3ji}(X, V) &\equiv D_{mkd} K_3[\Delta_{ji}(X, V)] = \frac{\partial^3}{\partial \gamma_{2d} \partial \gamma_{2k} \partial \gamma_{2m}} K_3[(\gamma_1', \gamma_2')']\Big|_{(\gamma_1', \gamma_2')' = \Delta_{ji}(X, V)\}}, \\
D_{qmkd} K_{3ji}(X, V) &\equiv D_{mkd} K_3[\Delta_{ji}(X, V)] = \frac{\partial^4}{\partial \gamma_{2d} \partial \gamma_{2k} \partial \gamma_{2m} \partial \gamma_{2q}} K_3[(\gamma_1', \gamma_2')']\Big|_{(\gamma_1', \gamma_2')' = \Delta_{ji}(X, V)}.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\mathsf{S}_{1n}^d(Y) &= \left[\mathsf{S}_1^d(Y_1)\ \ \mathsf{S}_1^d(Y_2)\ \ \cdots\ \ \mathsf{S}_1^d(Y_n)\right]', & \mathsf{S}_{2n}^d(Y) &= \left[\mathsf{S}_2^d(Y_1)\ \ \mathsf{S}_2^d(Y_2)\ \ \cdots\ \ \mathsf{S}_2^d(Y_n)\ \right]', \\
\mathsf{S}_{1n}^d(Z) &= \left[\mathsf{S}_1^d(Z_1)\ \ \mathsf{S}_1^d(Z_2)\ \ \cdots\ \ \mathsf{S}_1^d(Z_n)\right]', & \mathsf{S}_{2n}^d(Z) &= \left[\mathsf{S}_2^d(Z_1)\ \ \mathsf{S}_2^d(Z_2)\ \ \cdots\ \ \mathsf{S}_2^d(Z_n)\ \right]', \\
\mathbf{K}_{1i}(X_d) &= \left[K_{11i}(X_d)\ \ K_{12i}(X_d)\ \ \cdots\ \ K_{1ni}(X_d)\right]', & \mathbf{K}_{2i}^{(m)}(V_d) &= \left[K_{21i}^{(m)}(V_d)\ \ K_{22i}^{(m)}(V_d)\ \ \cdots\ \ K_{2ni}^{(m)}(V_d)\right]', \\
\mathbf{D}_d K_{3i} &= \left[D_d K_{31i}(X, V)\ \ \cdots\ \ D_d K_{3ni}(X, V)\right]', & \mathbf{M}_d &= [m_d(W_1)\ \ m_d(W_2)\ \ \cdots\ \ m_d(W_n)]', \\
\mathbf{M}_d^{l_n} &= [m_d^{l_n}(W_1)\ \ m_d^{l_n}(W_2)\ \ \cdots\ \ m_d^{l_n}(W_n)]', & \hat{\mathbf{M}}_d^{l_n} &= [\hat{m}_d^{l_n}(W_1)\ \ \hat{m}_d^{l_n}(W_2)\ \ \cdots\ \ \hat{m}_d^{l_n}(W_n)]',
\end{aligned}$$

and $I(-j)$ is an identity matrix of appropriate dimension with the jth diagonal entry set to zero. Also,

$$\begin{aligned}
\dot{\boldsymbol{\zeta}}_c &= diag(\{\zeta_{ci}\}_{i=1}^n), & \dot{\boldsymbol{\Theta}}_{1n}^d &= diag\big(\{\theta_{1i}^d\}_{i=1}^n\big), & \dot{\boldsymbol{\Theta}}_{2n}^d &= diag\big(\{\theta_{2i}^d\}_{i=1}^n\big), \\
\dot{\mathbf{p}}(X_d)^{-1} &= diag\big(\{p(X_{di})^{-1}\}_{i=1}^n\big), & \dot{\mathbf{p}}(V_d)^{-1} &= diag\big(\{p(V_{di})^{-1}\}_{i=1}^n\big), & \dot{\mathbf{u}}_n &= diag\Big(\{u_i\}_{i=1}^n\Big), \\
\dot{\boldsymbol{\eta}}_{2c}^d &= diag(\{\eta_{2cj}^d\}_{j=1}^n) & \dot{\boldsymbol{\eta}}_{1c}^d &= diag(\{\eta_{1cj}^d\}_{j=1}^n) & \dot{\boldsymbol{\rho}}_c &= diag(\{\rho_{ci}\}_{i=1}^n)
\end{aligned}$$

$$\dot{\mathbf{H}}_2^d(Z_c) = diag(\{H_2^d(Z_{ci})\}_{i=1}^n) \qquad \dot{\boldsymbol{V}}_d = diag(\{V_{di}\}_{i=1}^n) \qquad \dot{\mathbf{V}}_{dj} = V_{dj} I_{l_n}$$

Also conditioning on S_i indicates conditioning on everything indexed by i and conditioning on S_{-i} indicates conditioning on everything not indexed by i.

Proof of Lemma 1

Part i) By Assumption A5

$$E\left[\phi\left(\sum_{j=1}^{D} h_j(X_j) + \sum_{j=1}^{D} f_j(V_j)\right)\middle| X_d\right] = \sum_{d=1}^{D} E(\phi h_d(X_d)|X_d) + \sum_{j=1}^{D} E(\phi f_d(V_d)|X_d)$$

$$= h_d(X_d)\int \frac{g(X,V)}{p(X,V)}\frac{p(X,V)}{p(X_d)} d(X_{-d},V) + \int \left(\sum_{j\neq d}^{D} h_j(X_j) + \sum_{j=1}^{D} f_j(V_j)\right)\frac{g(X,V)}{p(X,V)}\frac{p(X,V)}{p(X_d)} d(X_{-d},V)$$

$$= h_d(X_d)\int g(X_{-d},V)d(X_{-d},V) + \sum_{j\neq d}^{D}\int h_j(X_j)g(X_{-d},V)d(X_{-d},V)$$
$$+ \sum_{j=1}^{D}\int f_j(V_j)g(X_{-d},V)d(X_{-d},V)$$

$$= h_d(X_d) + \sum_{j\neq d}^{D} E[h_j(X_j)] + \sum_{j=1}^{D} E[f_j(V_j)] = h_d(X_d).$$

In a similar manner, $E\left[\phi\left(\sum_{j=1}^{D} h_j(X_j) + \sum_{j=1}^{D} f_j(V_j)\right)\middle| V_d\right] = f_d(V_d)$. Additionally note that,

$$\begin{aligned} E[\phi(Y - Z'\beta_1)] &= \mu_Y - \mu_Z'\beta_1 \\ &= \beta_0 E[\phi] + \sum_{d=1}^{D} E[\phi\, h_d(X_j)] + \sum_{d=1}^{D} E[\phi f_d(V_j)] + E[\phi\varepsilon] \\ &= \beta_0 \int \frac{g(X,V)}{p(X,V)} p(X,V)dXdV + \sum_{d=1}^{D}\int h_d(X_d)\frac{g(X,V)}{p(X,V)} p(X,V)dXdV \\ &\qquad + \sum_{d=1}^{D}\int f_d(V_d)\frac{g(X,V)}{p(X,V)} p(X,V)dXdV + E[\phi E(\varepsilon|Z,X,V)] \\ &= \beta_0 + \sum_{d=1}^{D} E[h_d(X_d)] + \sum_{d=1}^{D} E[f_d(V_d)] \\ &= \beta_0. \end{aligned}$$

As a result, provided μ_Y and μ_Z exist, β_0 is identified whenever β_1 is. Next consider,

$$H_1^d(Y) = \beta_0 E[\phi|X_d] + E[\phi Z|X_d]'\beta_1 + E\left[\phi\left(\sum_{j=1}^{D} h_j(X_j) + \sum_{j=1}^{D} f_j(V_j)\right)\middle| X_d\right] + E[\phi E(\varepsilon|Z,X,V)|X_d]$$

$$= \beta_0 + H_1^d(Z)'\beta_1 + h_d(X_d).$$

and,

$$H_2^d(Y) = \beta_0 E[\phi|V_d] + E[\phi Z|V_d]'\beta_1 + E\left[\phi\left(\sum_{j=1}^{D} h_j(X_j) + \sum_{j=1}^{D} f_j(V_j)\right)\middle| V_d\right] + E[\phi E(\varepsilon|Z,X,V)|V_d]$$

$$= \beta_0 + H_2^d(Z)'\beta_1 + f_d(V_d).$$

Combining these statements gives,

$$\begin{aligned} H(Y) &= \sum_{d=1}^{D}\left[H_1^d(Y) + H_2^d(Y)\right], \\ &= 2D\beta_0 + \sum_{d=1}^{D}\left[H_1^d(Z)'\beta_1 + H_2^d(Z)'\right]\beta_1 + \sum_{d=1}^{D}\left[h_d(X_d) + f_d(V_d)\right] \\ &= 2D\beta_0 + H(Z)'\beta_1 + h(X) + f(V). \end{aligned}$$

Next, recalling that $\beta_0 = \mu_Y - \mu_Z'\beta_1$,

$$\begin{aligned} H^*(Y) &= H(Y) - (2D-1)\mu_Y \\ &= 2D\beta_0 + H(Z) + h(X) + f(V) - (2D-1)\mu_Y \\ &= \beta_0 + (2D-1)\mu_Y - (2D-1)\mu_Z'\beta_1 + H(Z)'\beta_1 + h(X) + f(V) - (2D-1)\mu_Y \\ &= \beta_0 + H^*(Z)'\beta_1 + h(X) + f(V). \end{aligned}$$

Now, subtracting the above from equation (6) gives,

$$Y - H^*(Y) = (Z - H^*(Z))'\beta_1 + \varepsilon.$$

Lastly, premultiplying both sides by $Z - H^*(Z)$, and taking an expectation gives,

$$\begin{aligned} &E\Big[(Z - H^*(Z))(Y - H^*(Y))\Big] \\ &\qquad = E\Big[(Z - H^*(Z))(Z - H^*(Z))'\Big]\beta_1 + E\Big[(Z - H^*(Z))E[\varepsilon|Z,X,V]\Big] \\ &\qquad = E\Big[(Z - H^*(Z))(Z - H^*(Z))'\Big]\beta_1. \end{aligned}$$

Hence, if $E\Big[(Z-H^*(Z))(Z-H^*(Z))'\Big]$ is positive definite, then β_1 is identified. Now by A3(i) $Z \in L^2(\Omega, \mathcal{A}, P)$ so that by that projection theorem $Z - E[Z|X,V]$ is orthogonal to the space of square integrable functions of X, V. In particular $E[Z|X,V] - H^*(Z)$, as a result,

$$\begin{aligned} &E\Big[(Z - H^*(Z))(Z - H^*(Z))'\Big] \\ &\quad = E\left[(Z - E[Z|X,V] + E[Z|X,V] - H^*(Z))(Z - E[Z|X,V] + E[Z|X,V] - H^*(Z))'\right] \\ &\quad = E\left[(Z - E[Z|X,V])(Z - E[Z|X,V])'\right] + E\left[(E[Z|X,V] - H^*(Z))(E[Z|X,V] - H^*(Z))'\right] \\ &\quad = E\left[\rho\rho'\right] + E\left[\eta\eta'\right]. \end{aligned}$$

Consequently β_1 is identified if either, $E[\rho\rho']$ or $E[\eta\eta']$ is positive definite. **Part ii)**Recall that,

$$Y - H^*(Y) = (Z - H^*(Z))'\beta_1 + \varepsilon$$

and note that for any measurable function $L(X,V) : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$, $E\left(L(X,V)\varepsilon\right) = E\left(L(X,V)E[\varepsilon|Z,X,V]\right) = 0$. As a result, if we let $L(X,V) = \sqrt{\phi(X,V)}$, $\sqrt{\phi(X,V)}(Y - H^*(Y)) = \sqrt{\phi(X,V)}(Z - H^*(Z))'\beta_1 + \sqrt{\phi(X,V)}\varepsilon$. Then, premultiplying by $(Z - H^*(Z))\sqrt{\phi(X,V)}$ and taking expectations,

$$E\Big[(Z - H^*(Z))\phi(X,V)(Y - H^*(Y))\Big]$$

$$
\begin{aligned}
&= E\Big[(Z - H^*(Z))\phi(X,V)(Z - H^*(Z))'\Big]\beta_1 + E\Big[(Z - H^*(Z))\phi(X,V)E[\varepsilon|Z,X,V]\Big] \\
&= E\Big[(Z - H^*(Z))\phi(X,V)(Z - H^*(Z))'\Big]\beta_1 \\
&= E[\zeta\phi\zeta']\beta_1.
\end{aligned}
$$

Consequently if $E[\zeta\phi\zeta']$ is positive definite then β_1 is identified.

□

Proof of Theorem 1

Under the assumptions A1 - A5 of this paper the proof of uniform convergence of $\hat{p}(X_{di})$ to $p(X_{di})$, M_{2n} is well established in the literature and will not be repeated here, and the proof of uniform convergence of $\hat{p}(\hat{V}_{di})$ to $p(V_{di})$ follows from a simplification of the proof of uniform convergence of $\hat{p}(X_i\hat{V}_i)$ to $p(X_i, V_i)$. Accordingly I will only prove that latter case.

$$
\begin{aligned}
\hat{p}(X_i,\hat{V}_i) - \hat{p}(X_i,V_i) &= [nh_1^D h_2^D]^{-1}\sum_{j=1}^{n}\Big(K_3[H^{-1}\{(X_j,\hat{V}_j) - (X_i,\hat{V}_i\}] - K_3[H^{-1}\{(X_j,V_j) - (X_i,V_i\}]\Big) \\
&= [nh_1^D h_2^D]^{-1}\sum_{j=1}^{n}\big[K_{3ji}(X,\hat{V}) - K_{3ji}(X,V)\big] \\
&= \sum_{d=1}^{D}[nh_1^D h_2^{D+1}]^{-1}\sum_{j=1}^{n}D_dK_{3ji}(X,V)\big[\hat{V}_{dj} - V_{dj}\big] - \sum_{d=1}^{D}[nh_1^D h_2^{D+1}]^{-1}\sum_{j=1}^{n}D_dK_{3ji}(X,V)\big[\hat{V}_{di} - V_{di}\big] \\
&\quad + \sum_{1\le k,d\le D}[2nh_1^D h_2^{D+2}]^{-1}\sum_{j\ne i}D_{kd}K_{3ji}(X,V)\prod_{\xi\in\{d,k\}}\big[(\hat{V}_{\xi j} - V_{\xi j}) - \big(\hat{V}_{\xi i} - V_{\xi i}\big)\big]^2 \\
&\quad + \sum_{1\le m,k,d\le D}[6nh_1^D h_2^{D+3}]^{-1}\sum_{j\ne i}D_{mkd}K_{3ji}(X,V)\prod_{\xi\in\{m,d,k\}}\big[(\hat{V}_{\xi j} - V_{\xi j}) - \big(\hat{V}_{\xi i} - V_{\xi i}\big)\big]^3 \\
&\quad + \sum_{1\le q,m,k,d\le D}[24nh_1^D h_2^{D+4}]^{-1}\sum_{j\ne i}D_{qmkd}\tilde{K}_{3ji}(X,V)\prod_{\xi\in\{q,m,d,k\}}\big[(\hat{V}_{\xi j} - V_{\xi j}) - \big(\hat{V}_{\xi i} - V_{\xi i}\big)\big]^4 \\
&\equiv A_1 - A_2 + A_3 + A_4 + A_5.
\end{aligned}
$$

$$
\begin{aligned}
A_1 &= \sum_{d=1}^{D}[nh_1^D h_2^{D+1}]^{-1}\sum_{j=1}^{n}D_dK_{3ji}(X,V)\big[\hat{V}_{dj} - V_{dj}\big] \\
&= -\sum_{d=1}^{D}[nh_1^D h_2^{D+1}]^{-1}\sum_{j=1}^{n}D_dK_{3ji}(X,V)\big[\hat{m}_d^{l_n}(W_j) - m_d(W_j)\big] \equiv -\sum_{d=1}^{D}A_{1d}.
\end{aligned}
$$

$$
\begin{aligned}
A_{1d} &= [nh_1^D h_2^{D+1}]^{-1}\sum_{j=1}^{n}D_dK_{3ji}\big[\hat{m}_d^{l_n}(W_j) - m_d(W_j)\big] = [nh_1^D h_2^{D+1}]^{-1}\mathbf{D}_dK_{3i}'\big[\hat{\mathbf{M}}_d^{l_n} - \mathbf{M}_d\big] \\
&= [nh_1^D h_2^{D+1}]^{-1}\mathbf{D}_dK_{3i}'\big[\mathbf{B}_n\big(\hat{\alpha}_d^{l_n} - \alpha_d^{l_n}\big)\big] + [nh_1^D h_2^{D+1}]^{-1}\mathbf{D}_dK_{3i}'\big[\mathbf{B}_n\alpha_d^{l_n} - \mathbf{M}_d\big] \\
&= A_{11d} + A_{12d}.
\end{aligned}
$$

$$
\begin{aligned}
A_{11d} &= [nh_1^D h_2^{D+1}]^{-1}\mathbf{D}_dK_{3i}'\big[\mathbf{B}_n\big(\hat{\alpha}_d^{l_n} - \alpha_d^{l_n}\big)\big] \\
&= [nh_1^D h_2^{D+1}]^{-1}\mathbf{D}_dK_{3i}'\mathbf{B}_nQ_{BB}^{-1}n^{-1}\mathbf{B}_n'\mathbf{V}_d \\
&\quad + [nh_1^D h_2^{D+1}]^{-1}\mathbf{D}_dK_{3i}'\mathbf{B}_nQ_{BB}^{-1}n^{-1}\mathbf{B}_n'\big[\mathbf{M}_d - \mathbf{M}_d^{l_n}\big]
\end{aligned}
$$

$$
\begin{aligned}
&+ [nh_1^D h_2^{D+1}]^{-1}\mathbf{D}_d K_{3i}'\mathbf{B}_n\big(Q_{nBB}^{-1} - Q_{BB}^{-1}\big)n^{-1}\mathbf{B}_n'\mathbf{V}_d \\
&+ [nh_1^D h_2^{D+1}]^{-1}\mathbf{D}_d K_{3i}'\mathbf{B}_n\big(Q_{nBB}^{-1} - Q_{BB}^{-1}\big)n^{-1}\mathbf{B}_n'\big[\mathbf{M}_d - \mathbf{M}_d^{l_n}\big] \\
\equiv\ & A_{111d} + A_{112d} + A_{113d} + A_{114d}.
\end{aligned}
$$

$$
\begin{aligned}
\big|A_{111d}\big| &= \big|[nh_1^D h_2^{D+1}]^{-1}\mathbf{D}_d K_{3i}'\mathbf{B}_n Q_{BB}^{-1} n^{-1}\mathbf{B}_n'\mathbf{V}_d\big| \\
&\le ||[nh_1^D h_2^{D+1}]^{-1}\mathbf{D}_d K_{3i}'\mathbf{B}_n||_E\ ||Q_{BB}^{-1}||_{sp}||n^{-1}\mathbf{B}_n'\mathbf{V}_d||_E \\
&= O_p(1)O(1)O_p(l_n/\sqrt{n}) = O_p(\sqrt{l_n}/\sqrt{n}).
\end{aligned}
$$

By Lemma 4 and Lemma 5 .

$$
\begin{aligned}
\big|A_{112d}\big| &= \big|[nh_1^D h_2^{D+1}]^{-1}\mathbf{D}_d K_{3i}'\mathbf{B}_n Q_{BB}^{-1} n^{-1}\mathbf{B}_n'\big[\mathbf{M}_d - \mathbf{M}_d^{l_n}\big]\big| \\
&\le ||[nh_1^D h_2^{D+1}]^{-1}\mathbf{D}_d K_{3i}'\mathbf{B}_n||_E\ ||Q_{BB}^{-1}||_{sp}||n^{-1}\mathbf{B}_n'\big[\mathbf{M}_d - \mathbf{M}_d^{l_n}\big]||_E \\
&= O_p(1)O(1)O_p(l_n^{-k}) = O_p(l_n^{-k}).
\end{aligned}
$$

By Lemma 4 and Lemma 5 .

$$
\begin{aligned}
\big|A_{113d}\big| &\le ||[nh_1^D h_2^{D+1}]^{-1}\mathbf{D}_d K_{3i}'\mathbf{B}_n||_E\ ||Q_{nBB}^{-1} - Q_{BB}^{-1}||_{sp}||n^{-1}\mathbf{B}_n'\mathbf{V}_d||_E \\
&= O_p(1)O_p(l_n/\sqrt{n})O_p(\sqrt{l_n}/\sqrt{n}) = o_p(\sqrt{l_n}/\sqrt{n}).
\end{aligned}
$$

By Lemmas 4 and 5.

$$
\begin{aligned}
\big|A_{114d}\big| &\le ||[nh_1^D h_2^{D+1}]^{-1}\mathbf{D}_d K_{3i}'\mathbf{B}_n||_E\ ||Q_{nBB}^{-1} - Q_{BB}^{-1}||_{sp}||n^{-1}\mathbf{B}_n'\big[\mathbf{M}_d - \mathbf{M}_d^{l_n}\big]||_E \\
&= O_p(1)O_p(l_n/\sqrt{n})O_p(l_n^{-k}) = o_p(l_n^{-k}).
\end{aligned}
$$

Consequently,

$$
A_{11d} = A_{111d} + A_{112d} + A_{113d} + A_{114d} = O_p\big(\sqrt{l_n}/\sqrt{n} + l_n^{-k}\big).
$$

Now,

$$
\begin{aligned}
A_{12d} &= [nh_1^D h_2^{D+1}]^{-1}\mathbf{D}_d K_{3i}'\big[\mathbf{B}_n\alpha_d^{l_n} - \mathbf{M}_d\big] \\
&= [nh_1^D h_2^{D+1}]^{-1}\sum_{j=1}^{n} D_d K_{3ji}\big[m_d^{l_n}(W_j) - m_d(W_j)\big].
\end{aligned}
$$

Note that,

$$
\begin{aligned}
E\big(A_{12d}^2\big) &= [n^2 h_1^D h_2^{D+2}]^{-1}\sum_{j=1}^{n} E\Big(E\Big[[h_1^D h_2^D]^{-1} D_d K_{3ji}(X,V)^2\Big|W_j\Big]\big[m_d^{l_n}(W_j) - m_d(W_j)\big]^2\Big) \\
&\quad + n^{-2}\sum_{j=1}^{n}\sum_{g\neq j}^{n} E\Big(E\Big[[h_1^D h_2^{D+1}]^{-1} D_d K_{3ji}(X,V)\Big|V_i, W_j, S_g\Big]\big[m_d^{l_n}(W_j) - m_d(W_j)\big] \\
&\qquad\qquad \times E\Big[[h_1^D h_2^{D+1}]^{-1} D_d K_{3gi}(X,V)\Big|V_i, W_g, S_j\Big]\big[m_d^{l_n}(W_g) - m_d(W_g)\big]\Big) \\
&\le O_p([nh_1^D h_2^{D+2}]^{-1} + 1)\sup_{W\in G_W}|m_d^{l_n}(W) - m_d(W)|^2 = O_p\left(\frac{l_n^{-2k}}{nh_1^D h_2^{D+2}} + l_n^{-2k}\right) = O_p(l_n^{-2k})(o(1)+1).
\end{aligned}
$$

by Lemma 3 i) and integration by parts. Consequently by Markov's Inequality $A_{12d} = O_p(l_n^{-k})$. In all,

$$
A_1 = \sum_{d=1}^{D}\big(A_{11d} + A_{12d}\big) = O_p\left(\sqrt{\frac{l_n}{n}} + l_n^{-k}\right).
$$

$$
\begin{aligned}
A_{2d} \leq & \sup_{W\in G_W} |m_d(W) - \hat{m}_d^{l_n}(W)| \Big\{ \sup_{V_{di}\in G_{V_d}} \Big| [nh_1^D h_2^{D+1}]^{-1} \sum_{j=1}^n \big(D_d K_{3ji}(X,V) - E\big[D_d K_{3ji}(X,V)\big]\big) \Big| \\
& \qquad + \sup_{V_{di}\in G_{V_d}} \Big| [nh_1^D h_2^{D+1}]^{-1} \sum_{j=1}^n E\big[D_d K_{3ji}(X,V)\big] \Big| \Big\} \\
= & O_p(L_n) \left\{ O_p\left(\left[\frac{\log(n)}{nh_1^D h_2^{D+2}} \right]^{1/2} \right) + O(1) \right\} = O_p(L_n).
\end{aligned}
$$

By Lemma 3 xvii).

$$
\begin{aligned}
A_3 + A_4 + A_5 \leq & C \max_{1\leq d\leq D} [2 \sup_{W\in G_W} |m_d^{l_n}(W) - m_d(W)|]^2 [nh_2^2]^{-1} \sum_{j\neq i} [h_1^D h_2^D]^{-1} |D_{kd} K_{3ji}(X,V)| \\
& + C \max_{1\leq d\leq D} \big[2 \sup_{W\in G_W} |m_d^{l_n}(W) - m_d(W)|\big]^3 [nh_2^3]^{-1} \sum_{j\neq i} [h_1^D h_2^D]^{-1} |D_{kmd} K_{3ji}(X,V)| \\
& + C \max_{1\leq d\leq D} \big[2 \sup_{W\in G_W} |m_d^{l_n}(W) - m_d(W)|\big]^4 [nh_2^8]^{-1} \sum_{j\neq i} [h_1^D h_2^{D-4}]^{-1} |D_{qkmd} \tilde{K}_{3ji}(X,V)| \\
= & O_p\Big(\frac{L_n^2}{h_2^2} + \frac{L_n^3}{h_2^3} + \frac{L_n^4}{h_2^8} \Big) = o_p(n^{-1/2}).
\end{aligned}
$$

By Lemma 3 vi) and vii). Now, in summary, noting that each of the preceding result applies uniformly one has,

$$
\sup_{X_i, V_i \in G_{XV}} |\hat{p}(X_i, \hat{V}_i) - \hat{p}(X_i, V_i)| = O_p(L_n) + o_p(n^{-1/2}).
$$

Consequently under the assumptions A2, and A3, Theorem 1.4 of Li & Racine (2000) states that,

$$
\begin{aligned}
\sup_{X_i, V_i \in G_{XV}} |\hat{p}(X_i, \hat{V}_i) - p(X_i, V_i)| \leq & \sup_{X_i, V_i \in G_{XV}} |\hat{p}(X_i, \hat{V}_i) - \hat{p}(X_i, V_i)| + \sup_{X_i, V_i \in G_{XV}} |\hat{p}(X_i, V_i) - p(X_i, V_i)| \\
= & O_p(L_n) + O_p(M_{2n}).
\end{aligned}
$$

A proof of the uniform rate of convergence of $\hat{\theta}_1^d(X_i, \hat{V}_i)$ to $\theta_1^d(X_i, V_i)$ and $\hat{\theta}_2^d(X_i, \hat{V}_i)$ to $\theta_2^d(X_i, V_i)$ follows from a trivial modification to the proof of the uniform rate of convergence of $\hat{\phi}(X_i, \hat{V}_i)$ to $\phi(X_i, V_i)$. Consequently I only provide a proof of the latter case. Note that by Lemma 7 ,

$$
\begin{aligned}
\hat{g}(X_i, \hat{V}_i) - g(X_i, V_i) \leq & 2 \sum_{d=1}^D \sup_{X_{-d}, V \in G_{X_{-d}V}} g(X_{-d}, V) \sup_{X_d \in G_{X_d}} |\hat{p}(X_d) - p(X_d)\big] \\
& + 2 \sum_{d=1}^D \sup_{X, V_{-d} \in G_{XV_{-d}}} g(X, V_{-d}) \sup_{V_d \in G_{V_d}} |\hat{p}(\hat{V}_d) - p(V_d)| + o_p(n^{-1/2}) \\
= & O_p(\mathcal{L}_{0n}),
\end{aligned}
$$

By previous results and Lemma 3 vi). Now consider,

$$
\begin{aligned}
\hat{\phi}(X_i, \hat{V}_i) - \phi(X_i, V_i) = & \hat{p}(X_i, \hat{V}_i)^{-1} \hat{g}(X_i, \hat{V}_i) - p(X_i, V_i)^{-1} g(X_i, V_i) \\
= & [\hat{p}(X_i, \hat{V}_i) p(X_i, V_i)]^{-1} \Big(p(X_i, V_i) \big[\hat{g}(X_i, \hat{V}_i) - g(X_i, V_i)\big] + g(X_i, V_i) \big[p(X_i, V_i) - \hat{p}(X_i, \hat{V}_i)\big] \Big) \\
= & \hat{p}(X_i, \hat{V}_i)^{-1} \big[\big(\hat{g}(X_i, \hat{V}_i) - g(X_i, V_i)\big) + \phi_i \big(p(X_i, V_i) - \hat{p}(X_i, \hat{V}_i)\big) \big]
\end{aligned}
$$

$$
\begin{aligned}
&\quad \times \left[\big(\hat{g}(X_i,\hat{V}_i) - g(X_i,V_i)\big) + \phi_i\big(p(X_i,V_i) - \hat{p}(X_i,\hat{V}_i)\big)\right] \\
&\le \Big[p(X_i,V_i)^{-1} + \big(p(X_i,V_i)^2 + p(X_i,V_i)\big[\hat{p}(X_i,\hat{V}_i) - p(X_i,V_i)\big]\big)^{-1}\big[\hat{p}(X_i,\hat{V}_i) - p(X_i,V_i)\big]\Big] \\
&\quad \times \left[\big(\hat{g}(X_i,\hat{V}_i) - g(X_i,V_i)\big) + \phi_i\big(p(X_i,V_i) - \hat{p}(X_i,\hat{V}_i)\big)\right] \\
&\equiv \big[A_1 + A_2\big]\big[\big(\hat{g}(X_i,\hat{V}_i) - g(X_i,V_i)\big) + \phi_i\big(p(X_i,V_i) - \hat{p}(X_i,\hat{V}_i)\big)\big].
\end{aligned}
$$

Where,

$$
A_1 = p(X_i,V_i)^{-1} \quad \text{and} \quad A_2 = \big(p(X_i,V_i)^2 + p(X_i,V_i)\big[\hat{p}(X_i,\hat{V}_i) - p(X_i,V_i)\big]\big)^{-1}\big[\hat{p}(X_i,\hat{V}_i) - p(X_i,V_i)\big]
$$

Furthermore,

$$
\begin{aligned}
&A_2\big[\big(\hat{g}(X_i,\hat{V}_i) - g(X_i,V_i)\big) + \phi_i\big(p(X_i,V_i) - \hat{p}(X_i,\hat{V}_i)\big)\big] \\
&\quad \le |A_2|\Big(\sup_{X,V\in G_{XV}} |\hat{g}(X_i,\hat{V}_i) - g(X_i,V_i)| + \sup_{X,V\in G_{XV}} \phi(X,V) \sup_{X,V\in G_{XV}} |p(X_i,V_i) - \hat{p}(X_i,\hat{V}_i)|\Big) \\
&\quad \le \Big(\inf_{X,V\in G_{XV}} p(X_i,V_i)^2 + \inf_{X,V\in G_{XV}} p(X_i,V_i) o_p(1)\Big)^{-1} \sup_{X,V\in G_{XV}} \big|\hat{p}(X_i,\hat{V}_i) - p(X_i,V_i)\big| O_p(\mathcal{L}_{0n}) \\
&\quad = O_p(\mathcal{L}_{0n}^2) = o_p(n^{-1/2}).
\end{aligned}
$$

By Assumption 3, and Lemma 3 xxvi). Also,

$$
\begin{aligned}
&A_1\big[\big(\hat{g}(X_i,\hat{V}_i) - g(X_i,V_i)\big) + \phi_i\big(p(X_i,V_i) - \hat{p}(X_i,\hat{V}_i)\big)\big] \\
&\quad \le |A_1|\Big(\sup_{X,V\in G_{XV}} |\hat{g}(X_i,\hat{V}_i) - g(X_i,V_i)| + \sup_{X,V\in G_{XV}} p(X,V) \sup_{X,V\in G_{XV}} |p(X_i,V_i) - \hat{p}(X_i,\hat{V}_i)|\Big) \\
&\quad \le \Big[\inf_{X,V\in G_{XV}} p(X,V)\Big]^{-1} O_p(\mathcal{L}_{0n}) = O(1) O_p(\mathcal{L}_{0n}).
\end{aligned}
$$

By Assumption 3. Hence,

$$
\hat{\phi}(X_i,\hat{V}_i) - \phi(X_i,V_i) = O_p(\mathcal{L}_{0n}) + o_p(n^{-1/2}).
$$

□

Proof of Theorem 2 : Let A be one of the component random variables in $[\,Y, X']'$ Note that,

$$
E\big(K_{2ji}^{(1)}(V_d) A_j \theta_{2j}^d\big) = E\big(K_{2ji}^{(1)}(V_d) p(V_{dj})^{-1} E\big[\phi_j A_j | V_{dj}, S_i\big]\big) = E\big(E\big[K_{2ji}^{(1)}(V_d) p(V_{dj})^{-1} H_2^d(A_j)\big|S_i\big]\big).
$$

Integration by parts gives,

$$
\begin{aligned}
&E\big[K_{2ji}^{(1)}(V_d) p(V_{dj})^{-1} H_2^d(A_j)\big|S_i\big] = \int K_{2ji}^{(1)}(V_d) p(V_{dj})^{-1} H_2^d(A_j) p(V_{dj}) dV_{dj} \\
&\quad = \Big[K_2(b_2^{-1}[V_{dj} - V_{di}]) E(\phi_j A_j | V_{dj})\Big]_{-\infty}^{\infty} - b_2 \int K_{2ji}(V_d) H_2^{(1)d}(A_j) dV_{dj} \\
&\quad \le \sup_{V_d \in G_{V_d}} |E\big(\phi A | V_d)|\Big[\lim_{\gamma\to\infty} K_2(\,\gamma\,) - \lim_{\gamma\to-\infty} K_2(\,\gamma\,)\Big] + \sup_{V_d\in G_{V_d}} |E^{(1)}\big(\phi A|V_d)| b_2^2 \int |K_2(\gamma)| d\gamma \quad = O(b_2^2).
\end{aligned}
$$

By assumptions A2 and A4. Thus,

$$
E\big(K_{2ji}^{(1)}(V_d) A_j \theta_{2j}^d\big) = E\Big(E\big[K_{2ji}^{(1)}(V_d) p(V_{dj})^{-1} H_2^d(A_j)\big|S_i\big]\Big) = O(b_2^2).
$$

$$
\begin{aligned}
\sup_{X_{di}\in G_{X_d}} &\Big|[(n-1)b_1]^{-1}\sum_{j\neq i}\big(\hat{\theta}^d_{1j}K_{1ji}(X_d)A_j-\theta^d_{1j}K_{1ji}(X_d)A_j\big)\Big| \\
&\leq \sup_{X,V\in G_{XV}}[(n-1)b_1]^{-1}\sum_{j\neq i}\big|K_{1ji}(X_d)\big|\big|\hat{\theta}^d_{1i}-\theta^d_{1i}\big|\big|A_j\big| \\
&\leq \sup_{X,V\in G_{XV}}\big|\hat{\theta}^d_1(X,\hat{V})-\theta^d_1(X,\hat{V})\big|\sup_{X_{di}\in G_{X_d}}[(n-1)b_1]^{-1}\sum_{j\neq i}\big|K_{1ji}(X_d)\big|\big|A_j\big|. \\
&= O_p(\mathcal{L}_{0n}).
\end{aligned}
$$

By Theorem 1. Furthermore, by Theorem 2.6 in Li and Racine (2007), under the assumptions of this paper,

$$
\sup_{X_{di}\in G_{X_d}}\Big|[(n-1)b_1]^{-1}\sum_{j\neq i}\theta^d_{1j}K_{1ji}(X_d)A_j-E\big[\phi_iA_i|X_{di}\big]\Big| = O_p\left(\left[\frac{\log(n)}{nb_1}\right]^{1/2}+b_1^{\nu_1}\right)=O_p(N_{1n}).
$$

Consequently, one has,

$$
\begin{aligned}
\sup_{X_{di}\in G_{X_d}}\big|\hat{H}^d_1(A_i)-H^d_1(A_i)\big| &= \sup_{X_{di}\in G_{X_d}}\Big|[(n-1)b_1]^{-1}\sum_{j\neq i}\hat{\theta}^d_{1j}K_{1ji}(X_d)A_j-E\big[\phi_iA_i|X_{di}\big]\Big| \\
&= O_p\big(\mathcal{L}_{0n}+N_{1n}\big)=O_p(L_n+M_n+N_{1n}) \\
&= O_p(\mathcal{L}_{1n}).
\end{aligned}
$$

Now,

$$
\begin{aligned}
[(n-1)b_2]^{-1}&\sum_{j\neq i}K_{2ji}(\hat{V}_d)\hat{\theta}^d_{2j}A_j-[(n-1)b_2]^{-1}\sum_{j\neq i}K_{2ji}(V_d)\theta^d_{2j}A_j \\
&\leq [(n-1)b_2]^{-1}\sum_{j\neq i}\big[K_{2ji}(\hat{V}_d)\hat{\theta}^d_{2j}A_j-K_{2ji}(\hat{V}_d)\theta^d_{2j}A_j+K_{2ji}(\hat{V}_d)\theta^d_{2j}A_j-K_{2ji}(V_d)\theta^d_{2j}A_j\big] \\
&\leq [(n-1)b_2]^{-1}\sum_{j\neq i}K_{2ji}(\hat{V}_d)\big(\hat{\theta}^d_{2j}-\theta^d_{2j}\big)A_j+[(n-1)b_2]^{-1}\sum_{j\neq i}\big(K_{2ji}(\hat{V}_d)-K_{2ji}(V_d)\big)\theta^d_{2j}A_j \\
&= [(n-1)b_2]^{-1}\sum_{j\neq i}K_{2ji}(V_d)\big(\hat{\theta}^d_{2j}-\theta^d_{2j}\big)A_j \\
&\qquad + [(n-1)b_2]^{-1}\sum_{j\neq i}\big(K_{2ji}(\hat{V}_d)-K_{2ji}(V_d)\big)\big(\hat{\theta}^d_{2j}-\theta^d_{2j}\big)A_j \\
&\qquad + [(n-1)b_2]^{-1}\sum_{j\neq i}\big(K_{2ji}(\hat{V}_d)-K_{2ji}(V_d)\big)\theta^d_{2j}A_j \\
&\equiv B_1+B_2+B_3.
\end{aligned}
$$

By Theorem 1

$$
\begin{aligned}
B_1 &= [(n-1)b_2]^{-1}\sum_{j\neq i}K_{2ji}(V_d)\big(\hat{\theta}^d_{2j}-\theta^d_{2j}\big)A_j \\
&\leq \sup_{X,V\in G_{XV}}|\hat{\theta}^d_2(X,\hat{V})-\theta^d_2(X,V)|[(n-1)b_2]^{-1}\sum_{j\neq i}|K_{2ji}(V_d)||A_j| = O_p(\mathcal{L}_{0n}).
\end{aligned}
$$

By Taylor expansion one has

$$
\big|B_2\big| = \Big|[(n-1)b_2]^{-1}\sum_{j\neq i}\big(K_{2ji}(\hat{V}_d)-K_{2ji}(V_d)\big)\big(\hat{\theta}^d_{2j}-\theta^d_{2j}\big)A_j\Big|
$$

$$
\begin{aligned}
&\leq \sup_{X,V\in G_{XV}} |\hat{\theta}_2^d(X,\hat{V}) - \theta_2^d(X,V)|[(n-1)b_2]^{-1}\sum_{j\neq i}|K_{2ji}(\hat{V}_d) - K_{2ji}(V_d)||A_j| \\
&= O_p(\mathcal{L}_{0n})\Big\{[(n-1)b_2^2]^{-1}\sum_{j\neq i}\big|K_{2ji}^{(1)}(V_d)\big|\big|A_j\big|\big|\big(\hat{m}_d^{l_n}(W_j) - m_d(W_j)\big) - \big(\hat{m}_d^{l_n}(W_i) - m_d(W_i)\big)\big| \\
&\qquad + [2(n-1)b_2^3]^{-1}\sum_{j\neq i}\big|K_{2ji}^{(2)}(V_d)\big|\big|A_j\big|\big|\big(\hat{m}_d^{l_n}(W_j) - m_d(W_j)\big) - \big(\hat{m}_d^{l_n}(W_i) - m_d(W_i)\big)\big|^2 \\
&\qquad + [6(n-1)b_2^4]^{-1}\sum_{j\neq i}\big|K_{2ji}^{(3)}(V_d)\big|\big|A_j\big|\big|\big(\hat{m}_d^{l_n}(W_j) - m_d(W_j)\big) - \big(\hat{m}_d^{l_n}(W_i) - m_d(W_i)\big)\big|^3 \\
&\qquad + [24(n-1)b_2^5]^{-1}\sum_{j\neq i}\big|\tilde{K}_{2ji}^{(4)}(V_d)\big|\big|A_j\big|\big|\big(\hat{m}_d^{l_n}(W_j) - m_d(W_j)\big) - \big(\hat{m}_d^{l_n}(W_i) - m_d(W_i)\big)\big|^4\Big\} \\
&= O_p(\mathcal{L}_{0n})\Big\{\sup_{X,V\in G_{XV}}\big|\hat{m}_d^{l_n}(W) - m_d(W)\big|[(n-1)b_2]^{-1}\sum_{j\neq i}\big|b_2^{-1}K_{2ji}^{(1)}(V_d)\big|\big|A_j\big| \\
&\qquad + \sup_{X,V\in G_{XV}}\big|\hat{m}_d^{l_n}(W) - m_d(W)\big|^2[(n-1)b_2^2]^{-1}\sum_{j\neq i}\big|b_2^{-1}K_{2ji}^{(2)}(V_d)\big|\big|A_j\big| \\
&\qquad + \sup_{X,V\in G_{XV}}\big|\hat{m}_d^{l_n}(W) - m_d(W)\big|^3[(n-1)b_2^3]^{-1}\sum_{j\neq i}\big|b_2^{-1}K_{2ji}^{(3)}(V_d)\big|\big|A_j\big| \\
&\qquad + \sup_{X,V\in G_{XV}}\big|\hat{m}_d^{l_n}(W) - m_d(W)\big|^4[(n-1)b_2^5]^{-1}\sum_{j\neq i}\big|\tilde{K}_{2ji}^{(4)}(V_d)\big|\big|A_j\big|\Big\} \\
&= O_p(\mathcal{L}_{0n})\Big\{O_p\big(L_nb_2^{-1}\big)\big|[(n-1)]^{-1}\sum_{j\neq i}\big|b_2^{-1}K_{2ji}^{(1)}(V_d)\big|\big|A_j\big| \\
&\qquad + O_p\big(L_n^2b_2^{-2}\big)\big|[(n-1)]^{-1}\sum_{j\neq i}\big|b_2^{-1}K_{2ji}^{(2)}(V_d)\big|\big|A_j\big| \\
&\qquad + O_p\big(L_n^3b_2^{-3}\big)\big|[(n-1)]^{-1}\sum_{j\neq i}\big|b_2^{-1}K_{2ji}^{(3)}(V_d)\big|\big|A_j\big| \\
&\qquad + O_p\big(L_n^4b_2^{-5}\big)\big|\sup_{\gamma\in\mathbb{R}}\big|K_2^{(4)}(\ \gamma\ )\big|[(n-1)]^{-1}\sum_{j\neq i}\big|A_j\big|\Big\} \\
&= O_p(\mathcal{L}_{0n})\Big\{O_p\big(L_nb_2^{-1}\big) + O_p\big(L_n^2b_2^{-2}\big) + O_p\big(L_n^3b_2^{-3}\big) + O_p\big(L_n^4b_2^{-5}\big)\Big\} = o_p(\mathcal{L}_{0n}).
\end{aligned}
$$

$$
\begin{aligned}
B_3 &= [(n-1)b_2]^{-1}\sum_{j\neq i}\big(K_{2ji}(\hat{V}_d) - K_{2ji}(V_d)\big)\theta_{2j}^dA_j \\
&= (-1)[(n-1)b_2]^{-1}\sum_{j\neq i}b_2^{-1}K_{2ji}^{(1)}(V_d)A_j\theta_{2j}^d\big(\hat{m}_d^{l_n}(W_j) - m_d(W_j)\big) \\
&\qquad + [(n-1)b_2]^{-1}\sum_{j\neq i}b_2^{-1}K_{2ji}^{(1)}(V_d)A_j\theta_{2j}^d\big(\hat{m}_d^{l_n}(W_i) - m_d(W_i)\big) \\
&\qquad + [(n-1)b_2^2]^{-1}\sum_{j\neq i}b_2^{-1}K_{2ji}^{(2)}(V_d)A_j\theta_{2j}^d\big[\big(\hat{m}_d^{l_n}(W_i) - m_d(W_i)\big) - \big(\hat{m}_d^{l_n}(W_j) - m_d(W_j)\big)\big]^2 \\
&\qquad + [(n-1)b_3^2]^{-1}\sum_{j\neq i}b_2^{-1}K_{2ji}^{(3)}(V_d)A_j\theta_{2j}^d\big[\big(\hat{m}_d^{l_n}(W_i) - m_d(W_i)\big) - \big(\hat{m}_d^{l_n}(W_j) - m_d(W_j)\big)\big]^3 \\
&\qquad + [(n-1)b_5^2]^{-1}\sum_{j\neq i}\tilde{K}_{2ji}^{(4)}(V_d)A_j\theta_{2j}^d\big[\big(\hat{m}_d^{l_n}(W_i) - m_d(W_i)\big) - \big(\hat{m}_d^{l_n}(W_j) - m_d(W_j)\big)\big]^4 \\
&\equiv -B_{31} + B_{32} + B_{33} + B_{34} + B_{35}.
\end{aligned}
$$

Now consider,

$$B_{31} = [(n-1)b_2]^{-1}\sum_{j\neq i} b_2^{-1}K_{2ji}^{(1)}(V_d)A_j\theta_{2j}^d\big(\hat{m}_d^{l_n}(W_j)-m_d(W_j)\big)$$
$$= [(n-1)b_2^2]^{-1}\mathbf{K}_{2i}^{(1)}(V_d)'\dot{\mathbf{\Theta}}_{2n}^d\dot{\mathbf{A}}_nI_n(-i)\big[\hat{\mathbf{M}}_d^{l_n}-\mathbf{M}_d\big].$$

Now as in the proof of Theorem 1, and given the results of Lemma 4

$$\begin{aligned}
\big|B_{31}\big| = \Big|&[(n-1)b_2^2]^{-1}\mathbf{K}_{2i}^{(1)}(V_d)'\dot{\mathbf{\Theta}}_{2n}^d\dot{\mathbf{A}}_nI_n(-i)\mathbf{B}_nQ_{BB}^{-1}n^{-1}\mathbf{B}_n\mathbf{V}_d\\
&+[(n-1)b_2^2]^{-1}\mathbf{K}_{2i}^{(1)}(V_d)'\dot{\mathbf{\Theta}}_{2n}^d\dot{\mathbf{A}}_nI_n(-i)\mathbf{B}_n\big(Q_{nBB}^{-1}-Q_{BB}^{-1}\big)n^{-1}\mathbf{B}_n\mathbf{V}_d\\
&+[(n-1)b_2^2]^{-1}\mathbf{K}_{2i}^{(1)}(V_d)'\dot{\mathbf{\Theta}}_{2n}^d\dot{\mathbf{A}}_nI_n(-i)\mathbf{B}_nQ_{BB}^{-1}n^{-1}\mathbf{B}_n\big[\mathbf{M}_d-\mathbf{M}_d^{l_n}\big]\\
&+[(n-1)b_2^2]^{-1}\mathbf{K}_{2i}^{(1)}(V_d)'\dot{\mathbf{\Theta}}_{2n}^d\dot{\mathbf{A}}_nI_n(-i)\mathbf{B}_n\big(Q_{nBB}^{-1}-Q_{BB}^{-1}\big)n^{-1}\mathbf{B}_n\big[\mathbf{M}_d-\mathbf{M}_d^{l_n}\big]\Big|\\
\leq\ &||[(n-1)b_2^2]^{-1}\mathbf{K}_{2i}^{(1)}(V_d)'\dot{\mathbf{\Theta}}_{2n}^d\dot{\mathbf{A}}_nI_n(-i)\mathbf{B}_n||_E||Q_{BB}^{-1}n^{-1}\mathbf{B}_n\mathbf{V}_d||_E\\
&+||[(n-1)b_2^2]^{-1}\mathbf{K}_{2i}^{(1)}(V_d)'\dot{\mathbf{\Theta}}_{2n}^d\dot{\mathbf{A}}_nI_n(-i)\mathbf{B}_n||_E||\big(Q_{nBB}^{-1}-Q_{BB}^{-1}\big)n^{-1}\mathbf{B}_n\mathbf{V}_d||_E\\
&+||[(n-1)b_2^2]^{-1}\mathbf{K}_{2i}^{(1)}(V_d)'\dot{\mathbf{\Theta}}_{2n}^d\dot{\mathbf{A}}_nI_n(-i)\mathbf{B}_n||_E||Q_{BB}^{-1}n^{-1}\mathbf{B}_n\big[\mathbf{M}_d-\mathbf{M}_d^{l_n}\big]||_E\\
&+||[(n-1)b_2^2]^{-1}\mathbf{K}_{2i}^{(1)}(V_d)'\dot{\mathbf{\Theta}}_{2n}^d\dot{\mathbf{A}}_nI_n(-i)\mathbf{B}_n||_E||\big(Q_{nBB}^{-1}-Q_{BB}^{-1}\big)n^{-1}\mathbf{B}_n\big[\mathbf{M}_d-\mathbf{M}_d^{l_n}\big]||_E\\
\leq\ &||[(n-1)b_2^2]^{-1}\mathbf{K}_{2i}^{(1)}(V_d)'\dot{\mathbf{\Theta}}_{2n}^d\dot{\mathbf{A}}_nI_n(-i)\mathbf{B}_n||_E||Q_{BB}^{-1}||_{sp}||n^{-1}\mathbf{B}_n\mathbf{V}_d||_E\\
&+||[(n-1)b_2^2]^{-1}\mathbf{K}_{2i}^{(1)}(V_d)'\dot{\mathbf{\Theta}}_{2n}^d\dot{\mathbf{A}}_nI_n(-i)\mathbf{B}_n||_E||Q_{nBB}^{-1}-Q_{BB}^{-1}||_{sp}||n^{-1}\mathbf{B}_n\mathbf{V}_d||_E\\
&+||[(n-1)b_2^2]^{-1}\mathbf{K}_{2i}^{(1)}(V_d)'\dot{\mathbf{\Theta}}_{2n}^d\dot{\mathbf{A}}_nI_n(-i)\mathbf{B}_n||_E||Q_{BB}^{-1}||_{sp}||n^{-1}\mathbf{B}_n\big[\mathbf{M}_d-\mathbf{M}_d^{l_n}\big]||_E\\
&+||[(n-1)b_2^2]^{-1}\mathbf{K}_{2i}^{(1)}(V_d)'\dot{\mathbf{\Theta}}_{2n}^d\dot{\mathbf{A}}_nI_n(-i)\mathbf{B}_n||_E||Q_{nBB}^{-1}-Q_{BB}^{-1}||_{sp}||n^{-1}\mathbf{B}_n\big[\mathbf{M}_d-\mathbf{M}_d^{l_n}\big]||_E\\
=\ &O_p(b_2^{-1})\left[O(1)O_p\left(\sqrt{\frac{l_n}{n}}\right)+O_p\left(\frac{l_n}{\sqrt{n}}\right)O_p\left(\sqrt{\frac{l_n}{n}}\right)+O(1)O_p\left(l_n^{-k}\right)+O_p\left(\frac{l_n}{\sqrt{n}}\right)O_p\left(l_n^{-k}\right)\right]\\
=\ &O_p\left(b_2^{-1}\left[\sqrt{\frac{l_n}{n}}+l_n^{-k}\right]\right).
\end{aligned}$$

By assumption A1.

$$\begin{aligned}
B_{32} &= [(n-1)b_2]^{-1}\sum_{j\neq i} b_2^{-1}K_{2ji}^{(1)}(V_d)A_j\theta_{2j}^d\big(\hat{m}_d^{l_n}(W_i)-m_d(W_i)\big)\\
&\leq \sup_{W\in G_W}\big|\hat{m}_d^{l_n}(W)-m_d(W)\big|\big|[(n-1)b_2]^{-1}\sum_{j\neq i} b_2^{-1}K_{2ji}^{(1)}(V_d)A_j\theta_{2j}^d\big|\\
&\leq O_p(L_n)\Big\{\sup_{V_d\in G_{V_d}}\Big|[(n-1)b_2]^{-1}\sum_{j\neq i}\Big[b_2^{-1}K_{2ji}^{(1)}(V_d)A_j\theta_{2j}^d-E\big(b_2^{-1}K_{2ji}^{(1)}(V_d)A_j\theta_{2j}^d\big)\Big]\Big|\\
&\qquad+\Big|\sup_{V_d\in G_{V_d}}[(n-1)b_2]^{-1}\sum_{j\neq i}E\big(b_2^{-1}K_{2ji}^{(1)}(V_d)A_j\theta_{2j}^d\big)\Big|.
\end{aligned}$$

From the preliminary portion of this proof, and a combination of Assumption A3 and the results of Lemma 6 one obtains,

$$B_{32} = O_p(L_n)\left(O_p\left(\left[\frac{\log(n)}{nb_2^3}\right]^{1/2}\right)+O(1)\right) = O_p(L_n).$$

$$
\begin{aligned}
B_{33}+&B_{34}+B_{35} \leq 2 \sup_{X,V\in G_{XV}} \left|\hat{m}_d^{l_n}(W)-m_d(W)\right|^2 \sup_{X,V\in G_{XV}} \theta_2^d(X,V)[(n-1)b_2^2]^{-1}\sum_{j\neq i}\left|b_2^{-1}K_{2ji}^{(2)}(V_d)\right|\left|A_j\right| \\
&+8/6 \sup_{X,V\in G_{XV}} \left|\hat{m}_d^{l_n}(W)-m_d(W)\right|^3 \sup_{X,V\in G_{XV}} \theta_2^d(X,V)[(n-1)b_2^3]^{-1}\sum_{j\neq i}^{n}\left|b_2^{-1}K_{2ji}^{(3)}(V_d)\right|\left|A_j\right| \\
&+16/24 \sup_{X,V\in G_{XV}} \left|\hat{m}_d^{l_n}(W)-m_d(W)\right|^4 \sup_{X,V\in G_{XV}} \theta_2^d(X,V)\sup_{\gamma\in\mathbb{R}}|K_2^{(4)}(\ \gamma\ )|[(n-1)b_2^5]^{-1}\sum_{j\neq i}|A_j| \\
&\leq O_p(L_n^2 b_2^{-2})+O_p(L_n^3 b_2^{-3})+O_p(L_n^4 b_2^{-5})=o_p(n^{-1/2}),
\end{aligned}
$$

by Lemma 3 vi) and vii). In all,

$$
\begin{aligned}
B_3 &= -B_{31}+B_{32}+B_{33}+B_{34}+B_{35} \\
&= O_p\left(b_2^{-1}\left[\sqrt{\frac{l_n}{n}}+l_n^{-k}\right]\right)+O_p(L_n)+o_p(n^{1/2}) \\
&= O_p\left(L_n+b_2^{-1}\left[\sqrt{\frac{l_n}{n}}+l_n^{-k}\right]\right).
\end{aligned}
$$

Also,

$$
\begin{aligned}
&[(n-1)b_2]^{-1}\sum_{j\neq i}K_{2ji}(\hat{V}_d)\hat{\theta}_{2i}^d A_j-[(n-1)b_2]^{-1}\sum_{j\neq i}K_{2ji}(V_d)\theta_{2i}^d A_j \\
&\quad = B_1+B_2+B_3 \\
&\quad = O_p\left(L_n+\mathcal{L}_{0n}+b_2^{-1}\left[\sqrt{\frac{l_n}{n}}+l_n^{-k}\right]\right)=O_p\left(L_n+M_n+b_2^{-1}\left[\sqrt{\frac{l_n}{n}}+l_n^{-k}\right]\right).
\end{aligned}
$$

Furthermore, by Theorem 2.6 in Li and Racine (2007), and under the assumptions A2 and A3,

$$
\sup_{V_{di}\in G_{V_d}}\left|[(n-1)b_2]^{-1}\sum_{j\neq i}\left(\theta_{2j}^d K_{2ji}(V_d)A_j-E\left[\phi_i A_i|V_{di}\right]\right)\right|=O_p\left(\left[\frac{\log(n)}{nb_2}\right]^{1/2}+b_2^{\nu_2}\right)=O_p(N_{2n}).
$$

Consequently, one has,

$$
\begin{aligned}
\sup_{V_{di}\in G_{V_d}}\left|\hat{H}_2^d(A_i)-H_2^d(A_i)\right| &= \sup_{V_{di}\in G_{V_d}}\left|[(n-1)b_2]^{-1}\sum_{j\neq i}\left(\hat{\theta}_{2j}^d K_{2ji}(V_d)A_j-E\left[\phi_i A_i|V_{di}\right]\right)\right| \\
&= O_p\left(L_n+M_n+b_2^{-1}\left[\sqrt{\frac{l_n}{n}}+l_n^{-k}\right]+N_{2n}\right)=O_p(\mathcal{L}_{2n}).
\end{aligned}
$$

$$
\begin{aligned}
\hat{\mu}_A-\mu_A &= n^{-1}\sum_{j=1}\left(\hat{\phi}_j A_j-E\left[\phi_j A_j\right]\right) \\
&\leq \sup_{X,V\in G_{XV}}\left|\hat{\phi}(X,\hat{V})-\phi(X,V)\right|n^{-1}\sum_{j=1}^{n}|A_j|+n^{-1}\sum_{j=1}^{n}\left(\phi_j A_j-E\left[\phi_j A_j\right]\right) \\
&= O_p(\mathcal{L}_{0n})+O_p(n^{-1/2})=O_p(\mathcal{L}_{0n})=O_p(\mathcal{L}_{1n}),
\end{aligned}
$$

by Markov's Inequality. In summary,

$$\begin{aligned}
\sup_{X,V\in G_{XV}} |\hat{H}^*(A_j) - H^*(A_j)| &\le \sum_{d=1}^{D} \sup_{X_d\in G_{X_d}} |\hat{H}_1^d(A_j) - H_1^d(A_j)| \\
&\quad + \sum_{d=1}^{D} \sup_{V_d\in G_{V_d}} |\hat{H}_2^d(A_j) - H_2^d(A_j)| + (2D-1)[\hat{\mu}_A - \mu_A] \\
&= O_p\big(\mathcal{L}_{1n} + \mathcal{L}_{2n}\big) = O_p(\mathcal{L}_n).
\end{aligned}$$

□

Proof of Theorem 3:

$$\begin{aligned}
n^{1/2}\big(\hat{\beta}_1 - \beta_1\big) &= \big[n^{-1}\hat{\boldsymbol{\zeta}}_n'\hat{\boldsymbol{\phi}}_n\hat{\zeta}_n\big]^{-1}\sqrt{n}\big(n^{-1}\hat{\boldsymbol{\zeta}}_n'\hat{\boldsymbol{\phi}}_n\big[\mathbf{Y}_n - \hat{\mathbf{H}}_n^*(Y) - \big(\mathbf{Z}_n - \hat{\mathbf{H}}_n^*(Z)\big)\beta_1\big]\big) \\
&= A^{-1}\sqrt{n}B.
\end{aligned}$$

where

$$\begin{aligned}
A &= n^{-1}\big[\mathbf{Z}_n - \hat{\mathbf{H}}_n^*(Z)\big]'\hat{\boldsymbol{\phi}}_n\big[\mathbf{Z}_n - \hat{\mathbf{H}}_n^*(Z)\big], \\
B &= n^{-1}\big[\mathbf{Z}_n - \hat{\mathbf{H}}_n^*(Z)\big]'\hat{\boldsymbol{\phi}}_n\big[\mathbf{Y}_n - \hat{\mathbf{H}}_n^*(Y) - \big(\mathbf{Z}_n - \hat{\mathbf{H}}_n^*(Z)\big)\beta_1\big].
\end{aligned}$$

Note that,

$$\begin{aligned}
A &= n^{-1}\big[\mathbf{Z}_n - \hat{\mathbf{H}}_n^*(Z)\big]'\hat{\boldsymbol{\phi}}_n\big[\mathbf{Z}_n - \hat{\mathbf{H}}_n^*(Z)\big] \\
&= n^{-1}\big[\mathbf{Z}_n - \mathbf{H}_n^*(Z)\big]'\boldsymbol{\phi}_n\big[\mathbf{Z}_n - \mathbf{H}_n^*(Z)\big] + n^{-1}\big[\mathbf{Z}_n - \mathbf{H}_n^*(Z)\big]'\boldsymbol{\phi}_n\big[\mathbf{H}_n^*(Z) - \hat{\mathbf{H}}_n^*(Z)\big] \\
&\quad + n^{-1}\big[\mathbf{Z}_n - \mathbf{H}_n^*(Z)\big]'\big(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_n\big)\big[\mathbf{Z}_n - \mathbf{H}_n^*(Z)\big] \\
&\quad + n^{-1}\big[\mathbf{Z}_n - \mathbf{H}_n^*(Z)\big]'\big(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_n\big)\big[\mathbf{H}_n^*(Z) - \hat{\mathbf{H}}_n^*(Z)\big] \\
&\quad + n^{-1}\big[\mathbf{H}_n^*(Z) - \hat{\mathbf{H}}_n^*(Z)\big]'\boldsymbol{\phi}_n\big[\mathbf{Z}_n - \mathbf{H}_n^*(Z)\big] + n^{-1}\big[\mathbf{H}_n^*(Z) - \hat{\mathbf{H}}_n^*(Z)\big]'\boldsymbol{\phi}_n\big[\mathbf{H}_n^*(Z) - \hat{\mathbf{H}}_n^*(Z)\big] \\
&\quad + n^{-1}\big[\mathbf{H}_n^*(Z) - \hat{\mathbf{H}}_n^*(Z)\big]'\big(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_n\big)\big[\mathbf{Z}_n - \mathbf{H}_n^*(Z)\big] \\
&\quad + n^{-1}\big[\mathbf{H}_n^*(Z) - \hat{\mathbf{H}}_n^*(Z)\big]'\big(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_n\big)\big[\mathbf{H}_n^*(Z) - \hat{\mathbf{H}}_n^*(Z)\big] \\
&\equiv A_1 + A_2 + A_3 + A_4 + A_5 + A_6 + A_7 + A_8.
\end{aligned}$$

First, recall that,

$$\Sigma_0 = E\Big([Z - H^*(Z)]'\phi([Z - H^*(Z)]\Big).$$

Let $c, m \in \{1, 2, \ldots, p\}$,

$$A_1(c,m) = n^{-1}\sum_{i=1}^{n}[Z_{ci} - H^*(Z_{ci})]\phi_i[Z_{mi} - H^*(Z_{mi})] = n^{-1}\sum_{i=1}^{n}\zeta_{ci}\phi_i\zeta_{mi}.$$

Define,

$$D_1(c,m) = A_1(c,m) - \Sigma_0(c,m) = n^{-1}\sum_{i=1}^{n}\big[\zeta_{ci}\phi_i\zeta_{mi} - E\big(\zeta_c\phi\zeta_m\big)\big].$$

and $D_1(c,:) = [D_1(c,1) \;\; D_1(c,2) \;\; \cdots \;\; D_1(c,p)]'$ so that,

$$D_1(c,:)'D_1(c,:) = \sum_{m=1}^{p}\Big(n^{-1}\sum_{i=1}^{n}\big[\zeta_{ci}\phi_i\zeta_{mi} - E\big(\zeta_c\phi\zeta_m\big)\big]\Big)^2.$$

Since for all $c, m \in \{1, 2, \ldots, p\}$, $E\big(\zeta_c \phi \zeta_m\big) = O(1)$ one has,

$$\begin{aligned}
||A_1 - \Sigma_0||^2 = ||D_1||^2 = \text{trace}\big(D_1 D_1'\big) &= \sum_{c=1}^{p} D_1(c,:)' D_1(c,:) \\
&= \sum_{c=1}^{p}\sum_{m=1}^{p} \Big(n^{-1}\sum_{i=1}^{n} \big[\zeta_{ci}\phi_i\zeta_{mi} - E\big(\zeta_c\phi\zeta_m\big)\big]\Big)^2 \\
&= \sum_{c=1}^{p} \Big(n^{-1}\sum_{i=1}^{n} \big[\zeta_{ci}^2\phi_i - E\big(\zeta_c^2\phi\big)\big]\Big)^2 + \sum_{c=1}^{p}\sum_{m\neq c} \Big(n^{-1}\sum_{i=1}^{n} \big[\zeta_{ci}\phi_i\zeta_{mi} - E\big(\zeta_c\phi\zeta_m\big)\big]\Big)^2 \\
&= \sum_{c=1}^{p} o_p(1)^2 + \sum_{c=1}^{p}\sum_{m\neq c} o_p(1)^2 = o_p(1).
\end{aligned}$$

Let $c, m \in \{1, 2, \ldots, p\}$ and define,

$$A_2(c,m) = n^{-1}\sum_{i=1}^{n} \zeta_{ci}\phi_i\big[H^*(Z_{mi}) - \hat{H}^*(Z_{mi})\big].$$

Also let $A_2(c,:) = [A_2(c,1) \quad A_2(c,2) \quad \cdots \quad A_2(c,p)]'$ so that by assumption A5, Theorem 2, and Markov's inequality,

$$\begin{aligned}
A_2(c,:)' A_2(c,:) &= \sum_{m=1}^{p} \Big(n^{-1}\sum_{i=1}^{n} \zeta_{ci}\phi_i\big[H^*(Z_{mi}) - \hat{H}^*(Z_{mi})\big]\Big)^2 \\
&\leq \sum_{m=1}^{p} \sup_{X,V\in G_{XV}} \big[H^*(Z_m) - \hat{H}^*(Z_m)\big]^2 \Big(n^{-1}\sum_{i=1}^{n} |\zeta_{ci}\phi_i|\Big)^2 \\
&= O_p(\mathcal{L}_n^2) O_p(1) = o_p(1).
\end{aligned}$$

Consequently, $||A_2|| = \text{trace}(A_2 A_2')^{1/2} = \Big[\sum_{c=1}^{p} A_2(c,:)' A_2(c,:)\Big]^{1/2} = o_p(1)$. Let $c, m \in \{1, 2, \ldots, p\}$ so that,

$$A_3(c,m) = n^{-1}\sum_{i=1}^{n} \zeta_{ci}(\hat{\phi}_i - \phi_i)\zeta_{mi}.$$

Also let $A_3(c,:) = [A_3(c,1) \quad A_3(c,2) \quad \cdots \quad A_3(c,p)]'$ so that by assumption A5, Theorem 2, and Markov's Inequality,

$$\begin{aligned}
A_3(c,:)' A_3(c,:) &= \sum_{m=1}^{p} \Big(n^{-1}\sum_{i=1}^{n} \zeta_{ci}(\hat{\phi}_i - \phi_i)\zeta_{mi}\Big)^2 \\
&\leq \sum_{m=1}^{p} \sup_{X,V\in G_{XV}} |\hat{\phi}(X,\hat{V}) - \phi(X,V)|^2 \Big(n^{-1}\sum_{i=1}^{n} |\zeta_{ci}\zeta_{mi}|\Big)^2 \\
&= O_p(\mathcal{L}_n^2) O_p(1) = o_p(1).
\end{aligned}$$

Consequently, $||A_3|| = \text{trace}(A_3 A_3')^{1/2} = \Big[\sum_{c=1}^{p} A_3(c,:)' A_3(c,:)\Big]^{1/2} = o_p(1)$. Let $c, m \in \{1, 2, \ldots, p\}$ so that,

$$A_4(c,m) = n^{-1}\sum_{i=1}^{n} \zeta_{ci}(\hat{\phi}_i - \phi_i)[H^*(Z_{mi}) - \hat{H}^*(Z_{mi})]$$

Also let $A_4(c,:) = [A_4(c,1) \;\; A_4(c,2) \;\; \cdots \;\; A_4(c,p)]'$ so that by assumption A5, Theorem 2, and Markov's Inequality,

$$\begin{aligned} A_4(c,:)'A_4(c,:) &= \sum_{m=1}^{p} \Big(n^{-1}\sum_{i=1}^{n} \zeta_{ci}(\hat{\phi}_i - \phi_i)[H^*(Z_{mi}) - \hat{H}^*(Z_{mi})]\Big)^2 \\ &\leq \sum_{m=1}^{p} \sup_{X,V\in G_{XV}} |\hat{\phi}(X,V) - \phi(X,V)|^2 \sup_{X,V\in G_{XV}} |H^*(Z_{mi}) - \hat{H}^*(Z_{mi})|^2 \Big(n^{-1}\sum_{i=1}^{n}|\zeta_{ci}|\Big) \\ &= O_p(\mathcal{L}_{0n}^2)O_p(\mathcal{L}_n^2)O_p(1) = o_p(1) \end{aligned}$$

Consequently, $||A_4|| = \text{trace}(A_4A_4')^{1/2} = \Big[\sum_{c=1}^{p} A_4(c,:)'A_4(c,:)\Big]^{1/2} = o_p(1)$.

$$A_5 = n^{-1}\big[\mathbf{H}_n^*(Z) - \hat{\mathbf{H}}_n^*(Z)\big]'\boldsymbol{\phi}_n\big[\mathbf{Z}_n - \mathbf{H}_n^*(Z)\big].$$

Note the proof of the order of A_5 is, mutatis mutandis, practically identical to the proof of the order of A_2, thus the arguments are not repeated here. Consequently one can conclude that, $||A_5|| = o_p(1)$.
Let $c, m \in \{1, 2, \ldots, p\}$ so that, $A_6(c,m) = n^{-1}\sum_{i=1}^{n}[H^*(Z_{ci}) - \hat{H}^*(Z_{ci})]\phi_i[H^*(Z_{mi}) - \hat{H}^*(Z_{mi})]$. Also let $A_6(c,:) = [A_6(c,1) \;\; A_6(c,2) \;\; \cdots \;\; A_6(c,p)]'$ so that by assumption A5, Theorem 2, and Markov's Inequality,

$$\begin{aligned} A_6(c,:)'A_6(c,:) &= \sum_{m=1}^{p} \Big(n^{-1}\sum_{i=1}^{n}[H^*(Z_{ci}) - \hat{H}^*(Z_{ci})]\phi_i[H^*(Z_{mi}) - \hat{H}^*(Z_{mi})]\Big)^2 \\ &\leq p \max_{1\leq a\leq p} \sup_{X,V\in G_{XV}} |H^*(Z_a) - \hat{H}^*(Z_a)|^4 n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}|\phi_i||\phi_j| \\ &= O_p(\mathcal{L}_n^4)O(1) = o_p(1). \end{aligned}$$

Consequently, $||A_6|| = \text{trace}(A_6A_6')^{1/2} = \Big[\sum_{c=1}^{p} A_6(c,:)'A_6(c,:)\Big]^{1/2} = o_p(1)$.

$$A_7 = n^{-1}\big[\mathbf{H}_n^*(Z) - \hat{\mathbf{H}}_n^*(Z)\big]'\big(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_n\big)\big[\mathbf{Z}_n - \mathbf{H}_n^*(Z)\big].$$

The proof of the order of A_7 is, mutatis mutandis, practically identical to the proof of the order of A_4, thus the arguments are not repeated here. Consequently one can conclude that, $||A_7|| = o_p(1)$. Let $c, m \in \{1, 2, \ldots, p\}$ so that,

$$A_8(c,m) = n^{-1}\sum_{i=1}^{n}[H^*(Z_{ci}) - \hat{H}^*(Z_{ci})]\big(\hat{\phi}_i - \phi_i\big)[H^*(Z_{mi}) - \hat{H}^*(Z_{mi})].$$

Also, let $A_8(c,:) = [A_8(c,1) \;\; A_8(c,2) \;\; \cdots \;\; A_8(c,p)]'$ so that by assumption A5, Theorem 2, and Markov's Inequality,

$$\begin{aligned} A_8(c,:)'A_8(c,:) &= \sum_{m=1}^{p} \Big(n^{-1}\sum_{i=1}^{n}[H^*(Z_{ci}) - \hat{H}^*(Z_{ci})]\big(\hat{\phi}_i - \phi_i\big)[H^*(Z_{mi}) - \hat{H}^*(Z_{mi})]\Big)^2 \\ &\leq p \max_{1\leq a\leq p} \sup_{X,V\in G_{XV}} |H^*(Z_a) - \hat{H}^*(Z_a)|^4 \sup_{X,V\in G_{XV}} |\hat{\phi}(X,V) - \phi(X,V)| \\ &= O_p(\mathcal{L}_n^4)O_p(\mathcal{L}_{0n}^2) = o_p(1). \end{aligned}$$

Consequently, $||A_8|| = \text{trace}(A_8A_8')^{1/2} = \Big[\sum_{c=1}^{p} A_8(c,:)'A_8(c,:)\Big]^{1/2} = o_p(1)$. In all,

$$||A - \Sigma_0|| \leq ||A_1 - \Sigma_0|| + ||A_2|| + ||A_3|| + ||A_4|| + ||A_5|| + ||A_6|| + ||A_7|| + ||A_8|| = o_p(1).$$

Consequently $A = \Sigma_0 + o_p(1)$. Now, recall from Lemma 1 that $\beta_0 = \mu_Y - \mu_Z'\beta_1$ and consider,

$$
\begin{aligned}
&Y_i - \hat{H}^*(Y_i) - \big(Z_i - \hat{H}^*(Z_i)\big)'\beta_1 = Y_i - Z_i\beta_1 - \big(\hat{H}^*(Y_i) - \hat{H}^*(Z_i)'\beta_1\big) \\
&\quad = Y_i - Z_i'\beta_1 - \beta_0 - h(X_i) - f(V_i) + \beta_0 + h(X_i) + f(V_i) - \big(\hat{H}^*(Y_i) - \hat{H}^*(Z_i)'\beta_1\big) \\
&\quad = u_i + \beta_0 + \sum_{d=1}^{D}\big[H_1^d(Y_i) - \beta_0 - H_1^d(Z_i)'\beta_1\big] + \sum_{d=1}^{D}\big[H_2^d(Y_i) - \beta_0 - H_2^d(Z_i)'\beta_1\big] \\
&\qquad\quad - \Big\{\sum_{d=1}^{D}\big[\hat{H}_1^d(Y_i) - \hat{H}_1^d(Z_i)'\beta_1\big] + \sum_{d=1}^{D}\big[\hat{H}_2^d(Y_i) - \hat{H}_2^d(Z_i)'\beta_1\big] + (2D-1)\big[\hat{\mu}_Z'\beta_1 - \hat{\mu}_Y\big]\Big\} \\
&\quad = u_i + \sum_{d=1}^{D}\big[H_1^d(Y_i) - \hat{H}_1^d(Y_i)\big] + \sum_{d=1}^{D}\big[H_2^d(Y_i) - \hat{H}_2^d(Y_i)\big] - \sum_{d=1}^{D}\big[H_1^d(Z_i) - \hat{H}_1^d(Z_i)\big]'\beta_1 \\
&\qquad\quad - \sum_{d=1}^{D}\big[H_2^d(Z_i) - \hat{H}_2^d(Z_i)\big]'\beta_1 - (2D-1)\big[\mu_Y - \hat{\mu}_Y\big] + (2D-1)\big[\mu_Z - \hat{\mu}_Z\big]'\beta_1 \\
&\quad = u_i + \sum_{d=1}^{D}\big[\mathsf{S}_1^d(Y_i) + \mathsf{S}_2^d(Y_i)\big] - \sum_{d=1}^{D}\big[\mathsf{S}_1^d(Z_i) + \mathsf{S}_2^d(Z_i)\big]\beta_1 + (2D-1)\Big(\big[\mu_Z - \hat{\mu}_Z\big]'\beta_1 - \big[\mu_Y - \hat{\mu}_Y\big]\Big).
\end{aligned}
$$

In vector notation,

$$
\begin{aligned}
\mathbf{Y}_n - \hat{\mathbf{H}}_n^*(Y) - \big(\hat{\mathbf{Z}}_n - \hat{\mathbf{H}}_n^*(Z)\big)\beta_1 = \mathbf{u}_n &+ \sum_{d=1}^{D}\big[\mathsf{S}_{1n}^d(Y) - \mathsf{S}_{1n}^d(Z)\beta_1\big] + \sum_{d=1}^{D}\big[\mathsf{S}_{2n}^d(Y) - \mathsf{S}_{2n}^d(Z)\beta_1\big] \\
&+ (2D-1)\Big(\big[\boldsymbol{\mu}_{Zn} - \hat{\boldsymbol{\mu}}_{Z_n}\big]'\beta_1 - \big[\boldsymbol{\mu}_{Yn} - \hat{\boldsymbol{\mu}}_{Yn}\big]\Big).
\end{aligned}
$$

Now applying this definition to B we get,

$$
\begin{aligned}
B &= n^{-1}\big[\mathbf{Z}_n - \hat{\mathbf{H}}_n^*(Z)\big]'\hat{\boldsymbol{\phi}}_n\big[\mathbf{Y}_n - \hat{\mathbf{H}}_n^*(Y) - \big(\mathbf{Z}_n - \hat{\mathbf{H}}_n^*(Z)\big)\beta_1\big] \\
&= n^{-1}\big[\mathbf{Z}_n - \hat{\mathbf{H}}_n^*(Z)\big]'\hat{\boldsymbol{\phi}}_n\mathbf{u}_n + \sum_{d=1}^{D} n^{-1}\big[\mathbf{Z}_n - \hat{\mathbf{H}}_n^*(Z)\big]'\hat{\boldsymbol{\phi}}_n\big[\mathsf{S}_{1n}^d(Y) - \mathsf{S}_{1n}^d(Z)\beta_1\big] \\
&\qquad + \sum_{d=1}^{D} n^{-1}\big[\mathbf{Z}_n - \hat{\mathbf{H}}_n^*(Z)\big]'\hat{\boldsymbol{\phi}}_n\big[\mathsf{S}_{2n}^d(Y) - \mathsf{S}_{2n}^d(Z)\beta_1\big] \\
&\qquad + (2D-1)n^{-1}\big[\mathbf{Z}_n - \hat{\mathbf{H}}_n^*(Z)\big]'\hat{\boldsymbol{\phi}}_n\Big(\big[\boldsymbol{\mu}_{Zn} - \hat{\boldsymbol{\mu}}_{Zn}\big]'\beta_1 - \big[\boldsymbol{\mu}_{Yn} - \hat{\boldsymbol{\mu}}_{Yn}\big]\Big) \\
&\equiv B_1 + \sum_{d=1}^{D} B_{2d} + \sum_{d=1}^{D} B_{3d} + B_4.
\end{aligned}
$$

First consider B_1,

$$
\begin{aligned}
B_1 &= n^{-1}\big[\mathbf{Z}_n - \hat{\mathbf{H}}_n^*(Z)\big]'\hat{\boldsymbol{\phi}}_n\mathbf{u}_n \\
&= n^{-1}\big[\mathbf{Z}_n - \mathbf{H}_n^*(Z)\big]'\boldsymbol{\phi}_n\mathbf{u}_n + n^{-1}\big[\mathbf{Z}_n - \mathbf{H}_n^*(Z)\big]'\big(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_n\big)\mathbf{u}_n \\
&\qquad + n^{-1}\big[\mathbf{H}_n^*(Z) - \hat{\mathbf{H}}_n^*(Z)\big]'\big(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_n\big)\mathbf{u}_n + n^{-1}\big[\mathbf{H}_n^*(Z) - \hat{\mathbf{H}}_n^*(Z)\big]'\boldsymbol{\phi}_n\mathbf{u}_n \\
&\equiv B_{11} + B_{12} + B_{13} + B_{14}.
\end{aligned}
$$

Consider, $B_{11} = n^{-1}\big[\mathbf{Z}_n - \mathbf{H}_n^*(Z)\big]'\boldsymbol{\phi}_n\mathbf{u}_n = n^{-1}\sum_{i=1}^n[Z_i - H^*(Z_i)]\phi_i u_i$. Now, since, $\{Z_i, X_i, V_i\}_{i=1}^n$ is i.i.d,

$$
E\big[\big(Z_i - H^*(Z_i)\big)\phi_i u_i\big] = E\big[\big(Z_i - H^*(Z_i)\big)\phi_i E\big(u_i|Z_i, X_i, V_i\big)\big] = 0,
$$

and, $V\big[\big(Z_i - H^*(Z_i)\big)\phi_i u_i\big] = E\big[E\big(u_i^2|Z_i, X_i, V_i\big)\big(Z_i - H^*(Z_i)\big)\phi_i\big(Z_i - H^*(Z_i)\big)'\big] = \Sigma_1 = O(1)$. Consequently by CLT,

$$\sqrt{n}B_{11} = n^{-1/2}\sum_{i=1}^{n}[Z_i - H^*(Z_i)]\phi_i u_i \xrightarrow{d} N\big(0, \Sigma_1\big).$$

Next consider B_{12} where by Lemma 10 ,

$$\begin{aligned} B_{12} &= n^{-1}\big[\mathbf{Z}_n - \mathbf{H}_n^*(Z)\big]'\big(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_n\big)\mathbf{u}_n = n^{-1}\sum_{i=1}^{n}\big[Z_i - H^*(Z_i)\big]u_i\big(\hat{\phi}_i - \phi_i\big) = n^{-1}\sum_{i=1}^{n}\zeta_i u_i\big(\hat{\phi}_i - \phi_i\big) \\ &= o_p(n^{-1/2}). \end{aligned}$$

Now consider B_{13} where, by Theorem's 1,2 and Lemma 3 xxvi) and xxvii),

$$\begin{aligned} B_{13} &= n^{-1}\big[\mathbf{H}_n^*(Z) - \hat{\mathbf{H}}_n^*(Z)\big]'\big(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_n\big)\mathbf{u}_n = n^{-1}\sum_{i=1}^{n}\big[H^*(Z_i) - \hat{H}^*(Z_i)\big]\big(\hat{\phi}_i - \phi_i\big)u_i. \\ &\le \sup_{XV\in G_{XV}}|H^*(Z) - \hat{H}^*(Z)| \sup_{XV\in G_{XV}}|\hat{\phi}(X,\hat{V}) - \phi(X,V)|n^{-1}\sum_{i=1}^{n}|u_i| \\ &= O_p(\mathcal{L}_n)O_p(\mathcal{L}_{0n})O_p(1) = o_p(n^{-1/2}). \end{aligned}$$

Furthermore consider B_{14} where,

$$\begin{aligned} B_{14} &= n^{-1}\big[\mathbf{H}_n^*(Z) - \hat{\mathbf{H}}_n^*(Z)\big]'\boldsymbol{\phi}_n\mathbf{u}_n \\ &= n^{-1}\sum_{i=1}^{n}\phi_i u_i\Big[\sum_{d=1}^{D}(-1)\big[\hat{H}_1^d(Z_i) - H_1^d(Z_i)\big] \\ &\quad - \sum_{d=1}^{D}\big[\hat{H}_2^d(Z_i) - H_2^d(Z_i)\big] + (2D+1)\big[\big(\mu_Z - \hat{\mu}_Z\big)'\beta_1 - (\mu_Y - \hat{\mu}_Y)\big]\Big] \\ &= (-1)\sum_{d=1}^{D}n^{-1}\sum_{i=1}^{n}\phi_i u_i\big[\hat{H}_1^d(Z_i) - H_1^d(Z_i)\big] - \sum_{d=1}^{D}n^{-1}\sum_{i=1}^{n}\phi_i u_i\big[\hat{H}_2^d(Z_i) - H_2^d(Z_i)\big] \\ &\qquad + (2D+1)\big[\big(\mu_Z - \hat{\mu}_Z\big)'\beta_1 - (\mu_Y - \hat{\mu}_Y)\big]n^{-1}\sum_{i=1}^{n}\phi_i u_i \\ &= (-1)\sum_{d=1}^{D}B_{141d} - \sum_{d=1}^{D}B_{142d} + B_{143}. \end{aligned}$$

As for B_{141d}, by Lemma 11 i.), ii.), iii.), and iv.),

$$\begin{aligned} B_{141d}(c) &= n^{-1}\sum_{i=1}^{n}\phi_i u_i\big[\hat{H}_1^d(Z_{ci}) - H_1^d(Z_{ci})\big] \\ &= n^{-1}\sum_{i=1}^{n}\phi_i u_i\left[[(n-1)b_1]^{-1}\sum_{j\neq i}K_{1ji}(X_d)\hat{\theta}_{1j}^d Z_{cj} - H_1^d(Z_{ci})\right] \\ &= [n(n-1)]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}b_1^{-1}\phi_i u_i K_{1ji}(X_d)\big(\hat{\theta}_{1j}^d - \theta_{1j}^d\big)Z_{cj} \end{aligned}$$

$$+ [n(n-1)]^{-1}\sum_{i=1}^{n}\sum_{j\neq i} b_1^{-1}\phi_i u_i K_{1ji}(X_d)\theta_{1j}^d\rho_{cj} + [n(n-1)]^{-1}\sum_{i=1}^{n}\sum_{j\neq i} b_1^{-1}\phi_i u_i K_{1ji}(X_d)\theta_{1j}^d\eta_{1cj}^d$$
$$+ [n(n-1)]^{-1}\sum_{i=1}^{n}\sum_{j\neq i} b_1^{-1}\phi_i u_i K_{1ji}(X_d)\theta_{1j}^d\big(H_1^d(Z_{cj}) - H_1^d(Z_{ci})\big)$$
$$= o_p(n^{-1/2}).$$

For $c \in \{1, 2, \cdots, p\}$. Now consider $B_{142d}(c)$ where,

$$B_{142d}(c) = n^{-1}\sum_{i=1}^{n}\phi_i u_i\big[\hat{H}_2^d(Z_{ci}) - H_2^d(Z_{ci})\big] = n^{-1}\sum_{i=1}^{n}\phi_i u_i\Big\{[(n-1)b_2]^{-1}\sum_{j\neq i}K_{2ji}(\hat{V}_d)\hat{\theta}_{2j}^d Z_{cj} - H_2^d(Z_{ci})\Big\}$$
$$= n^{-1}\sum_{i=1}^{n}\phi_i u_i\Big\{[(n-1)b_2]^{-1}\sum_{j\neq i}K_{2ji}(\hat{V}_d)\big(\hat{\theta}_{2j}^d - \theta_{2j}^d + \theta_{2j}^d\big)Z_{cj} - H_2^d(Z_{ci})\Big\}$$
$$= n^{-1}\sum_{i=1}^{n}\phi_i u_i[(n-1)b_2]^{-1}\sum_{j\neq i}K_{2ji}(\hat{V}_d)\Big\{\big(\hat{\theta}_{2j}^d - \theta_{2j}^d\big)Z_{cj} + \theta_{2j}^d\big(Z_{cj} - E[Z_{cj}|X_j, V_j]\big)$$
$$+ \theta_{2j}^d\big(E[Z_{cj}|X_j, V_j] - H_2^d(Z_{cj})\big) + \theta_{2j}^d\big(H_2^d(Z_{cj}) - H_2^d(Z_{ci})\big)\Big\}$$
$$= n^{-1}\sum_{i=1}^{n}\phi_i u_i[(n-1)b_2]^{-1}\sum_{j\neq i}K_{2ji}[b_2^{-1}(\hat{V}_{dj} - \hat{V}_{di})]\mathsf{C}_{2ji}^d(c)$$
$$= [n(n-1)]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\phi_i u_i b_2^{-1}K_{2ji}(V_d)\mathsf{C}_{2ji}^d(c)$$
$$- [n(n-1)b_2]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\phi_i u_i b_2^{-1}K_{2ji}^{(1)}(V_d)[\hat{m}_d^{l_n}(W_j) - m_d(W_j)]\mathsf{C}_{2ji}^d(c)$$
$$+ [n(n-1)b_2]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\phi_i u_i b_2^{-1}K_{2ji}^{(1)}(V_d)[\hat{m}_d^{l_n}(W_i) - m_d(W_i)]\mathsf{C}_{2ji}^d(c)$$
$$+ [2n(n-1)b_2^2]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\phi_i u_i b_2^{-1}K_{2ji}^{(2)}(V_d)[\hat{m}_d^{l_n}(W_i) - m_d(W_i) - \big(\hat{m}_d^{l_n}(W_j) - m_d(W_j)\big)\big]^2\mathsf{C}_{2ji}^d(c)$$
$$+ [6n(n-1)b_2^3]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\phi_i u_i b_2^{-1}K_{2ji}^{(3)}(V_d)[\hat{m}_d^{l_n}(W_i) - m_d(W_i) - \big(\hat{m}_d^{l_n}(W_j) - m_d(W_j)\big)\big]^3\mathsf{C}_{2ji}^d(c)$$
$$+ [24n(n-1)b_2^5]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\phi_i u_i K_{2ji}^{(4)}(\tilde{V}_d)[\hat{m}_d^{l_n}(W_i) - m_d(W_i) - \big(\hat{m}_d^{l_n}(W_j) - m_d(W_j)\big)\big]^4\mathsf{C}_{2ji}^d(c)$$
$$\equiv B_{1421d}(c) - B_{1422d}(c) + B_{1423d}(c) + B_{1424d}(c) + B_{1425d}(c) + B_{1426d}(c).$$

The order of $B_{1421d}(c)$ follows from orders derived from a trivial modification of the proof of Lemma 11 i.), ii.), iii.), and iv.) as a result the order of the arguments will not be repeated here. Thus, one can conclude, $B_{1421d}(c) = o_p(n^{-1/2})$. Note that,

$$E\Big[\phi_i u_i H_2^d(Z_{ci})b_2^{-1}K_{2ji}^{(1)}(V_d)\Big] = E\Big[\phi_i u_i H_2^d(Z_{ci})b_2^{-1}K_{2ji}^{(1)}(V_d)E(u_i|Z_i, V_i, X_i, S_j)\Big] = 0,$$

and also, $\phi_i u_i H_2^d(Z_{ci})K_{2ji}^{(1)}(V_d) = K_{2ji}^{(1)}(V_d)\theta_{2i}^d\big[p(V_{di})u_i H_2^d(Z_{ci})\big]$ where,

$$E\big[|p(V_{di})u_i H_2^d(Z_{ci})|^2|V_{di}\big] \leq \Big(\sup_{V_d\in G_{V_d}} p(V_d)|H_2^d(Z_{ci})|\Big)^2 E\big[u_i^2|Z_i, X_i, V_i, S_j\big] < \infty,$$

Consequently, by a trivial modification of Lemma 6

$$\sup_{V_{dj}\in G_{V_d}}\left|[nb_2^2]^{-1}\sum_{i=1}^{n}\left(\phi_i u_i H_2^d(Z_{ci})K_{2ji}^{(1)}(V_d)-E\Big[\phi_i u_i H_2^d(Z_{ci})K_{2ji}^{(1)}(V_d)\Big]\right)\right|=O_p\left(\left[\frac{\log(n)}{nb_2^3}\right]^{1/2}\right).$$

In the trivial case where one sets $H_2^d(Z_{ci})=1$,

$$\sup_{V_{dj}\in G_{V_d}}\left|[nb_2^2]^{-1}\sum_{i=1}^{n}\left(\phi_i u_i K_{2ji}^{(1)}(V_d)-E\Big[\phi_i u_i K_{2ji}^{(1)}(V_d)\Big]\right)\right|=O_p\left(\left[\frac{\log(n)}{nb_2^3}\right]^{1/2}\right).$$

Bearing the preceding two resulting in mind consider,

$$\begin{aligned}
B_{1422d}(c)&=[n(n-1)b_2]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\phi_i u_i b_2^{-1}K_{2ji}^{(1)}(V_d)[\hat{m}_d^{l_n}(W_j)-m_d(W_j)]\mathsf{C}_{2ji}^d(c)\\
&=(n-1)^{-1}\sum_{j\neq i}[\hat{m}_d^{l_n}(W_j)-m_d(W_j)]\big(\hat{\theta}_{2j}^d-\theta_{2j}^d\big)Z_{cj}[nb_2^2]^{-1}\sum_{i=1}^{n}\left(\phi_i u_i K_{2ji}^{(1)}(V_d)-E\Big[\phi_i u_i K_{2ji}^{(1)}(V_d)\Big]\right)\\
&\quad+(n-1)^{-1}\sum_{j\neq i}[\hat{m}_d^{l_n}(W_j)-m_d(W_j)]\theta_{2j}^d\rho_{cj}[nb_2^2]^{-1}\sum_{i=1}^{n}\left(\phi_i u_i K_{2ji}^{(1)}(V_d)-E\Big[\phi_i u_i K_{2ji}^{(1)}(V_d)\Big]\right)\\
&\quad+(n-1)^{-1}\sum_{j\neq i}[\hat{m}_d^{l_n}(W_j)-m_d(W_j)]\theta_{2j}^d\eta_{2cj}^d[nb_2^2]^{-1}\sum_{i=1}^{n}\left(\phi_i u_i K_{2ji}^{(1)}(V_d)-E\Big[\phi_i u_i K_{2ji}^{(1)}(V_d)\Big]\right)\\
&\quad+(n-1)^{-1}\sum_{j\neq i}[\hat{m}_d^{l_n}(W_j)-m_d(W_j)]\theta_{2j}^d H_2^d(Z_{cj})[nb_2^2]^{-1}\sum_{i=1}^{n}\left(\phi_i u_i K_{2ji}^{(1)}(V_d)-E\Big[\phi_i u_i K_{2ji}^{(1)}(V_d)\Big]\right)\\
&\quad-(n-1)^{-1}\sum_{j\neq i}[\hat{m}_d^{l_n}(W_j)-m_d(W_j)]\theta_{2j}^d[nb_2^2]^{-1}\sum_{i=1}^{n}\left(\phi_i u_i H_2^d(Z_{ci})K_{2ji}^{(1)}(V_d)-E\Big[\phi_i u_i H_2^d(Z_{ci})K_{2ji}^{(1)}(V_d)\Big]\right)\\
&\leq\sup_{V_{dj}\in G_{V_d}}\left|[nb_2^2]^{-1}\sum_{i=1}^{n}\left(\phi_i u_i K_{2ji}^{(1)}(V_d)-E\Big[\phi_i u_i K_{2ji}^{(1)}(V_d)\Big]\right)\right|\sup_{W\in G_W}\left|\hat{m}_d^{l_n}(W)-m_d(W)\right|\\
&\qquad\qquad\times\Big\{\sup_{XV\in G_{XV}}|\hat{\theta}_2^d(X,\hat{V})-\theta_2^d(X,V)|(n-1)^{-1}\sum_{j\neq i}|Z_{cj}|\\
&\qquad\qquad\qquad+\sup_{XV\in G_{XV}}\theta_2^d(X,V)(n-1)^{-1}\sum_{j\neq i}\big(|\rho_{cj}|+|\eta_{2cj}^d|+|H_2^d(Z_{cj})|\big)\Big\}\\
&\qquad+\sup_{XV\in G_{XV}}\theta_2^d(X,V)\sup_{W\in G_W}\left|\hat{m}_d^{l_n}(W)-m_d(W)\right|\\
&\qquad\qquad\times\sup_{V_{dj}\in G_{V_d}}\left|[nb_2^2]^{-1}\sum_{i=1}^{n}\left(\phi_i u_i H_2^d(Z_{ci})K_{2ji}^{(1)}(V_d)-E\Big[\phi_i u_i H_2^d(Z_{ci})K_{2ji}^{(1)}(V_d)\Big]\right)\right|\\
&=O_p(L_n)O_p\left(\left[\frac{\log(n)}{nb_2^3}\right]^{1/2}\right)\Big\{O_p(\mathcal{L}_{0n})+O(1)\Big\}=O_p(L_n b_2^{-1})O_p\left(\left[\frac{\log(n)}{nb_2^{1/2}}\right]^{1/2}\right)=o_p(n^{-1/2}),
\end{aligned}$$

by Theorem 1, Lemma 3 , and Lemma 6 . Consider

$$\begin{aligned}
B_{1423d}(c)&=[n(n-1)b_2]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\phi_i u_i b_2^{-1}K_{2ji}^{(1)}(V_d)[\hat{m}_d^{l_n}(W_i)-m_d(W_i)]\mathsf{C}_{2ji}^d(c)\\
&=[n(n-1)b_2]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\phi_i u_i b_2^{-1}K_{2ji}^{(1)}(V_d)[\hat{m}_d^{l_n}(W_i)-m_d(W_i)](\hat{\theta}_{2j}^d-\theta_{2j}^d)Z_{cj}
\end{aligned}$$

$$+ [n(n-1)b_2]^{-1} \sum_{i=1}^{n} \sum_{j \neq i} \phi_i u_i b_2^{-1} K_{2ji}^{(1)}(V_d)[\hat{m}_d^{l_n}(W_i) - m_d(W_i)]\theta_{2j}^d \rho_{cj}$$

$$+ [n(n-1)b_2]^{-1} \sum_{i=1}^{n} \sum_{j \neq i} \phi_i u_i b_2^{-1} K_{2ji}^{(1)}(V_d)[\hat{m}_d^{l_n}(W_i) - m_d(W_i)]\theta_{2j}^d \eta_{2cj}^d$$

$$+ [n(n-1)b_2]^{-1} \sum_{i=1}^{n} \sum_{j \neq i} \phi_i u_i b_2^{-1} K_{2ji}^{(1)}(V_d)[\hat{m}_d^{l_n}(W_i) - m_d(W_i)]\theta_{2j}^d \big(H_2^d(Z_{cj}) - H_2^d(Z_{ci})\big)$$

$$= (n-1)^{-1} \sum_{j \neq i} \mathsf{C}_{2j}^{d*}(c)[nb_2^2]^{-1} \sum_{i=1}^{n} \phi_i u_i K_{2ji}^{(1)}(V_d)[\hat{m}_d^{l_n}(W_i) - m_d(W_i)]$$

$$- (n-1)^{-1} \sum_{j \neq i} \theta_{2j}^d [nb_2^2]^{-1} \sum_{i=1}^{n} \phi_i u_i H_2^d(Z_{ci}) K_{2ji}^{(1)}(V_d)[\hat{m}_d^{l_n}(W_i) - m_d(W_i)]$$

$$= (n-1)^{-1} \sum_{j=1}^{n} \mathsf{C}_{2j}^{d*}(c)[nb_2^2]^{-1} \sum_{i \neq j} \phi_i u_i K_{2ji}^{(1)}(V_d)[\hat{m}_d^{l_n}(W_i) - m_d(W_i)]$$

$$- (n-1)^{-1} \sum_{j=1}^{n} \theta_{2j}^d [nb_2^2]^{-1} \sum_{i \neq j} \phi_i u_i H_2^d(Z_{ci}) K_{2ji}^{(1)}(V_d)[\hat{m}_d^{l_n}(W_i) - m_d(W_i)]$$

$$= (n-1)^{-1} \sum_{j=1}^{n} \mathsf{C}_{2j}^{d*}(c)[nb_2^2]^{-1} \mathbf{K}_{2j}(V_d)' \boldsymbol{\phi}_n \dot{\boldsymbol{u}}_n I(-j)[\hat{\mathbf{M}}_d^{l_n} - \mathbf{M}_d]$$

$$- (n-1)^{-1} \sum_{j=1}^{n} \theta_{2j}^d [nb_2^2]^{-1} \mathbf{K}_{2j}(V_d)' \boldsymbol{\phi}_n \dot{\boldsymbol{u}}_n \dot{\mathbf{H}}_2^d(Z_c) I(-j)\big[\hat{\mathbf{M}}_d^{l_n} - \mathbf{M}_d\big]$$

$$\leq n^{-1} \sum_{j=1}^{n} \left|\mathsf{C}_{2j}^{d*}(c)\right| \, ||[(n-1)b_2^2]^{-1} \mathbf{K}_{2j}(V_d)' \boldsymbol{\phi}_n \dot{\boldsymbol{u}}_n I(-j) \mathbf{B}_n||_E O_p\left(\sqrt{\frac{l_n}{n}} + l_n^{-k}\right)$$

$$- [n(n-1)]^{-1} \sum_{j=1}^{n} \sum_{i \neq j} \mathsf{C}_{2j}^{d*}(c) \phi_i u_i b_2^{-2} K_{2ji}^{(1)}(V_d)[m_d^{l_n}(W_i) - m_d(W_i)]$$

$$- n^{-1} \sum_{j=1}^{n} ||[(n-1)b_2^2]^{-1} \mathbf{K}_{2j}(V_d)' \boldsymbol{\phi}_n \dot{\boldsymbol{u}}_n \dot{\mathbf{H}}_2^d(Z_c) I(-j) \mathbf{B}_n||_E O_p\left(\sqrt{\frac{l_n}{n}} + l_n^{-k}\right)$$

$$+ [n(n-1)]^{-1} \sum_{j=1}^{n} \sum_{i \neq j} \phi_i u_i b_2^{-2} H_2^d(Z_{ci}) K_{2ji}^{(1)}(V_d)[m_d^{l_n}(W_i) - m_d(W_i)]$$

$$\equiv B_{14231d}(c) + B_{14232d}(c) + B_{14233d}(c) + B_{14234d}(c).$$

Note that,

$$B_{14231d}(c) = n^{-1} \sum_{j=1}^{n} \left|\mathsf{C}_{2j}^{d*}(c)\right| \, ||[(n-1)b_2^2]^{-1} \mathbf{K}_{2j}(V_d)' \boldsymbol{\phi}_n \dot{\boldsymbol{u}}_n I(-j) \mathbf{B}_n||_E O_p\left(\sqrt{\frac{l_n}{n}} + l_n^{-k}\right).$$

consider,

$$E\big(||[(n-1)b_2^2]^{-1} \mathbf{K}_{2j}(V_d)' \boldsymbol{\phi}_n \dot{\boldsymbol{u}}_n I(-j) \mathbf{B}_n||_E^2\big) = E\Big(||[(n-1)b_2^2]^{-1} \sum_{i \neq j} \mathbf{B}_n(W_i) K_{2ji}(V_d) \phi_i u_i||_E\Big)$$

$$= E\Big([(n-1)b_2^2]^{-2} \sum_{i \neq j} \sum_{g \neq j} \mathbf{B}_n(W_g)' \mathbf{B}_n(W_i) K_{2gi}(V_d) K_{2ji}(V_d) \phi_g u_g \phi_i u_i\Big)$$

$$
\begin{aligned}
&= [(n-1)^2 b_2^3]^{-1} \sum_{i \neq j} E\Big(\mathbf{B}_n(W_i)'\mathbf{B}_n(W_i) b_2^{-1} K_{2ji}(V_d)^2 \phi_j^2 E\big[u_j^2 | W_j, X_j, V_j, S_i\big]\Big) \\
&\quad + [(n-1)b_2^2]^{-2} \sum_{i \neq j} \sum_{\substack{g \neq j \\ g \neq i}} E\Big(\mathbf{B}_n(W_g)'\mathbf{B}_n(W_i) K_{2gi}(V_d) K_{2ji}(V_d) \phi_g \phi_i \\
&\qquad\qquad \times E[u_g | W_g, X_g, V_g, S_{-g}] E[u_i | W_i, X_i, V_i, S_{-i}]\Big) \\
&\leq \sup_{X, V \in G_{XV}} |\phi(X,V)| O([(n-1)b_2^3]^{-1}) E\Big(\mathbf{B}_n(W_i)'\mathbf{B}_n(W_i) E[b_2^{-1} K_{2ji}(V_d)^2 | S_i]\Big) \\
&= O([(n-1)b_2^3]^{-1}) E\Big(\mathbf{B}_n(W_i)'\mathbf{B}_n(W_i)\Big) = O\left(\frac{l_n}{(n-1)b_2^3}\right).
\end{aligned}
$$

Consequently by Lemma 3 , and Markov's Inequality,

$$
\begin{aligned}
B_{14231d}(c) &= O_p\left(\frac{l_n}{n b_2^{3/2}}\right) n^{-1} \sum_{j=1}^n \big|\mathsf{C}_{2j}^{d*}(c)\big| \\
&= o_p(n^{-1/2}) n^{-1} \sum_{j=1}^n \big|\mathsf{C}_{2j}^{d*}(c)\big| \\
&\leq o_p(n^{-1/2}) \Big\{ \sup_{XV \in G_{XV}} |\hat{\theta}_2^d(X, \hat{V}) - \theta_2^d(X,V)| n^{-1} \sum_{j=1}^n |Z_{cj}| \\
&\qquad + n^{-1} \sum_{j=1}^n |\theta_{2j}^d \rho_{cj}| + n^{-1} \sum_{j=1}^n |\theta_{2j}^d \eta_{2cj}^d| + n^{-1} \sum_{j=1}^n |\theta_{2j}^d H_2^d(Z_{cj})| \Big\} \\
&= o_p(n^{-1/2}).
\end{aligned}
$$

Let $Q_{2j}^d(c) = \rho_{cj} + \eta_{2cj}^d + H_2^d(Z_{cj})$ and note that,

$$
\begin{aligned}
B_{14232d}(c) &= [n(n-1)]^{-1} \sum_{j=1}^n \sum_{i \neq j} \mathsf{C}_{2j}^{d*}(c) \phi_i u_i b_2^{-2} K_{2ji}^{(1)}(V_d) [m_d^{l_n}(W_i) - m_d(W_i)] \\
&= [n(n-1)]^{-1} \sum_{j=1}^n \sum_{i \neq j} (\hat{\theta}_{2j}^d - \theta_{2j}^d) Z_{cj} \phi_i u_i b_2^{-2} K_{2ji}^{(1)}(V_d) [m_d^{l_n}(W_i) - m_d(W_i)] \\
&\quad + [n(n-1)]^{-1} \sum_{j=1}^n \sum_{i \neq j} \theta_{2j}^d \rho_{cj} \phi_i u_i b_2^{-2} K_{2ji}^{(1)}(V_d) [m_d^{l_n}(W_i) - m_d(W_i)] \\
&\quad + [n(n-1)]^{-1} \sum_{j=1}^n \sum_{i \neq j} \theta_{2j}^d \eta_{2cj}^d \phi_i u_i b_2^{-2} K_{2ji}^{(1)}(V_d) [m_d^{l_n}(W_i) - m_d(W_i)] \\
&\quad + [n(n-1)]^{-1} \sum_{j=1}^n \sum_{i \neq j} \theta_{2j}^d H_2^d(Z_{cj}) \phi_i u_i b_2^{-2} K_{2ji}^{(1)}(V_d) [m_d^{l_n}(W_i) - m_d(W_i)] \\
&\leq \sup_{X, V \in G_{XV}} |\hat{\theta}_2^d(X, \hat{V}) - \theta_2^d(X,V)| \sup_{W \in G_W} |m_d^{l_n}(W) - m_d(W)| \\
&\qquad \times n^{-1} \sum_{j=1}^n |Z_{cj}| [(n-1)b_2]^{-1} \sum_{i \neq j} |\phi_i u_i| \big| b_2^{-1} K_{2ji}^{(1)}(V_d) \big| \\
&\quad + [n(n-1)]^{-1} \sum_{j=1}^n \sum_{i \neq j} \theta_{2j}^d \big[\rho_{cj} + \eta_{2cj}^d + H_2^d(Z_{cj})\big] \phi_i u_i b_2^{-2} K_{2ji}^{(1)}(V_d) [m_d^{l_n}(W_i) - m_d(W_i)]
\end{aligned}
$$

$$
\begin{aligned}
&\equiv O_p(\mathcal{L}_{0n})O(L_n)O_p(b_2^{-1})\\
&\qquad + [n(n-1)]^{-1}\sum_{j=1}^{n}\sum_{i\neq j}^{n}\theta_{2j}^{d}Q_{2j}^{d}(c)\phi_i u_i b_2^{-2}K_{2ji}^{(1)}(V_d)[m_d^{l_n}(W_i)-m_d(W_i)]\\
&\equiv o_p(n^{-1/2})
\end{aligned}
$$

By Theorem 1, Lemma 3 vi.), xxvi.) and Lemma 11 v.). The proof of the order of $B_{14233d}(c)$ and $B_{14234d}(c)$ follows mutatis mutandis (omitting $\mathsf{C}_{2j}^{d*}(c)$) directly from the proof of the order of $B_{14231d}(c)$, and $B_{14232d}(c)$ thus,

$$
B_{14233d}(c) + B_{14234d}(c) = o_p(n^{-1/2}).
$$

As a result,

$$
B_{1423d}(c) = B_{14231d}(c) + B_{14232d}(c) + B_{14233d}(c) + B_{14234d}(c) = o_p(n^{-1/2}).
$$

By Lemma 3 vi) and vii) we note that,

$$
\begin{aligned}
&B_{1424d}(c) + B_{1425d}(c) + B_{1426d}(c)\\
&= [2n(n-1)b_2^2]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\phi_i u_i K_{2ji}^{(2)}(V_d)\big[\hat{m}_d^{l_n}(W_i)-m_d(W_i)-\big(\hat{m}_d^{l_n}(W_j)-m_d(W_j)\big)\big]^2\mathsf{C}_{2ji}^{d}(c)\\
&\qquad + [6n(n-1)b_2^3]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\phi_i u_i K_{2ji}^{(3)}(V_d)\big[\hat{m}_d^{l_n}(W_i)-m_d(W_i)-\big(\hat{m}_d^{l_n}(W_j)-m_d(W_j)\big)\big]^3\mathsf{C}_{2ji}^{d}(c)\\
&\qquad + [24n(n-1)b_2^5]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\phi_i u_i K_{2ji}^{(4)}(\tilde{V}_d)\big[\hat{m}_d^{l_n}(W_i)-m_d(W_i)-\big(\hat{m}_d^{l_n}(W_j)-m_d(W_j)\big)\big]^4\mathsf{C}_{2ji}^{d}(c)\\
&\le \sup_{W\in G_W} 2^2|\hat{m}_d^{l_n}(W)-m_d(W)|^2[2n(n-1)b_2^2]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}|\phi_i u_i|\big|b_2^{-1}K_{2ji}^{(2)}(V_d)\big|\big|\mathsf{C}_{2ji}^{d}(c)\big|\\
&\qquad + \sup_{W\in G_W} 2^3|\hat{m}_d^{l_n}(W)-m_d(W)|^3[6n(n-1)b_2^3]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}|\phi_i u_i|\big|b_2^{-1}K_{2ji}^{(3)}(V_d)\big|\big|\mathsf{C}_{2ji}^{d}(c)\big|\\
&\qquad + \sup_{W\in G_W} 2^4|\hat{m}_d^{l_n}(W)-m_d(W)|^4\sup_{\gamma\in\mathbb{R}}\big|K_{2ji}^{(4)}(\gamma)\big|[24n(n-1)b_2^5]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}|\phi_i u_i|\big|\mathsf{C}_{2ji}^{d}(c)\big|\\
&= O_p\Big([L_n b_2^{-1}]^2\Big)[n(n-1)]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}|\phi_i u_i|\big|b_2^{-1}K_{2ji}^{(2)}(V_d)\big|\big|\mathsf{C}_{2ji}^{d}(c)\big|\\
&\qquad + O_p\Big([L_n b_2^{-1}]^3\Big)[n(n-1)]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}|\phi_i u_i|\big|b_2^{-1}K_{2ji}^{(3)}(V_d)\big|\big|\mathsf{C}_{2ji}^{d}(c)\big|\\
&\qquad + O_p\Big([L_n^4 b_2^{-5}]\Big)[n(n-1)]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}|\phi_i u_i|\big|\mathsf{C}_{2ji}^{d}(c)\big|\\
&= o_p(n^{-1/2})[n(n-1)]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}|\phi_i u_i|\Big(\big|b_2^{-1}K_{2ji}^{(2)}(V_d)\big| + \big|b_2^{-1}K_{2ji}^{(3)}(V_d)\big| + 1\Big)\big|\mathsf{C}_{2ji}^{d}(c)\big|\\
&\le o_p(n^{-1/2})[n(n-1)]^{-1}\sup_{X,V\in G_{XV}}|\hat{\theta}_2^{d}(X,\hat{V})-\theta_2^{d}(X,V)|\\
&\qquad \times\sum_{i=1}^{n}\sum_{j\neq i}|\phi_i u_i|\Big(\big|b_2^{-1}K_{2ji}^{(2)}(V_d)\big| + \big|b_2^{-1}K_{2ji}^{(3)}(V_d)\big| + 1\Big)|Z_{cj}|
\end{aligned}
$$

$$
\begin{aligned}
&+ o_p(n^{-1/2})[n(n-1)]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}|\phi_i u_i|\Big(\big|b_2^{-1}K_{2ji}^{(2)}(V_d)\big| + \big|b_2^{-1}K_{2ji}^{(3)}(V_d)\big| + 1\Big)\big|\theta_{2j}^d Q_{2j}^d(c)\big| \\
&+ o_p(n^{-1/2})[n(n-1)]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}|\phi_i u_i|\Big(\big|b_2^{-1}K_{2ji}^{(2)}(V_d)\big| + \big|b_2^{-1}K_{2ji}^{(3)}(V_d)\big| + 1\Big)\big|\theta_{2j}^d H_2^d(Z_{ci})\big| \\
&= o_p(n^{-1/2})
\end{aligned}
$$

By Theorem 1, Lemma 11 vi.) vii.), and viii.). Consequently,

$$
B_{142d}(c) = B_{1421d}(c) + B_{1422d}(c) + B_{1423d}(c) + B_{1424d}(c) + B_{1425d}(c) + B_{1426d}(c) = o_p(n^{1/2}).
$$

Now,

$$
\begin{aligned}
B_{143} &= (2D+1)\big[\big(\mu_Z - \hat{\mu}_Z\big)\beta_1 - (\mu_Y - \hat{\mu}_Y)\big]n^{-1}\sum_{i=1}^{n}\phi_i u_i \\
&\leq (2D+1)\Big[|\mu_Y - \hat{\mu}_Y| + \max_{1\leq c\leq p}|\mu_{Z_c} - \hat{\mu}_{Z_c}|\sum_{c=1}^{p}|\beta_1|\Big]\Big|n^{-1}\sum_{i=1}^{n}\phi_i u_i\Big| \\
&= O_p\big(\mathcal{L}_{0n}\big)\big(1 + ||\beta_1||_E\big)\Big|\sum_{i=1}^{n}\phi_i u_i\Big| = O_p\big(\mathcal{L}_{0n}\big)O(1)o_p(n^{-1/2}) = o_p(n^{-1/2}).
\end{aligned}
$$

Hence,

$$
B_{14} = (-1)\sum_{d=1}^{D}B_{141d} - \sum_{d=1}^{D}B_{142d} + B_{143} = o_p(n^{-1/2}).
$$

Thus in summary,

$$
\sqrt{n}B_1 = \sqrt{n}B_{11} + \sqrt{n}B_{12} + \sqrt{n}B_{13} + \sqrt{n}B_{14} = \sqrt{n}B_{11} + o_p(1).
$$

Consequently, $\sqrt{n}B_1 \xrightarrow{d} N(0, \Sigma_1)$.

$$
\begin{aligned}
B_2 &= \sum_{d=1}^{D} n^{-1}\big[\mathbf{Z}_n - \hat{\mathbf{H}}_n^*(Z)\big]'\hat{\boldsymbol{\phi}}_n\big[\mathsf{S}_{1n}^d(Y) - \mathsf{S}_{1n}^d(Z)\beta_1\big] \\
&= \sum_{d=1}^{D} n^{-1}\big[\mathbf{Z}_n - \mathbf{H}_n^*(Z) + \mathbf{H}_n^*(Z) - \hat{\mathbf{H}}_n^*(Z)\big]'\big(\boldsymbol{\phi}_n + \hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_n\big)\big[\mathsf{S}_{1n}^d(Y) - \mathsf{S}_{1n}^d(Z)\beta_1\big] \\
&= \sum_{d=1}^{D}\Big\{n^{-1}\big[\mathbf{Z}_n - \mathbf{H}_n^*(Z)\big]'\boldsymbol{\phi}_n\big[\mathsf{S}_{1n}^d(Y) - \mathsf{S}_{1n}^d(Z)\beta_1\big] \\
&\qquad + n^{-1}\big[\mathbf{Z}_n - \mathbf{H}_n^*(Z)\big]'\big(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_n\big)\big[\mathsf{S}_{1n}^d(Y) - \mathsf{S}_{1n}^d(Z)\beta_1\big] \\
&\qquad + n^{-1}\big[\mathbf{H}_n^*(Z) - \hat{\mathbf{H}}_n^*(Z)\big]'\boldsymbol{\phi}_n\big[\mathsf{S}_{1n}^d(Y) - \mathsf{S}_{1n}^d(Z)\beta_1\big] \\
&\qquad + n^{-1}\big[\mathbf{H}_n^*(Z) - \hat{\mathbf{H}}_n^*(Z)\big]'\big(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_n\big)\big[\mathsf{S}_{1n}^d(Y) - \mathsf{S}_{1n}^d(Z)\beta_1\big]\Big\} \\
&\equiv \sum_{d=1}^{D}\big[B_{21d} + B_{22d} + B_{23d} + B_{24d}\big].
\end{aligned}
$$

Consider,

$$
B_{22d} = n^{-1}\big[\mathbf{Z}_n - \mathbf{H}_n^*(Z)\big]\big(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_n\big)\big[\mathsf{S}_{1n}^d(Y) - \mathsf{S}_{1n}^d(Z)\beta_1\big]
$$

$$
\begin{aligned}
&= n^{-1}\sum_{i=1}^{n}\big[Z_i - H^*(Z_i)\big]\big(\hat{\phi}_i - \phi_i\big)\Big(H_1^d(Y_i) - \hat{H}_1^d(Y_i) - \sum_{c=1}^{p}\big[H_1^d(Z_{ci}) - \hat{H}_1^d(Z_{ci})\big]\beta_{1c}\Big) \\
&\leq \sup_{XV\in G_{XV}} |\hat{\phi}(X,\hat{V}) - \phi(X,V)|\Big[\sup_{X_d\in G_{X_d}}\big|H_1^d(Y_i) - \hat{H}_1^d(Y_i)\big| \\
&\qquad\qquad + \max_{1\leq j\leq p}\sup_{X_d\in G_{X_d}}\big|H_1^d(Z_j) - \hat{H}_1^d(Z_j)\big|\sum_{c=1}^{p}|\beta_{1c}|\Big]n^{-1}\sum_{i=1}^{n}\big|Z_i - H^*(Z_i)\big| \\
&\leq O_p\big(\mathcal{L}_{0n}\big)O_p\big(\mathcal{L}_{1n}\big)\big(1 + ||\beta_1||_E\big)O_p(1) = O_p\big(\mathcal{L}_{0n}\mathcal{L}_{1n}\big) = o_p(n^{-1/2}).
\end{aligned}
$$

By Theorem 1, Theorem 2, Lemma 3 xxvi) and xxvii), Assumption A3, and Markov's inequality.

$$
\begin{aligned}
B_{23d}(c) &= n^{-1}\big[\mathbf{H}_n^*(Z_c) - \hat{\mathbf{H}}_n^*(Z_c)\big]'\boldsymbol{\phi}_n\big[\mathsf{S}_{1n}^d(Y) - \mathsf{S}_{1n}^d(Z)\beta_1\big] \\
&= n^{-1}\sum_{i=1}^{n}\big[H^*(Z_{ci}) - \hat{H}^*(Z_{ci})\big]\phi_i\Big(H_1^d(Y_i) - \hat{H}_1^d(Y_i) - \sum_{c=1}^{p}\big[H_1^d(Z_{ci}) - \hat{H}_1^d(Z_{ci})\big]\beta_{1c}\Big) \\
&\leq \sup_{XV\in G_{XV}} |H^*(Z_c) - \hat{H}^*(Z_c)|\Big[\sup_{X_d\in G_{X_d}}\big|H_1^d(Y) - \hat{H}_1^d(Y)\big| \\
&\qquad\qquad + \max_{1\leq j\leq p}\sup_{X_d\in G_{X_d}}\big|H_1^d(Z_j) - \hat{H}_1^d(Z_j)\big|\sum_{c=1}^{p}|\beta_{1c}|\Big]n^{-1}\sum_{i=1}^{n}\phi_i \\
&\leq O_p\big(\mathcal{L}_n\mathcal{L}_{1n}\big)\big(1 + ||\beta_1||_E\big)O(1) = O_p\big(\mathcal{L}_n\mathcal{L}_{1n}\big) = o_p(n^{-1/2}),
\end{aligned}
$$

By Theorem 2, Lemma 3 xxvi) and xxvii), Assumption A3, and Markov's inequality.

$$
\begin{aligned}
B_{24d}(c) &= n^{-1}\big[\mathbf{H}_n^*(Z_c) - \hat{\mathbf{H}}_n^*(Z_c)\big]'\big(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_n\big)\big[\mathsf{S}_{1n}^d(Y) - \mathsf{S}_{1n}^d(Z)\beta_1\big] \\
&= n^{-1}\sum_{i=1}^{n}\big[H^*(Z_{ci}) - \hat{H}^*(Z_{ci})\big]\big(\hat{\phi}_i - \phi_i\big)\Big(H_1^d(Y_i) - \hat{H}_1^d(Y_i) - \sum_{c=1}^{p}\big[H_1^d(Z_{ci}) - \hat{H}_1^d(Z_{ci})\big]\beta_{1c}\Big) \\
&\leq \sup_{XV\in G_{XV}} |\hat{\phi}(X,\hat{V}) - \phi(X,V)|\sup_{XV\in G_{XV}} |H^*(Z_c) - \hat{H}^*(Z_c)|\Big[\sup_{X_d\in G_{X_d}}\big|H_1^d(Y) - \hat{H}_1^d(Y)\big| \\
&\qquad\qquad + \max_{1\leq j\leq p}\sup_{X_d\in G_{X_d}}\big|H_1^d(Z_j) - \hat{H}_1^d(Z_j)\big|\sum_{c=1}^{p}|\beta_{1c}|\Big] \\
&= O_p\big(\mathcal{L}_n\big)O_p\big(\mathcal{L}_{0n}\big)O_p\big(\mathcal{L}_{1n}\big)\big(1 + ||\beta_1||_E\big) = O_p\big(\mathcal{L}_n\big)O_p\big(\mathcal{L}_{0n}\big)O_p\big(\mathcal{L}_{1n}\big) = o_p(n^{-1/2}),
\end{aligned}
$$

By Theorem 1, Theorem 2, Lemma 3 xxvi) and xxvii), Assumption A3, and Markov's inequality.

$$
\begin{aligned}
B_{21d} &= n^{-1}\big[\mathbf{Z}_n - \mathbf{H}_n^*(Z)\big]\boldsymbol{\phi}_n\big[\mathsf{S}_{1n}^d(Y) - \mathsf{S}_{1n}^d(Z)\beta_1\big] \\
&= n^{-1}\sum_{i=1}^{n}\big[Z_i - H^*(Z_i)\big]\phi_i\Big(H_1^d(Y_i) - \hat{H}_1^d(Y_i) - \sum_{c=1}^{p}\big[H_1^d(Z_{ci}) - \hat{H}_1^d(Z_{ci})\big]\beta_{1c}\Big) \\
&= (-1)n^{-1}\sum_{i=1}^{n}\big[Z_i - H^*(Z_i)\big]\phi_i\big(\hat{H}_1^d(Y_i) - H_1^d(Y_i)\big) \\
&\qquad + \sum_{c=1}^{p}\beta_{1c}\, n^{-1}\sum_{i=1}^{n}\big[Z_i - H^*(Z_i)\big]\phi_i\big[\hat{H}_1^d(Z_{ci}) - H_1^d(Z_{ci})\big] \\
&= B_{211d} - \sum_{c=1}^{p}\beta_{1c}B_{212dc}.
\end{aligned}
$$

Note that by Lemma 2 , $E[\phi_i\zeta_{ai}K_{1ji}(X_d)|X_{di},S_j] = K_{1ji}(X_d)E[\phi_i\zeta_{ai}|X_{di}] = 0$. Furthermore, since by Assumption 4, $E\big[\phi_i^2\zeta_{ci}^2|X_{di}\big] = p(X_d)^2E\big[[\theta_{1i}^d]^2\zeta_{ci}^2|X_{di}\big]$ is uniformly bounded we have by Lemma 6 ,

$$\sup_{X_{di}\in G_{X_d}}\Big|n^{-1}\sum_{i=1}^n\Big[\phi_i\zeta_{ai}b_1^{-1}K_{1ji}(X_d) - E\Big(\phi_i\zeta_{ai}b_1^{-1}K_{1ji}(X_d)\Big)\Big]\Big| = O_p\left(\left[\frac{\log(n)}{nb_1}\right]^{1/2}\right).$$

Consider,

$$\begin{aligned}
B_{212dc}(a) &= n^{-1}\sum_{i=1}^n\big[Z_{ai} - H^*(Z_{ai})\big]\phi_i\big[\hat{H}_1^d(Z_{ci}) - H_1^d(Z_{ci})\big]\\
&= n^{-1}\sum_{i=1}^n\phi_i\zeta_{ai}\Big\{[(n-1)b_1]^{-1}\sum_{j\neq i}K_{1ji}(X_d)\hat{\theta}_{1j}^dZ_{ci} - H_1^d(Z_{ci})\Big\}\\
&= n^{-1}\sum_{i=1}^n\phi_i\zeta_{ai}[(n-1)b_1]^{-1}\sum_{j\neq i}K_{1ji}(X_d)\Big[\big(\hat{\theta}_{1j}^d - \theta_{1j}^d\big)Z_{cj} + \theta_{1j}^d\big(Z_{cj} - E[Z_{cj}|X_j,V_j]\big)\\
&\qquad + \theta_{1j}^d\big(E[Z_{cj}|X_j,V_j] - H_1^d(Z_{cj})\big) + \theta_{1j}^d\big(H_1^d(Z_{cj}) - H_1^d(Z_{ci})\big)\Big]\\
&= [n(n-1)]^{-1}\sum_{j\neq i}\big(\hat{\theta}_{1j}^d - \theta_{1j}^d\big)Z_{cj}\sum_{i=1}^n\phi_i\zeta_{ai}b_1^{-1}K_{1ji}(X_d)\\
&\quad + [n(n-1)]^{-1}\sum_{i=1}^n\sum_{j\neq i}\phi_i\zeta_{ai}b_1^{-1}K_{1ji}(X_d)\theta_{1j}^d\rho_{cj} + [n(n-1)]^{-1}\sum_{i=1}^n\sum_{j\neq i}\phi_i\zeta_{ai}b_1^{-1}K_{1ji}(X_d)\theta_{1j}^d\eta_{1cj}^d\\
&\quad + [n(n-1)]^{-1}\sum_{i=1}^n\sum_{j\neq i}\phi_i\zeta_{ai}b_1^{-1}K_{1ji}(X_d)\theta_{1j}^d\big(H_1^d(Z_{cj}) - H_1^d(Z_{ci})\big)\\
&= o_p(n^{-1/2})
\end{aligned}$$

Lemma 11 ix.), x.),xi.),xii.). The proof of the order of B_{211d} is, mutatis mutandis, virtually identical to proof of the order of B_{212d}, thus one can conclude that, $B_{211d} = o_p(n^{-1/2})$. Furthermore, by Assumption A3,

$$B_{21d} = B_{211d} - \sum_{c=1}^p\beta_{1c}B_{212dc} \leq O_p(n^{-1/2})(1 + ||\beta_1||_E) = o_p(n^{-1/2}).$$

Consequently, $B_2 = \sum_{d=1}^D\big[B_{21d} + B_{22d} + B_{23d} + B_{24d}\big] = o_p(n^{-1/2})$.

$$\begin{aligned}
B_{3d} &= n^{-1}\big[\mathbf{Z}_n - \hat{\mathbf{H}}_n^*(Z)\big]'\hat{\boldsymbol{\phi}}_n\big[\mathsf{S}_{2n}^d(Y) - \mathsf{S}_{2n}^d(Z)\beta_1\big]\\
&= n^{-1}\sum_{i=1}^n\big[Z_i - H^*(Z_i) + H^*(Z_i) - \hat{H}^*(Z_i)\big]\big(\phi_i + \hat{\phi}_i - \phi_i\big)\big[\mathsf{S}_2^d(Y_i) - \mathsf{S}_2^d(Z_i)'\beta_1\big]\\
&= n^{-1}\sum_{i=1}^n\big[Z_i - H^*(Z_i)\big]\phi_i\Big(H_2^d(Y_i) - \hat{H}_2^d(Y_i) - \sum_{c=1}^p\big[H_2^d(Z_{ci}) - \hat{H}_2^d(Z_{ci})\big]\beta_{1c}\Big)\\
&\quad + n^{-1}\sum_{i=1}^n\big[Z_i - H^*(Z_i)\big]\big(\hat{\phi}_i - \phi_i\big)\Big(H_2^d(Y_i) - \hat{H}_2^d(Y_i) - \sum_{c=1}^p\big[H_2^d(Z_{ci}) - \hat{H}_2^d(Z_{ci})\big]\beta_{1c}\Big)\\
&\quad + n^{-1}\sum_{i=1}^n\big[H^*(Z_i) - \hat{H}^*(Z_i)\big]\phi_i\Big(H_2^d(Y_i) - \hat{H}_2^d(Y_i) - \sum_{c=1}^p\big[H_2^d(Z_{ci}) - \hat{H}_2^d(Z_{ci})\big]\beta_{1c}\Big)\\
&\quad + n^{-1}\sum_{i=1}^n\big[H^*(Z_i) - \hat{H}^*(Z_i)\big]\big(\hat{\phi}_i - \phi_i\big)\Big(H_2^d(Y_i) - \hat{H}_2^d(Y_i) - \sum_{c=1}^p\big[H_2^d(Z_{ci}) - \hat{H}_2^d(Z_{ci})\big]\beta_{1c}\Big)
\end{aligned}$$

$$\equiv B_{31d} + B_{32d} + B_{33d} + B_{34d}$$

Consider,

$$\begin{aligned}
B_{32d} &= n^{-1}\big[\mathbf{Z}_n - \mathbf{H}_n^*(Z)\big]\big(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_n\big)\big[\mathsf{S}_{2n}^d(Y) - \mathsf{S}_{2n}^d(Z)\beta_1\big] \\
&= n^{-1}\sum_{i=1}^n \big[Z_i - H^*(Z_i)\big]\big(\hat{\phi}_i - \phi_i\big)\Big(H_2^d(Y_i) - \hat{H}_2^d(Y_i) - \sum_{c=1}^p \big[H_2^d(Z_{ci}) - \hat{H}_2^d(Z_{ci})\big]\beta_{1c}\Big) \\
&\le \sup_{XV\in G_{XV}} |\hat{\phi}(X,\hat{V}) - \phi(X,V)|\Big[\sup_{V_d\in G_{V_d}} \big|H_2^d(Y) - \hat{H}_2^d(Y)\big| \\
&\qquad + \max_{1\le j\le p}\sup_{V_d\in G_{V_d}} \big|H_2^d(Z_j) - \hat{H}_2^d(Z_j)\big|\sum_{c=1}^p |\beta_{1c}|\Big] n^{-1}\sum_{i=1}^n \big|Z_i - H^*(Z_i)\big| \\
&\le O_p\big(\mathcal{L}_{0n}\big)O_p\big(\mathcal{L}_{2n}\big)\big(1 + ||\beta_1||_E\big)O_p(1) = O_p\big(\mathcal{L}_{0n}\mathcal{L}_{2n}\big) = o_p(n^{-1/2}),
\end{aligned}$$

by Theorems 1 and 2, Lemma 3 xxvi) and xxvii), Assumption A3, and Markov's Inequality.

$$\begin{aligned}
B_{33d}(c) &= n^{-1}\big[\mathbf{H}_n^*(Z_c) - \hat{\mathbf{H}}_n^*(Z_c)\big]'\boldsymbol{\phi}_n\big[\mathsf{S}_{2n}^d(Y) - \mathsf{S}_{2n}^d(Z)\beta_1\big] \\
&= n^{-1}\sum_{i=1}^n \big[H^*(Z_{ci}) - \hat{H}^*(Z_{ci})\big]\phi_i\Big(H_2^d(Y_i) - \hat{H}_2^d(Y_i) - \sum_{c=1}^p \big[H_2^d(Z_{ci}) - \hat{H}_2^d(Z_{ci})\big]\beta_{1c}\Big) \\
&\le \sup_{XV\in G_{XV}} |H^*(Z_c) - \hat{H}^*(Z_c)|\Big[\sup_{V_d\in G_{V_d}} \big|H_2^d(Y) - \hat{H}_2^d(Y)\big| \\
&\qquad + \max_{1\le j\le p}\sup_{V_d\in G_{V_d}} \big|H_2^d(Z_j) - \hat{H}_2^d(Z_j)\big|\sum_{c=1}^p |\beta_{1c}|\Big] n^{-1}\sum_{i=1}^n \phi_i \\
&\le O_p\big(\mathcal{L}_n\mathcal{L}_{2n}\big)\big(1 + ||\beta_1||_E\big)O(1) = O_p\big(\mathcal{L}_n\mathcal{L}_{2n}\big) = o_p(n^{-1/2}),
\end{aligned}$$

by Theorem 2, Lemma 3 xxvi) and xxvii), and Assumption A3.

$$\begin{aligned}
B_{34d}(c) &= n^{-1}\big[\mathbf{H}_n^*(Z_c) - \hat{\mathbf{H}}_n^*(Z_c)\big]'\big(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_n\big)\big[\mathsf{S}_{2n}^d(Y) - \mathsf{S}_{2n}^d(Z)\beta_1\big] \\
&= n^{-1}\sum_{i=1}^n \big[H^*(Z_{ci}) - \hat{H}^*(Z_{ci})\big]\big(\hat{\phi}_i - \phi_i\big)\Big(H_2^d(Y_i) - \hat{H}_1^d(Y_i) - \sum_{c=1}^p \big[H_2^d(Z_{ci}) - \hat{H}_2^d(Z_{ci})\big]\beta_{1c}\Big) \\
&\le \sup_{XV\in G_{XV}} |\hat{\phi}(X,\hat{V}) - \phi(X,V)| \sup_{XV\in G_{XV}} |H^*(Z_c) - \hat{H}^*(Z_c)|\Big[\sup_{V_d\in G_{V_d}} \big|H_2^d(Y) - \hat{H}_2^d(Y)\big| \\
&\qquad + \max_{1\le j\le p}\sup_{V_d\in G_{V_d}} \big|H_2^d(Z_j) - \hat{H}_2^d(Z_j)\big|\sum_{c=1}^p |\beta_{1c}|\Big] \\
&= O_p\big(\mathcal{L}_n\big)O_p\big(\mathcal{L}_{0n}\big)O_p\big(\mathcal{L}_{2n}\big)\big(1 + ||\beta_1||_E\big) = O_p\big(\mathcal{L}_n\big)O_p\big(\mathcal{L}_{0n}\big)O_p\big(\mathcal{L}_{2n}\big) = o_p(n^{-1/2}),
\end{aligned}$$

by Theorems 1 and 2, Lemma 3 xxvi) and xxvii), and Assumption A3.

$$\begin{aligned}
B_{31d} &= n^{-1}\big[\mathbf{Z}_n - \mathbf{H}_n^*(Z)\big]\boldsymbol{\phi}_n\big[\mathsf{S}_{2n}^d(Y) - \mathsf{S}_{2n}^d(Z)\beta_1\big] \\
&= n^{-1}\sum_{i=1}^n \big[Z_i - H^*(Z_i)\big]\phi_i\Big(H_2^d(Y_i) - \hat{H}_2^d(Y_i) - \sum_{c=1}^p \big[H_2^d(Z_{ci}) - \hat{H}_2^d(Z_{ci})\big]\beta_{1c}\Big) \\
&= (-1)n^{-1}\sum_{i=1}^n \big[Z_i - H^*(Z_i)\big]\phi_i\big(\hat{H}_2^d(Y_i) - H_2^d(Y_i)\big) \\
&\qquad + \sum_{c=1}^p \beta_{1c}\, n^{-1}\sum_{i=1}^n \big[Z_i - H^*(Z_i)\big]\phi_i\big[\hat{H}_2^d(Z_{ci}) - H_2^d(Z_{ci})\big] \equiv -B_{311d} + \sum_{c=1}^p \beta_{1c}B_{312dc}.
\end{aligned}$$

Consider,

$$
\begin{aligned}
B_{312dc}(a) &= n^{-1}\sum_{i=1}^{n}\big[Z_{ai}-H^{*}(Z_{ai})\big]\phi_i\big[\hat{H}_2^d(Z_{ci})-H_2^d(Z_{ci})\big]\\
&= n^{-1}\sum_{i=1}^{n}\phi_i\zeta_{ai}\Big\{[(n-1)b_2]^{-1}\sum_{j\neq i}K_{2ji}(\hat{V}_d)\hat{\theta}_{2j}^d Z_{ci}-H_2^d(Z_{ci})\Big\}\\
&= n^{-1}\sum_{i=1}^{n}\phi_i\zeta_{ai}[(n-1)b_2]^{-1}\sum_{j\neq i}K_{2ji}(\hat{V}_d)\Big[\big(\hat{\theta}_{2j}^d-\theta_{2j}^d\big)Z_{cj}+\theta_{2j}^d\big(Z_{cj}-E[Z_{cj}|X_j,V_j]\big)\\
&\qquad\qquad +\theta_{2j}^d\big(E[Z_{cj}|X_j,V_j]-H_2^d(Z_{cj})\big)+\theta_{2j}^d\big(H_2^d(Z_{cj})-H_2^d(Z_{ci})\big)\Big]\\
&= n^{-1}\sum_{i=1}^{n}\phi_i\zeta_{ai}[(n-1)b_2]^{-1}\sum_{j\neq i}K_{2ji}[b_2^{-1}(\hat{V}_{dj}-\hat{V}_{di})]\mathsf{C}_{2ji}^d(c)\\
&= [n(n-1)]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\phi_i\zeta_{ai}b_2^{-1}K_{2ji}(V_d)\mathsf{C}_{2ji}^d(c)\\
&\quad -[n(n-1)b_2]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\phi_i\zeta_{ai}b_2^{-1}K_{2ji}^{(1)}(V_d)[\hat{m}_d^{l_n}(W_j)-m_d(W_j)]\mathsf{C}_{2ji}^d(c)\\
&\quad +[n(n-1)b_2]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\phi_i\zeta_{ai}b_2^{-1}K_{2ji}^{(1)}(V_d)[\hat{m}_d^{l_n}(W_i)-m_d(W_i)]\mathsf{C}_{2ji}^d(c)\\
&\quad +[2n(n-1)b_2^2]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\phi_i\zeta_{ai}b_2^{-1}K_{2ji}^{(2)}(V_d)[\hat{m}_d^{l_n}(W_i)-m_d(W_i)-\big(\hat{m}_d^{l_n}(W_j)-m_d(W_j)\big)\big]^2\mathsf{C}_{2ji}^d(c)\\
&\quad +[6n(n-1)b_2^3]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\phi_i\zeta_{ai}b_2^{-1}K_{2ji}^{(3)}(V_d)[\hat{m}_d^{l_n}(W_i)-m_d(W_i)-\big(\hat{m}_d^{l_n}(W_j)-m_d(W_j)\big)\big]^3\mathsf{C}_{2ji}^d(c)\\
&\quad +[24n(n-1)b_2^5]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\phi_i\zeta_{ai}K_{2ji}^{(4)}(\tilde{V}_d)[\hat{m}_d^{l_n}(W_i)-m_d(W_i)-\big(\hat{m}_d^{l_n}(W_j)-m_d(W_j)\big)\big]^4\mathsf{C}_{2ji}^d(c)\\
&\equiv B_{3121dc}(a)+B_{3122dd}(a)+B_{3123dc}(a)+B_{3124dc}(a)+B_{3125dc}(c)+B_{3126dc}(a).
\end{aligned}
$$

Note that proof of the order of B_{3121dc} is, mutatis mutandis, virtually identical to the proof of the order of B_{212dc} thuse the arguments are not repeated here and one can conclude that, $B_{3121dc}=o_p(n^{-1/2})$. Note that by Lemma 2 , Assumptions A4, and A5,

$$
\begin{aligned}
E\big[\phi_i\zeta_{ai}K_{2ji}^{(1)}(V_d)\big] &= E\big[K_{2ji}^{(1)}(V_d)E\big(\phi_i\zeta_{ai}|V_{di},S_j\big)\big]=0,\\
E\big[H_2^d(Z_{ci})\phi_i\zeta_{ai}K_{2ji}^{(1)}(V_d)\big] &= E\big[H_2^d(Z_{ci})K_{2ji}^{(1)}(V_d)E\big(\phi_i\zeta_{ai}|V_{di},S_j\big)\big]=0,
\end{aligned}
$$

and,

$$
\begin{aligned}
&\sup_{V_{di}\in G_{V_d}}E\big[|\phi_i\zeta_{ai}\big|^2V_{di}\big]=O(1),\\
&\sup_{V_{di}\in G_{V_d}}E\big[|H_2^d(Z_{ci})\phi_i\zeta_{ai}\big|^2V_{di}\big]\le\sup_{V_{di}\in G_{V_d}}H_2^d(Z_{ci})^2\sup_{V_{di}\in G_{V_d}}E\big[|\phi_i^2\zeta_{ai}^2\big|V_{di}\big]=O(1).
\end{aligned}
$$

Consequently, by a trivial modification of Lemma 6 ,

$$
\sup_{V_{dj}\in G_{V_d}}\Big|[nb_2^2]^{-1}\sum_{i=1}^{n}\Big[\phi_i\zeta_{ai}K_{2ji}^{(1)}(V_d)-E\Big(\phi_i\zeta_{ai}K_{2ji}^{(1)}(V_d)\Big)\Big]\Big|=O_p\left(\left[\frac{\log(n)}{nb_2^3}\right]^{1/2}\right),
$$

$$\sup_{V_{dj}\in G_{V_d}}\left|[nb_2^2]^{-1}\sum_{i=1}^n\Big[H_2^d(Z_{ci})\phi_i\zeta_{ai}K_{2ji}^{(1)}(V_d)-E\Big(H_2^d(Z_{ci})\phi_i\zeta_{ai}K_{2ji}^{(1)}(V_d)\Big)\Big]\right|=O_p\left(\left[\frac{\log(n)}{nb_2^3}\right]^{1/2}\right).$$

Consider,

$$\begin{aligned}
B_{3122dc}(a)&=[n(n-1)b_2]^{-1}\sum_{i=1}^n\sum_{j\neq i}\phi_i\zeta_{ai}b_2^{-1}K_{2ji}^{(1)}(V_d)[\hat m_d^{l_n}(W_j)-m_d(W_j)]\mathsf{C}_{2ji}^d(c)\\
&=(n-1)^{-1}\sum_{j\neq i}[\hat m_d^{l_n}(W_j)-m_d(W_j)]\mathsf{C}_{2j}^{d*}[nb_2^2]^{-1}\sum_{i=1}^n\Big[\phi_i\zeta_{ai}K_{2ji}^{(1)}(V_d)-E\Big(\phi_i\zeta_{ai}K_{2ji}^{(1)}(V_d)\Big)\Big]\\
&\quad-(n-1)^{-1}\sum_{j\neq i}[\hat m_d^{l_n}(W_j)-m_d(W_j)][nb_2^2]^{-1}\sum_{i=1}^n\Big[H_2^d(Z_{ci})\phi_i\zeta_{ai}K_{2ji}^{(1)}(V_d)-E\Big(H_2^d(Z_{ci})\phi_i\zeta_{ai}K_{2ji}^{(1)}(V_d)\Big)\Big]\\
&\leq\sup_{W\in G_W}[\hat m_d^{l_n}(W)-m_d(W)]\sup_{V_{dj}\in G_{V_d}}\left|[nb_2^2]^{-1}\sum_{i=1}^n\Big[\phi_i\zeta_{ai}K_{2ji}^{(1)}(V_d)-E\Big(\phi_i\zeta_{ai}K_{2ji}^{(1)}(V_d)\Big)\Big]\right|(n-1)^{-1}\sum_{j\neq i}|\mathsf{C}_{2j}^{d*}|\\
&\quad+\sup_{W\in G_W}[\hat m_d^{l_n}(W)-m_d(W)]\sup_{V_{dj}\in G_{V_d}}\left|[nb_2^2]^{-1}\sum_{i=1}^n\Big[H_2^d(Z_{ci})\phi_i\zeta_{ai}K_{2ji}^{(1)}(V_d)-E\Big(H_2^d(Z_{ci})\phi_i\zeta_{ai}K_{2ji}^{(1)}(V_d)\Big)\Big]\right|\\
&=O_p(L_n)O_p\left(\left[\frac{\log(n)}{nb_2^3}\right]^{1/2}\right)\sum_{j\neq i}|\mathsf{C}_{2j}^{d*}|+O_p(L_n)O_p\left(\left[\frac{\log(n)}{nb_2^3}\right]^{1/2}\right)\\
&\leq O_p(L_nb_2^{-1})O_p\left(\left[\frac{\log(n)}{nb_2}\right]^{1/2}\right)\Big\{1+\sup_{XV\in G_{XV}}|\hat\theta_{2j}^d(X,\hat V)-\theta_{2j}^d(X,V)|n^{-1}\sum_{j=1}^n|Z_{cj}|\\
&\qquad+n^{-1}\sum_{j=1}^n|\theta_{2j}^d\rho_{cj}|+n^{-1}\sum_{j=1}^n|\theta_{2j}^d\eta_{2cj}^d|+n^{-1}\sum_{j=1}^n|\theta_{2j}^dH_2^d(Z_{cj})|\Big\}\\
&=o_p(n^{-1/2})O_p(1)=o_p(n^{-1/2}),
\end{aligned}$$

by Theorem 1, Lemma 6 , Lemma 3 vi) and xxiv), and Markov's Inequality.

$$B_{3123dc}(a)=[n(n-1)b_2]^{-1}\sum_{i=1}^n\sum_{j\neq i}\phi_i\zeta_{ai}b_2^{-1}K_{2ji}^{(1)}(V_d)[\hat m_d^{l_n}(W_i)-m_d(W_i)]\mathsf{C}_{2ji}^d(c).$$

Recall that by assumption A2,

$$\lim_{\gamma\to\infty}K_2^{(1)}(\gamma)\gamma=0,\quad\text{and}\quad\lim_{\gamma\to-\infty}K_2^{(1)}(\gamma)\gamma=0.$$

Thus,

$$\begin{aligned}
&E\big[b_2^{-1}K_{2ji}^{(1)}(V_d)\theta_{2j}^d\phi_i\zeta_{ai}H_2^{(1)d}(Z_{ci})[b_2^{-1}(V_{dj}-V_{di})]\big|S_{-j}\big]\\
&\qquad=\phi_i\zeta_{ai}H_2^{(1)d}(Z_{ci})\int b_2^{-1}K_{2ji}^{(1)}(V_d)[b_2^{-1}(V_{dj}-V_{di})]g(X_j,V_{-dj})p(X_j,V_j)^{-1}p(X_j,V_j)dX_jdV_j\\
&\qquad=\phi_i\zeta_{ai}H_2^{(1)d}(Z_{ci})\int b_2^{-1}K_{2ji}^{(1)}(V_d)[b_2^{-1}(V_{dj}-V_{di})]dV_{dj}\int g(X_j,V_{-dj})dV_{-dj}dX_j\\
&\qquad=\phi_i\zeta_{ai}H_2^{(1)d}(Z_{ci})\int b_2^{-1}K_2^{(1)}(\gamma)b_2d\gamma\\
&\qquad=\phi_i\zeta_{ai}H_2^{(1)d}(Z_{ci})\Big[\lim_{\gamma\to\infty}K_2^{(1)}(\gamma)\gamma-\lim_{\gamma\to-\infty}K_2^{(1)}(\gamma)\gamma-\int K_2(\gamma)d\gamma\Big]
\end{aligned}$$

$$= -\phi_i \zeta_{ai} H_2^{(1)d}(Z_{ci}).$$

Eliminating the $E[Z_{cj}|X_j, V_j]$ terms from $\mathsf{C}_{2ji}^d(c)$ one has,

$$\begin{aligned}
B_{3123dc}(a) &= [n(n-1)b_2]^{-1} \sum_{i=1}^n \sum_{j\neq i} \phi_i \zeta_{ai} b_2^{-1} K_{2ji}^{(1)}(V_d)[\hat{m}_d^{l_n}(W_i) - m_d(W_i)] \mathsf{C}_{2ji}^d(c) \\
&= [n(n-1)b_2]^{-1} \sum_{i=1}^n \sum_{j\neq i} \phi_i \zeta_{ai} b_2^{-1} K_{2ji}^{(1)}(V_d)[\hat{m}_d^{l_n}(W_i) - m_d(W_i)] \big(\hat{\theta}_{2j}^d - \theta_{2j}^d\big) Z_{cj} \\
&\quad + [n(n-1)b_2]^{-1} \sum_{i=1}^n \sum_{j\neq i} \phi_i \zeta_{ai} b_2^{-1} K_{2ji}^{(1)}(V_d)[\hat{m}_d^{l_n}(W_i) - m_d(W_i)] \theta_{2j}^d \big(Z_{cj} - H_2^d(Z_{cj})\big) \\
&\quad + [n(n-1)b_2]^{-1} \sum_{i=1}^n \sum_{j\neq i} \phi_i \zeta_{ai} b_2^{-1} K_{2ji}^{(1)}(V_d)[\hat{m}_d^{l_n}(W_i) - m_d(W_i)] \theta_{2j}^d \big(H_2^d(Z_{cj}) - H_2^d(Z_{ci})\big) \\
&\equiv B_{31231dc}(a) + B_{31232dc}(a) + B_{31233dc}(a).
\end{aligned}$$

Consider,

$$\begin{aligned}
B_{31231dc}(a) &= [n(n-1)b_2]^{-1} \sum_{i=1}^n \sum_{j\neq i}^n \phi_i \zeta_{ai} b_2^{-1} K_{2ji}^{(1)}(V_d)[\hat{m}_d^{l_n}(W_i) - m_d(W_i)] \big(\hat{\theta}_{2j}^d - \theta_{2j}^d\big) Z_{cj} \\
&\le \sup_{XV\in G_{X,V}} \big|\hat{\theta}_2^d(X,\hat{V}) - \theta_2^d(X,V)\big| \sup_{W\in G_W} |\hat{m}_d^{l_n}(W) - m_d(W)| \\
&\qquad \times \sup_{V_{di}\in G_{V_d}} [(n-1)b_2]^{-1} \sum_{j\neq i} |b_2^{-1} K_{2ji}^{(1)}(V_d)| n^{-1} \sum_{i=1}^n |\phi_i \zeta_{ai}| \\
&= O_p\big(\mathcal{L}_{0n}\big) O_p(L_n) O_p(b_2^{-1}) O_p(1) = O_p\big(\mathcal{L}_{0n} L_n b_2^{-1}\big) = o_p(n^{-1/2}).
\end{aligned}$$

by Theorem 1 and Lemma 3 . Note that,

$$\begin{aligned}
E\big[K_{2ji}^{(1)}(V_d)\theta_{2j}^d\big(Z_{cj} - H_2^d(Z_{cj})\big)\big] &= E\big[K_{2ji}^{(1)}(V_d)p(V_{dj})^{-1}\phi_j\big(Z_{cj} - H_2^d(Z_{cj})\big)\big] \\
&= E\big[K_{2ji}^{(1)}(V_d)p(V_{dj})^{-1}E\big(\phi_j Z_{cj} - \phi_j H_2^d(Z_{cj})\big|V_{dj}, V_{di}\big)\big] \\
&= E\big[K_{2ji}^{(1)}(V_d)p(V_{dj})^{-1}\big(H_d^2(Z_{cj}) - H_2^d(Z_{cj})E[\phi_j|V_{dj}, V_{di}]\big)\big] \\
&= E\big[K_{2ji}^{(1)}(V_d)p(V_{dj})^{-1}H_d^2(Z_{cj})\big(1 - \int g(X_j, V_j)p(X_j, V_j)^{-1}p(X_j, V_j)p(V_{dj})^{-1}dX_j, dV_{-dj}\big)\big] \\
&= E\big[K_{2ji}^{(1)}(V_d)p(V_{dj})^{-1}H_d^2(Z_{cj})\big(1 - \int g(X_j, V_{-dj})dX_j, dV_{-dj}\big)\big] = 0
\end{aligned}$$

By Assumption A5, $E\big(|Z_{cj} - H_2^d(Z_{cj})|^2\big|V_{dj}\big) = O(1)$. Then by Lemma 6 ,

$$\begin{aligned}
&\sup_{V_{di}\in G_{V_d}} \Big|[(n-1)b_2^2]^{-1} \sum_{j\neq i} \Big[K_{2ji}^{(1)}(V_d)\theta_{2j}^d\big[Z_{cj} - H_2^d(Z_{cj})\big] - E\Big(K_{2ji}^{(1)}(V_d)\theta_{2j}^d\big[Z_{cj} - H_2^d(Z_{cj})\big]\Big)\Big]\Big| \\
&\quad = O_p\left(\left[\frac{\log(n)}{(n-1)b_2^3}\right]^{1/2}\right).
\end{aligned}$$

Consider,

$$B_{31232dc}(a) = [n(n-1)b_2]^{-1} \sum_{i=1}^n \sum_{j\neq i} \phi_i \zeta_{ai} b_2^{-1} K_{2ji}^{(1)}(V_d)[\hat{m}_d^{l_n}(W_i) - m_d(W_i)] \theta_{2j}^d \big(Z_{cj} - H_2^d(Z_{cj})\big)$$

$$
\begin{aligned}
&= n^{-1}\sum_{i=1}^{n}\phi_i\zeta_{ai}[\hat{m}_d^{l_n}(W_i)-m_d(W_i)][(n-1)b_2^2]^{-1}\sum_{j\neq i}K_{2ji}^{(1)}(V_d)\theta_{2j}^d\big[Z_{cj}-H_2^d(Z_{cj})\big]\\
&\leq \sup_{V_{di}\in G_{V_d}}\Big|[(n-1)b_2^2]^{-1}\sum_{j\neq i}\Big[K_{2ji}^{(1)}(V_d)\theta_{2j}^d\big[Z_{cj}-H_2^d(Z_{cj})\big]-E\Big(K_{2ji}^{(1)}(V_d)\theta_{2j}^d\big[Z_{cj}-H_2^d(Z_{cj})\big]\Big)\Big]\Big|\\
&\qquad\times\sup_{W\in G_W}\big|\hat{m}_d^{l_n}(W)-m_d(W)\big|n^{-1}\sum_{i=1}^{n}|\phi_i\zeta_{ai}|\\
&= O_p\left(\left[\frac{\log(n)}{(n-1)b_2^3}\right]^{1/2}\right)O_p(L_n)O_p(1)=O_p(L_nb_2^{-1})O_p\left(\left[\frac{\log(n)}{(n-1)b_2}\right]^{1/2}\right)=o_p(n^{-1/2}),
\end{aligned}
$$

by Lemma 3 x) and xxvii) and Lemma 6 , and Markov's Inequality. Now consider

$$
\begin{aligned}
B_{31233dc}(a) &= n^{-1}\sum_{i=1}^{n}\phi_i\zeta_{ai}[\hat{m}_d^{l_n}(W_i)-m_d(W_i)][(n-1)b_2^2]^{-1}\sum_{j\neq i}K_{2ji}^{(1)}(V_d)\theta_{2j}^d\big(H_2^d(Z_{cj})-H_2^d(Z_{ci})\big)\\
&\leq n^{-1}\sum_{i=1}^{n}\phi_i\zeta_{ai}[\hat{m}_d^{l_n}(W_i)-m_d(W_i)]H_2^{(1)d}(Z_{ci})[(n-1)b_2]^{-1}\sum_{j\neq i}K_{2ji}^{(1)}(V_d)\theta_{2j}^d[b_2^{-1}(V_{dj}-V_{di})]\\
&\qquad+\sup_{W\in G_W}|\hat{m}_d^{l_n}(W)-m_d(W)|]\big(O_p(b_2^{\nu_2-1})+o_p(n^{-1/2})\big)n^{-1}\sum_{i=1}^{n}|\phi_i\zeta_{ci}|\\
&= [n(n-1)b_2]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\phi_i\zeta_{ai}K_{2ji}^{(1)}(V_d)\theta_{2j}^dH_2^{(1)d}(Z_{ci})[b_2^{-1}(V_{dj}-V_{di})][\hat{m}_d^{l_n}(W_i)-m_d(W_i)]\\
&\qquad+O_p(L_nb_2^{-1})O_p(b_2^{\nu_2})+O_p(L_n)o_p(n^{-1/2})\\
&\equiv B^*_{31233dc}(a)+o_p(n^{-1/2}).
\end{aligned}
$$

By Lemma 11 xv.), Lemma 3 vi) and xxiii) and Markov's Inequality. So that furthermore,

$$
\begin{aligned}
{B^*}_{31233dc}(a) &= [n(n-1)b_2]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\phi_i\zeta_{ai}K_{2ji}^{(1)}(V_d)\theta_{2j}^dH_2^{(1)d}(Z_{ci})[b_2^{-1}(V_{dj}-V_{di})][m_d(W_i)-\hat{m}_d^{l_n}(W_i)]\\
&= [n(n-1)b_2]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\phi_i\zeta_{ai}K_{2ji}^{(1)}(V_d)\theta_{2j}^dH_2^{(1)d}(Z_{ai})[b_2^{-1}(V_{dj}-V_{di})][\hat{m}_d^{l_n}(W_i)-m_d^{l_n}(W_i)]\\
&\qquad+[n(n-1)b_2]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\phi_i\zeta_{ai}K_{2ji}^{(1)}(V_d)\theta_{2j}^dH_2^{(1)d}(Z_{ci})[b_2^{-1}(V_{dj}-V_{di})][m_d^{l_n}(W_i)-m_d(W_i)]\\
&= [n(n-1)b_2]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\phi_i\zeta_{ai}K_{2ji}^{(1)}(V_d)\theta_{2j}^dH_2^{(1)d}(Z_{ci})[b_2^{-1}(V_{dj}-V_{di})][m_d^{l_n}(W_i)-m_d(W_i)]\\
&\qquad-[n(n-1)b_2]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\phi_i\zeta_{ai}K_{2ji}^{(1)}(V_d)\theta_{2j}^dH_2^{(1)d}(Z_{ci})[b_2^{-1}(V_{dj}-V_{di})]\mathbf{B}_n(W_i)'Q_{BB}^{-1}n^{-1}\mathbf{B}_n'\mathbf{V}_d\\
&\qquad+[n(n-1)b_2]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\phi_i\zeta_{ai}K_{2ji}^{(1)}(V_d)\theta_{2j}^dH_2^{(1)d}(Z_{ci})[b_2^{-1}(V_{dj}-V_{di})]\mathbf{B}_n(W_i)'\\
&\qquad\qquad\times Q_{BB}^{-1}n^{-1}\mathbf{B}_n'\big(\mathbf{M}_d^{l_n}-\mathbf{M}_d\big)\\
&\qquad-[n(n-1)b_2]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\phi_i\zeta_{ai}K_{2ji}^{(1)}(V_d)\theta_{2j}^dH_2^{(1)d}(Z_{ci})[b_2^{-1}(V_{dj}-V_{di})]\mathbf{B}_n(W_i)'
\end{aligned}
$$

$$
\begin{aligned}
&\qquad\qquad\qquad\qquad\qquad\qquad \times \big(Q_{nBB}^{-1} - Q_{BB}^{-1}\big) n^{-1} \mathbf{B}_n' \mathbf{V}_d \\
&\qquad + [n(n-1)b_2]^{-1} \sum_{i=1}^{n} \sum_{j \neq i} \phi_i \zeta_{ai} K_{2ji}^{(1)}(V_d) \theta_{2j}^d H_2^{(1)d}(Z_{ci}) [b_2^{-1}(V_{dj} - V_{di})] \mathbf{B}_n(W_i)' \\
&\qquad\qquad\qquad\qquad\qquad\qquad \times \big(Q_{nBB}^{-1} - Q_{BB}^{-1}\big) n^{-1} \mathbf{B}_n' \big(\mathbf{M}_d^{l_n} - \mathbf{M}_d\big) \\
&\equiv o_p(n^{-1/2}) - o_p(n^{-1/2}) + R_{1dc}(a) - R_{2dc}(a) + R_{3dc}(a).
\end{aligned}
$$

By Lemma 11 xiii.) and xiv.). Note that,

$$
\begin{aligned}
&E\Big(||(n-1)^{-1}\theta_{2j}^d b_2^{-1} \mathbf{K}_{2j}^{(1)d}(V_d)' \boldsymbol{\phi}_i \dot{\zeta}_c \dot{\mathbf{H}}_2^{(1)d}(Z_{ci}) [b_2^{-1}(\dot{\mathbf{V}}_{dj} - \dot{\mathbf{V}}_d)] \mathbf{B}_n||_E^2\Big) \\
&\quad = E\Big(||(n-1)^{-1} \sum_{i \neq j} \mathbf{B}_n(W_i) \phi_i \zeta_{ci} H_2^{(1)d}(Z_{ci}) b_2^{-1} K_{2ji}^{(1)}(V_d) \theta_{2j}^d [b_2^{-1}(V_{dj} - V_{di})]||_E^2\Big) \\
&\quad = E\Big([\theta_{2j}^d]^2 b_2^{-1} (n-1)^{-2} \sum_{i \neq j} \mathbf{B}_n(W_i)' \mathbf{B}_n(W_i) [\phi_i \zeta_{ci}]^2 H_2^{(1)d}(Z_{ci}) b_2^{-1} K_{2ji}^{(1)}(V_d)^2 [b_2^{-1}(V_{dj} - V_{di})]^2\Big) \\
&\qquad + E\Big([\theta_{2j}^d]^2 (n-1)^{-2} \sum_{i \neq j} \sum_{\substack{g \neq j \\ g \neq i}} \mathbf{B}_n(W_i)' \phi_i \zeta_{ci} H_2^{(1)d}(Z_{ci}) b_2^{-1} K_{2ji}^{(1)}(V_d)^2 [b_2^{-1}(V_{dj} - V_{di})] \\
&\qquad\qquad\qquad\qquad\qquad \times \mathbf{B}_n(W_g) \phi_g \zeta_{cg} H_2^{(1)d}(Z_{cg}) b_2^{-1} K_{2jg}^{(1)}(V_d)^2 [b_2^{-1}(V_{dj} - V_{dg})]\Big) \\
&\quad \leq \sup_{X,V \in G_{XV}} \theta_2^d(X,V)^2 \sup_{V_d, W \in G_{V_d,W}} E[\phi^2 \zeta_c^2 | V_d, W] \sup_{V_d \in G_{V_d}} H_2^{(1)d}(Z_c)^2 \\
&\qquad \times \Bigg\{ b_2^{-1}(n-1)^{-2} \sum_{i \neq j} E\Big(\mathbf{B}_n(W_i)' \mathbf{B}_n(W_i) E\Big[b_2^{-1} K_{2ji}^{(1)}(V_d)^2 [b_2^{-1}(V_{dj} - V_{di})]^2 \Big| W_i\Big]\Big) \\
&\qquad\qquad + (n-1)^{-2} \sum_{i \neq j} \sum_{\substack{g \neq j \\ g \neq i}} E\Big(\big|\mathbf{B}_n(W_i)'\big| E\Big[b_2^{-1} K_{2ji}^{(1)}(V_d)^2 [b_2^{-1}(V_{dj} - V_{di})] \Big| S_{-i}, W_i\Big] \Bigg\} \\
&\qquad\qquad\qquad\qquad\qquad \times \big|\mathbf{B}_n(W_g)\big| E\Big[b_2^{-1} K_{2jg}^{(1)}(V_d)^2 [b_2^{-1}(V_{dj} - V_{dg})] \Big| S_{-g}, W_g\Big]\Big) \\
&\quad = O\big([(n-1)b_2]^{-1}\big) E\Big(\mathbf{B}_n(W_i)' \mathbf{B}_n(W_i)\Big) + E\Big(\big|\mathbf{B}_n(W_i)'\big|\big|\mathbf{B}_n(W_g)\big|\Big) \\
&\quad = O\big(l_n [(n-1)b_2]^{-1}\big) + O(1) = O(1).
\end{aligned}
$$

By Lemma 11 xiii.) and xiv.) Consequently by Assumption 1, and Markov's inequality,

$$
||(n-1)^{-1}\theta_{2j}^d b_2^{-1} \mathbf{K}_{2j}^{(1)d}(V_d)' \boldsymbol{\phi}_i \dot{\zeta}_c \dot{\mathbf{H}}_2^{(1)d}(Z_{ci}) [b_2^{-1}(\dot{\mathbf{V}}_{dj} - \dot{\mathbf{V}}_d)] \mathbf{B}_n||_E = O_p(1).
$$

Consider,

$$
\begin{aligned}
&R_{1dc}(a) - R_{2dc}(a) + R_{3dc}(a) \\
&\quad = [n(n-1)b_2]^{-1} \sum_{i=1}^{n} \sum_{j \neq i} \phi_i \zeta_{ai} K_{2ji}^{(1)}(V_d) \theta_{2j}^d H_2^{(1)d}(Z_{ci}) [b_2^{-1}(V_{dj} - V_{di})] \mathbf{B}_n(W_i)' \\
&\qquad\qquad\qquad\qquad\qquad\qquad \times Q_{BB}^{-1} n^{-1} \mathbf{B}_n' \big(\mathbf{M}_d^{l_n} - \mathbf{M}_d\big) \\
&\qquad - [n(n-1)b_2]^{-1} \sum_{i=1}^{n} \sum_{j \neq i} \phi_i \zeta_{ai} K_{2ji}^{(1)}(V_d) \theta_{2j}^d H_2^{(1)d}(Z_{ci}) [b_2^{-1}(V_{dj} - V_{di})] \mathbf{B}_n(W_i)' \\
&\qquad\qquad\qquad\qquad\qquad\qquad \times \big(Q_{nBB}^{-1} - Q_{BB}^{-1}\big) n^{-1} \mathbf{B}_n' \mathbf{V}_d
\end{aligned}
$$

$$
\begin{aligned}
&+ [n(n-1)b_2]^{-1} \sum_{i=1}^{n} \sum_{j \neq i} \phi_i \zeta_{ai} K_{2ji}^{(1)}(V_d) \theta_{2j}^d H_2^{(1)d}(Z_{ci}) [b_2^{-1}(V_{dj} - V_{di})] \mathbf{B}_n(W_i)' \\
&\qquad\qquad\qquad\qquad \times \big(Q_{nBB}^{-1} - Q_{BB}^{-1}\big) n^{-1} \mathbf{B}_n' \big(\mathbf{M}_d^{l_n} - \mathbf{M}_d\big) \\
&\leq n^{-1} \sum_{j=1}^{n} \Big\{ ||(n-1)^{-1} \theta_{2j}^d b_2^{-1} \mathbf{K}_{2j}^{(1)d}(V_d)' \boldsymbol{\phi}_i \dot{\zeta}_c \dot{\mathbf{H}}_2^{(1)d}(Z_{ci}) [b_2^{-1}(\dot{\mathbf{V}}_{dj} - \dot{\mathbf{V}}_d)] \\
&\qquad\qquad\qquad\qquad \times \mathbf{B}_n ||_E ||Q_{BB}^{-1}||_{sp} ||n^{-1} \mathbf{B}_n' \big(\mathbf{M}_d^{l_n} - \mathbf{M}_d\big)||_E \\
&+ ||(n-1)^{-1} \theta_{2j}^d b_2^{-1} \mathbf{K}_{2j}^{(1)d}(V_d)' \boldsymbol{\phi}_i \dot{\zeta}_c \dot{\mathbf{H}}_2^{(1)d}(Z_{ci}) [b_2^{-1}(\dot{\mathbf{V}}_{dj} - \dot{\mathbf{V}}_d)] \mathbf{B}_n ||_E \\
&\qquad\qquad\qquad\qquad \times ||Q_{nBB}^{-1} - Q_{BB}^{-1}||_{sp} ||n^{-1} \mathbf{B}_n' \mathbf{V}_d||_E \\
&+ ||(n-1)^{-1} \theta_{2j}^d b_2^{-1} \mathbf{K}_{2j}^{(1)d}(V_d)' \boldsymbol{\phi}_i \dot{\zeta}_c \dot{\mathbf{H}}_2^{(1)d}(Z_{ci}) [b_2^{-1}(\dot{\mathbf{V}}_{dj} - \dot{\mathbf{V}}_d)] \mathbf{B}_n ||_E \\
&\qquad\qquad\qquad\qquad \times ||Q_{nBB}^{-1} - Q_{BB}^{-1}||_{sp} ||n^{-1} \mathbf{B}_n' \big(\mathbf{M}_d^{l_n} - \mathbf{M}_d\big)||_E \Big\} \\
&= n^{-1} \sum_{j=1}^{n} \Big\{ O(1) O(1) O_p(l_n^{-k}) + O(1) O_p\left(\frac{l_n}{\sqrt{n}}\right) O_p\left(\left[\frac{l_n}{n}\right]^{1/2}\right) + O(1) O_p\left(\frac{l_n}{\sqrt{n}}\right) O_p(l_n^{-k}) \Big\} \\
&= o_p(n^{-1/2}).
\end{aligned}
$$

By Lemma 3 i.), ii.), vi.). Now consider,

$$
\begin{aligned}
B_{31233dc}(a) &= B_{31233dc}^*(a) + o_p(n^{-1/2}) \\
&= o_p(n^{-1/2}) + R_{1dc}(a) - R_{2dc}(a) + R_{3dc}(a) + o_p(n^{-1/2}) \\
&= o_p(n^{-1/2}),
\end{aligned}
$$

As a result,

$$
B_{3123dc}(a) = B_{31231dc}(a) + B_{31232dc}(a) + B_{31233dc}(a) = o_p(n^{-1/2}).
$$

Furthermore,

$$
\begin{aligned}
&B_{3124dc}(a) + B_{3125dc}(c) + B_{3126dc}(a) \\
&\leq \big[2b_2^{-1} \sup_{W \in G_W} \big|\hat{m}_d^{l_n}(W) - m_d(W)\big|\big]^2 [n(n-1)]^{-1} \sum_{i=1}^{n} \sum_{j \neq i} \big|\phi_i \zeta_{ai} b_2^{-1} K_{2ji}^{(2)}(V_d) \mathsf{C}_{2ji}^d(c)\big| \\
&\quad + \big[2b_2^{-1} \sup_{W \in G_W} \big|\hat{m}_d^{l_n}(W) - m_d(W)\big|\big]^3 [n(n-1)]^{-1} \sum_{i=1}^{n} \sum_{j \neq i} \big|\phi_i \zeta_{ai} b_2^{-1} K_{2ji}^{(3)}(V_d) \mathsf{C}_{2ji}^d(c)\big| \\
&\quad + \big[2b_2^{-1} \sup_{W \in G_W} \big|\hat{m}_d^{l_n}(W) - m_d(W)\big|\big]^3 b_2^{-1} [n(n-1)]^{-1} \sum_{i=1}^{n} \sum_{j \neq i} \big|\phi_i \zeta_{ai} K_{2ji}^{(4)}(\tilde{V}_d) \mathsf{C}_{2ji}^d(c)\big| \\
&= O_p\Big([L_n b_2^{-1}]^2\Big) [n(n-1)]^{-1} \sum_{i=1}^{n} \sum_{j \neq i} |\phi_i u_i| \big|b_2^{-1} K_{1ji}^{(2)}(V_d)\big| \big|\mathsf{C}_{2ji}^d(c)\big| \\
&\quad + O_p\Big([L_n b_2^{-1}]^3\Big) [n(n-1)]^{-1} \sum_{i=1}^{n} \sum_{j \neq i} |\phi_i u_i| \big|b_2^{-1} K_{1ji}^{(3)}(V_d)\big| \big|\mathsf{C}_{2ji}^d(c)\big| \\
&\quad + O_p\Big([L_n^4 b_2^{-5}]\Big) [n(n-1)]^{-1} \sum_{i=1}^{n} \sum_{j \neq i} |\phi_i u_i| \big|\mathsf{C}_{2ji}^d(c)\big|.
\end{aligned}
$$

Note that the proof of the order of $B_{3124dc}(a) + B_{3125dc}(c) + B_{3126dc}(a)$ is, mutatis mutandis, (exchanging u_i for ζ_{ai}) practically identical to the proof of the order of $B_{1424d}(c) + B_{1425d}(c) + B_{1426d}(c)$. Thus the arguments will not be repeated here. As a result, one can conclude,

$$B_{3124dc}(a)+B_{3125dc}(c)+B_{3126dc}(a)=o_p(n^{-1/2}).$$

Furthermore,

$$B_{312dc}=B_{3121dc}+B_{3122dc}+B_{3123dc}+B_{3124dc}+B_{3125dc}(c)+B_{3126dc}=o_p(n^{-1/2}).$$

Note that the proof of the order of B_{311d} is, mutatis mutandis, (exchanging Z_{ci} for Y_i) practically identical to the proof of the order of B B_{312dc}. Thus will not be repeated here. As a result, one can conclude,$B_{311d}=o_p(n^{-1/2})$. Consequently,

$$B_{31d}=B_{311d}-\sum_{c=1}^{p}\beta_{1c}B_{312dc}\leq o_p(n^{-1/2})\big[1+\sum_{c=1}^{p}|\beta_{1c}|\big]=o_p(n^{-1/2})\big[1+||\beta_1||_E\big]=o_p(n^{-1/2}).$$

Also, $B_3=\sum_{d=1}^{d}\big[B_{31d}+B_{32d}+B_{33d}+B_{34d}\big]=o_p(n^{-1/2})$. Now note that,

$$\begin{aligned}
B_4&=(2D-1)n^{-1}\big[\mathbf{Z}_n-\hat{\mathbf{H}}^*_n(Z)\big]'\hat{\boldsymbol{\phi}}_n\Big(\big[\boldsymbol{\mu}_Z-\hat{\boldsymbol{\mu}}_Z\big]\beta_1-\big[\boldsymbol{\mu}_Y-\hat{\boldsymbol{\mu}}_Y\big]\Big)\\
&=(2D-1)n^{-1}\big[\mathbf{Z}_n-\mathbf{H}^*_n(Z)\big]'\boldsymbol{\phi}_n\Big(\big[\boldsymbol{\mu}_Z-\hat{\boldsymbol{\mu}}_Z\big]\beta_1-\big[\boldsymbol{\mu}_Y-\hat{\boldsymbol{\mu}}_Y\big]\Big)\\
&\qquad+(2D-1)n^{-1}\big[\mathbf{Z}_n-\mathbf{H}^*_n(Z)\big]'\big(\hat{\boldsymbol{\phi}}_n-\boldsymbol{\phi}_n\big)\Big(\big[\boldsymbol{\mu}_Z-\hat{\boldsymbol{\mu}}_Z\big]\beta_1-\big[\boldsymbol{\mu}_Y-\hat{\boldsymbol{\mu}}_Y\big]\Big)\\
&\qquad+(2D-1)n^{-1}\big[\mathbf{H}^*_n(Z)-\hat{\mathbf{H}}^*_n(Z)\big]'\boldsymbol{\phi}_n\Big(\big[\boldsymbol{\mu}_Z-\hat{\boldsymbol{\mu}}_Z\big]\beta_1-\big[\boldsymbol{\mu}_Y-\hat{\boldsymbol{\mu}}_Y\big]\Big)\\
&\qquad+(2D-1)n^{-1}\big[\mathbf{H}^*_n(Z)-\hat{\mathbf{H}}^*_n(Z)\big]'\big(\hat{\boldsymbol{\phi}}_n-\boldsymbol{\phi}_n\big)\Big(\big[\boldsymbol{\mu}_Z-\hat{\boldsymbol{\mu}}_Z\big]\beta_1-\big[\boldsymbol{\mu}_Y-\hat{\boldsymbol{\mu}}_Y\big]\Big)\\
&\equiv(2D+1)\big[B_{41}+B_{42}+B_{43}+B_{44}\big].
\end{aligned}$$

Note that by Lemma 2 ,

$$E\Big(\Big|n^{-1}\sum_{i=1}^{n}\phi_i\zeta_{ci}\Big|^2\Big)=n^{-2}\sum_{i=1}^{n}E\big(\phi^2\zeta_{ci}^2\big)+[n(n-1)]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}E\big(\phi_i\zeta_{ci}\big)E\big(\phi_j\zeta_{cj}\big)=O(n^{-1}).$$

Consequently by Markov's Inequality, $|n^{-1}\sum_{i=1}^{n}\phi_i\zeta_{ci}\big|=O_p(n^{-1/2})$.

$$\begin{aligned}
B_{41}&=n^{-1}\big[\mathbf{Z}_n-\mathbf{H}^*_n(Z)\big]'\boldsymbol{\phi}_n\Big(\big[\boldsymbol{\mu}_Z-\hat{\boldsymbol{\mu}}_Z\big]\beta_1-\big[\boldsymbol{\mu}_Y-\hat{\boldsymbol{\mu}}_Y\big]\Big)\\
&=n^{-1}\sum_{i=1}^{n}\big[Z_i-H^*(Z_i)\big]\phi_i\Big(\sum_{j=1}^{p}\big[\mu_{Z_j}-\hat{\mu}_{Z_j}\big]\beta_{1j}-\big[\mu_Y-\hat{\mu}_Y\big]\Big)\\
&\leq\Big|\sum_{j=1}^{p}\big[\mu_{Z_j}-\hat{\mu}_{Z_j}\big]\beta_{1j}-\big[\mu_Y-\hat{\mu}_Y\big]\Big|\Big|n^{-1}\sum_{i=1}^{n}\phi_i\zeta_i\Big|\\
&=O_p(\mathcal{L}_{0n})O_p(n^{-1/2})\big(1+||\beta_1||_E\big)=o_p(n^{-1/2}).
\end{aligned}$$

By Theorem 2, Lemma 3 xxvi.) and Markov's Inequality,

$$\begin{aligned}
B_{42}&=n^{-1}\big[\mathbf{Z}_n-\mathbf{H}^*_n(Z)\big]'\big(\hat{\boldsymbol{\phi}}_n-\boldsymbol{\phi}_n\big)\Big(\big[\boldsymbol{\mu}_Z-\hat{\boldsymbol{\mu}}_Z\big]\beta_1-\big[\boldsymbol{\mu}_Y-\hat{\boldsymbol{\mu}}_Y\big]\Big)\\
&=n^{-1}\sum_{i=1}^{n}\big[Z_i-H^*(Z_i)\big]\big(\hat{\phi}_i-\phi_i\big)\Big(\sum_{j=1}^{p}\big[\mu_{Z_j}-\hat{\mu}_{Z_j}\big]\beta_{1j}-\big[\mu_Y-\hat{\mu}_Y\big]\Big)
\end{aligned}$$

$$\leq \Big|\sum_{j=1}^{p}\big[\mu_{Z_j}-\hat{\mu}_{Z_j}\big]\beta_{1j}-[\mu_Y-\hat{\mu}_Y]\Big| \sup_{X,V\in G_{XV}}\big|\hat{\phi}(X,\hat{V})-\phi(X,V)\big| n^{-1}\sum_{i=1}^{n}|\zeta_i|$$
$$= O_p(\mathcal{L}_{0n}^2)O_p(1)\big(1+||\beta_1||_E\big)=o_p(n^{-1/2}).$$

By Theorem 1 and 2, Lemma 3 xxvi.) and Markov's Inequality,

$$\begin{aligned}
B_{43} &= n^{-1}\big[\mathbf{H}_n^*(Z)-\hat{\mathbf{H}}_n^*(Z)\big]'\boldsymbol{\phi}_n\Big(\big[\boldsymbol{\mu}_Z-\hat{\boldsymbol{\mu}}_Z\big]\beta_1-\big[\boldsymbol{\mu}_Y-\hat{\boldsymbol{\mu}}_Y\big]\Big)\\
&= n^{-1}\sum_{i=1}^{n}\big[H^*(Z_i)-\hat{H}^*(Z_i)\big]\phi_i\Big(\sum_{j=1}^{p}\big[\mu_{Z_j}-\hat{\mu}_{Z_j}\big]\beta_{1j}-[\mu_Y-\hat{\mu}_Y]\Big)\\
&\leq \Big|\sum_{j=1}^{p}\big[\mu_{Z_j}-\hat{\mu}_{Z_j}\big]\beta_{1j}-[\mu_Y-\hat{\mu}_Y]\Big| \sup_{X,V\in G_{X,V}}\big|H^*(Z_i)-\hat{H}^*(Z_i)\big| n^{-1}\sum_{i=1}^{n}|\phi_i|\\
&O_p(\mathcal{L}_{0n})O_p(\mathcal{L}_n)\big(1+||\beta_1||_E\big)=o_p(n^{-1/2}),
\end{aligned}$$

By Theorem 2, Lemma 3 xxvi.) xxvii.) and Markov's Inequality,

$$\begin{aligned}
B_{44} &= n^{-1}\big[\mathbf{H}_n^*(Z)-\hat{\mathbf{H}}_n^*(Z)\big]'\big(\hat{\boldsymbol{\phi}}_n-\boldsymbol{\phi}_n\big)\Big(\big[\boldsymbol{\mu}_Z-\hat{\boldsymbol{\mu}}_Z\big]\beta_1-\big[\boldsymbol{\mu}_Y-\hat{\boldsymbol{\mu}}_Y\big]\Big)\\
&= n^{-1}\sum_{i=1}^{n}\big[H^*(Z_i)-\hat{H}^*(Z_i)\big]\big(\hat{\phi}_i-\phi_i\big)\Big(\sum_{j=1}^{p}\big[\mu_{Z_j}-\hat{\mu}_{Z_j}\big]\beta_{1j}-[\mu_Y-\hat{\mu}_Y]\Big)\\
&\leq \Big|\sum_{j=1}^{p}\big[\mu_{Z_j}-\hat{\mu}_{Z_j}\big]\beta_{1j}-[\mu_Y-\hat{\mu}_Y]\Big| \sup_{X,V\in G_{XV}}\big|\hat{\phi}(X,\hat{V})-\phi(X,V)\big| \sup_{X,V\in G_{X,V}}\big|H^*(Z_i)-\hat{H}^*(Z_i)\big|\\
&= O_p(\mathcal{L}_{0n}^2)O_p(\mathcal{L}_n)\big(1+||\beta_1||_E\big)=o_p(n^{-1/2}),
\end{aligned}$$

by Theorems 1, 2 and Lemma 3 xxvi.) and xxvii.) Consequently,

$$B_4=(2D-1)\big[B_{41}+B_{42}+B_{43}+B_{44}\big]=o_p(n^{-1/2}).$$

In summary,

$$\sqrt{n}B=\sqrt{n}B_1+\sqrt{n}B_2+\sqrt{n}B_3+\sqrt{n}B_4=\sqrt{n}B_1+o_p(1)$$

Hence, by the Cramer Wold Device,

$$\sqrt{n}(\hat{\beta}_1-\beta_1)=A^{-1}\sqrt{n}B\xrightarrow{d}N(0,\Sigma_0^{-1}\Sigma_1\Sigma_0^{-1}).$$

Now given rate of convergence for $\hat{\beta}_1$, I show a rate of convergence for $\hat{\beta}_0$ Note that by assumption A.5(i) $\beta_0 = E(Y_i)-E(Z_i')\beta_1-E(h(X_i))-E(f(V_i))-E(u_i)$, consequently, by Assumption A3 and Markov's inequality;

$$\begin{aligned}
\hat{\beta}_0-\beta_0 &= n^{-1}\sum_{i=1}^{n}Y_i-\hat{\beta}_1'n^{-1}\sum_{i=1}^{n}Z_i-E(Y_i)+E(Z_i')\beta_1\\
&= n^{-1}\sum_{i=1}^{n}(Y_i-E(Y_i))-(\hat{\beta}_1'-\beta_1')n^{-1}\sum_{i=1}^{n}[Z_i-E(Z_i)]-(\hat{\beta}_1'-\beta_1')E(Z_i)-\beta_1'n^{-1}\sum_{i=1}^{n}(Z_i-E(Z_i))\\
&\leq \left|n^{-1}\sum_{i=1}^{n}(Y_i-E(Y_i))\right|+\left(\max_{1\leq j\leq p}\left|n^{-1}\sum_{i=1}^{n}Z_{ji}-E(Z_{ji})\right|+\max_{1\leq j\leq p}E(|Z_i|)\right)\sum_{j=1}^{p}|\hat{\beta}_1-\beta_1|
\end{aligned}$$

$$+ ||\beta_1||_E \max_{1\le j\le p}\left|n^{-1}\sum_{i=1}^{n} Z_{ji} - E(Z_{ji})\right|$$
$$= O_p(n^{-1/2}) + \big(O_p(n^{-1/2}) + O(1)\big)O_p(n^{-1/2}) + O(1)O_p(n^{-1/2}) = O_p(n^{-1/2})$$

□

Proof of Theorem 4

Proof of (*i*) : First I adopt the following notation equivalencies which will help compress notation in this proof. Let $\tilde{V}_{di} = \lambda\hat{V}_{di} + (1-\lambda)V_{di}$ and define.

$$\frac{d\mathbf{P}_n^{k_n}(V_d)'}{dV_d} = \left[\begin{array}{cccc} \frac{dB_{1,a}(V_d)}{dV_d} & \frac{dB_{2,a}(V_d)}{dV_d} & \cdots & \frac{dB_{k_n+2a,a}(V_d)}{dV_d}\end{array}\right]$$
$$\mathbf{P}_n^{(1)}(\tilde{V}_i)(\hat{V}_i - V_i)' = \left[\begin{array}{cccc} \frac{d\mathbf{P}_n^{k_n}(\tilde{V}_{1i})'}{d\tilde{V}_{1i}}(\hat{V}_{1i}-V_{1i}) & \frac{d\mathbf{P}_n^{k_n}(\tilde{V}_{2i})'}{d\tilde{V}_{2i}}(\hat{V}_{2i}-V_{2i}) & \cdots & \frac{d\mathbf{P}_n^{k_n}(\tilde{V}_{Di})'}{d\tilde{V}_{Di}}(\hat{V}_{Di}-V_{Di})\end{array}\right]$$
$$\ddot{\mathbf{P}}_n^{(1)}(\tilde{V}_i)(\hat{V}_i - V_i)' = \left[\begin{array}{cc} \vec{0}'_{D(k_n+2a)} & \mathbf{P}_n^{(1)}(\tilde{V}_i)(\hat{V}_i - V_i)'\end{array}\right]$$

where $\vec{0}'_{D(k_n+2a)}$ is a $D(k_n+2a)$ dimensional vector of zeros.

Note : (a) by Assumption A1 and Lemma B.4.iv of Ozabaci et al. (2013); $||Q_{PP}^{-1}||_{sp} = O(1)$, and $||\hat{Q}_{nPP}^{-1} - Q_{PP}^{-1}||_{sp} = O_p(\tau_n)$. (b) $O_p(\tau_n) = O_p(1)$ by Assumption A7. (c) $|\mathbf{P}_n(X_i,V_i)|$ is the element wise absolute value of $\mathbf{P}_n(X_i,V_i)$. (d) As mentioned on page 151 of Newey (1997) for B-Spline basis functions.

$$\sup_{V_d\in G_{V_d}} \left|\left|\frac{d\mathbf{P}_n^{k_n}(V_d)'}{dV_d}\right|\right|_E = O(k_n^{3/2})$$

Now consider that by the remarks to Assumption 6;

$$\sup_{X,V\in G_{XV}} |\mathbf{P}(X,V)'\Gamma_n - h(X) - f(V)| = O_p(k_n^{-a})$$

Now consider,

$$\begin{aligned}
\hat{\Gamma}_n - \Gamma_n &= \hat{Q}_{n,pp}^{-1}n^{-1}\hat{\mathbf{P}}_n'\big(\mathbf{Y}_n - \hat{\beta}_0\mathbf{1}_n - \mathbf{Z}_n\hat{\beta}_1\big) - \Gamma_n\\
&= \hat{Q}_{n,pp}^{-1}n^{-1}\hat{\mathbf{P}}_n'\big((\beta_0-\hat{\beta}_0)\mathbf{1}_n - \mathbf{Z}_n(\hat{\beta}_1-\beta_1) + \mathbf{h}_n + \mathbf{f}_n + \mathbf{u}_n - \hat{\mathbf{P}}_n\Gamma_n\big)\\
&= \hat{Q}_{n,pp}^{-1}n^{-1}\hat{\mathbf{P}}_n'\big((\beta_0-\hat{\beta}_0)\mathbf{1}_n + \mathbf{Z}_n(\beta_1-\hat{\beta}_1)\big) + \hat{Q}_{n,pp}^{-1}n^{-1}\hat{\mathbf{P}}_n'\big(\mathbf{h}_n + \mathbf{f}_n + \mathbf{u}_n - \hat{\mathbf{P}}_n\Gamma_n\big)\\
&\equiv A_1 + A_2
\end{aligned}$$

By Theorem 3.1 of Ozabaci et al. (2013) $||A_2||_E = O_p\left(\sqrt{\frac{l_n}{n}} + \sqrt{\frac{k_n}{n}} + l_n^{-k} + k_n^{-a}\right)$. Now consider,

$$\begin{aligned}
A_1 &= \hat{Q}_{n,pp}^{-1}n^{-1}\hat{\mathbf{P}}_n'\big((\beta_0-\hat{\beta}_0)\mathbf{1}_n + \mathbf{Z}_n(\beta_1-\hat{\beta}_1)\big)\\
&= Q_{pp}^{-1}n^{-1}\mathbf{P}_n'\big((\beta_0-\hat{\beta}_0)\mathbf{1}_n + \mathbf{Z}_n(\beta_1-\hat{\beta}_1)\big)\\
&\quad + Q_{pp}^{-1}n^{-1}(\hat{\mathbf{P}}_n - \mathbf{P}_n')\big((\beta_0-\hat{\beta}_0)\mathbf{1}_n + \mathbf{Z}_n(\beta_1-\hat{\beta}_1)\big)\\
&\quad + (\hat{Q}_{n,pp}^{-1} - Q_{pp}^{-1})n^{-1}\mathbf{P}_n'\big((\beta_0-\hat{\beta}_0)\mathbf{1}_n + \mathbf{Z}_n(\beta_1-\hat{\beta}_1)\big)\\
&\quad + (\hat{Q}_{n,pp}^{-1} - Q_{pp}^{-1})n^{-1}(\hat{\mathbf{P}}_n - \mathbf{P}_n')\big((\beta_0-\hat{\beta}_0)\mathbf{1}_n + \mathbf{Z}_n(\beta_1-\hat{\beta}_1)\big)
\end{aligned}$$

As a result,

$$||A_1||_E \le |\beta_0 - \hat{\beta}_0|\Big(\big|\big|Q_{pp}^{-1}\big|\big|_{sp} + \big|\big|Q_{pp}^{-1} - \hat{Q}_{n,pp}^{-1}\big|\big|_{sp}\Big)$$

$$
\begin{aligned}
&\times \Big(\big|\big|n^{-1}\sum_{i=1}^{n}\mathbf{P}_n(X_i,V_i)\big|\big|_E + \big|\big|n^{-1}\sum_{i=1}^{n}(\hat{\mathbf{P}}_n(X_i,V_i)-\mathbf{P}_n(X_i,V_i))\big|\big|_E\Big) \\
&+\sum_{j=1}^{p}|\beta_{1j}-\hat{\beta}_{1j}|\Big(\big|\big|Q_{pp}^{-1}\big|\big|_{sp}+\big|\big|Q_{pp}^{-1}-\hat{Q}_{n,pp}^{-1}\big|\big|_{sp}\Big) \\
&\times \Big(\big|\big|n^{-1}\sum_{i=1}^{n}\mathbf{P}_n(X_i,V_i)Z_{ji}\big|\big|_E + \big|\big|n^{-1}\sum_{i=1}^{n}(\hat{\mathbf{P}}_n(X_i,V_i)-\mathbf{P}_n(X_i,V_i))Z_{ji}\big|\big|_E\Big) \\
&\le O_p(n^{-1/2})O(1+\tau_n)\Big(\big|\big|n^{-1}\sum_{i=1}^{n}\mathbf{P}_n(X_i,V_i)\big|\big|_E + \big|\big|n^{-1}\sum_{i=1}^{n}\ddot{\mathbf{P}}_n^{(1)}(\tilde{V}_i)(\hat{V}_i-V_i)\big|\big|_E\Big) \\
&+O_p(n^{-1/2})O(1+\tau_n)\Big(\big|\big|n^{-1}\sum_{i=1}^{n}\mathbf{P}_n(X_i,V_i)Z_{ji}\big|\big|_E + \big|\big|n^{-1}\sum_{i=1}^{n}\ddot{\mathbf{P}}_n^{(1)}(\tilde{V}_i)(\hat{V}_i-V_i)Z_{ji}\big|\big|_E\Big) \\
&= O_p(n^{-1/2})\Big(\big|\big|n^{-1}\sum_{i=1}^{n}\mathbf{P}_n(X_i,V_i)\big|\big|_E + n^{-1}\sum_{i=1}^{n}\big|\big|\ddot{\mathbf{P}}_n^{(1)}(\tilde{V}_i)(\hat{V}_i-V_i)\big|\big|_E\Big) \\
&+O_p(n^{-1/2})\Big(\big|\big|n^{-1}\sum_{i=1}^{n}\mathbf{P}_n(X_i,V_i)Z_{ji}\big|\big|_E + n^{-1}\sum_{i=1}^{n}\big|\big|\ddot{\mathbf{P}}_n^{(1)}(\tilde{V}_i)(\hat{V}_i-V_i)Z_{ji}\big|\big|_E\Big)
\end{aligned}
$$

Now we examine each remaining piece of of A_1,

$$
\begin{aligned}
E\Big(||n^{-1}\sum_{i=1}^{n}\mathbf{P}_n(X_i,V_i)||_E^2\Big) &= n^{-2}\sum_{i=1}^{n}\sum_{g=1}^{n}E\Big(\mathbf{P}_n(X_i,V_i)'\mathbf{P}_n(X_g,V_g)\Big) \\
&= n^{-2}\sum_{i=1}^{n}E\Big(\mathbf{P}_n(X_i,V_i)'\mathbf{P}_n(X_i,V_i)\Big) + n^{-2}\sum_{i=1}^{n}\sum_{g\neq i}E\Big(\mathbf{P}_n(X_i,V_i)'\Big)E\Big(\mathbf{P}_n(X_g,V_g)\Big) \\
&= O(k_n/n)+O(1)=O(1)
\end{aligned}
$$

Also,

$$
\begin{aligned}
E\Big(||n^{-1}\sum_{i=1}^{n}\mathbf{P}_n(X_i,V_i)Z_{ji}||_E^2\Big) &= n^{-2}\sum_{i=1}^{n}\sum_{g=1}^{n}E\Big(\mathbf{P}_n(X_i,V_i)'\mathbf{P}_n(X_g,V_g)Z_{ji}Z_{gi}\Big) \\
&\le \sup_{X,V\in G_{XV}}E[Z_j^2|X,V]n^{-2}\sum_{i=1}^{n}E\Big(\mathbf{P}_n(X_i,V_i)'\mathbf{P}_n(X_i,V_i)\Big) \\
&\quad + \sup_{X,V\in G_{XV}}E[|Z_j||X,V]^2 n^{-2}\sum_{i=1}^{n}\sum_{g\neq i}E\Big(|\mathbf{P}_n(X_i,V_i)'|\Big)E\Big(|\mathbf{P}_n(X_g,V_g)|\Big) \\
&= O(1)\big[O(k_n/n)+O(1)\big]=O(1)
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
E\Big(\big|\big|\ddot{\mathbf{P}}_n^{(1)}(\tilde{V}_i)(\hat{V}_i-V_i)\big|\big|_E^2\Big) &\le \sum_{j=1}^{D}E\Big(\Big|\Big|\frac{d\mathbf{P}_n^{k_n}(\tilde{V}_{ji})'}{d\tilde{V}_{1i}}(\hat{V}_{ji}-V_{ji})\Big|\Big|_E^2\Big) \\
&\le \sum_{j=1}^{D}\sup_{V_j\in G_{V_j}}\Big|\Big|\frac{d\mathbf{P}_n^{k_n}(V_j)'}{dV_j}\Big|\Big|_E E\big(|\hat{V}_{ji}-V_{ji}|^2\big) \\
&= O(k_n^3)O_p\left(\frac{l_n}{n}+l_n^{-2k}\right)
\end{aligned}
$$

By Markov's inequality. Now by Assumption A.5/A.7, $k_n = l_n$, thus we have,

$$||\ddot{\mathbf{P}}_n^{(1)}(\tilde{V}_i)(\hat{V}_i - V_i)||_E^2 = O_p\Big(\Big[\frac{k_n^3 l_n}{n}\Big]^{1/2} + k_n^{3/2} l_n^{-k}\Big) = O_p\Big(\Big[\frac{l_n^4}{n}\Big]^{1/2} + l_n^{3/2-k} = o_p(1)$$

Furthermore by Assumption A.2

$$\begin{aligned}
E\Big(||\ddot{\mathbf{P}}_n^{(1)}(\tilde{V}_i)(\hat{V}_i - V_i)Z_{ji}||_E^2\Big) &\leq \sum_{j=1}^{D} E\Big(\Big|\Big|\frac{d\mathbf{P}_n^{k_n}(\tilde{V}_{ji})'}{d\tilde{V}_{1i}}(\hat{V}_{ji} - V_{ji})Z_{ji}\Big|\Big|_E^2\Big) \\
&\leq \sum_{j=1}^{D} \sup_{Z_j \in G_{Z_j}} |Z_j|^2 \sup_{V_j \in G_{V_j}} \Big|\Big|\frac{d\mathbf{P}_n^{k_n}(V_j)'}{dV_j}\Big|\Big|_E E\big(|\hat{V}_{ji} - V_{ji}|^2\big) \\
&= O(k_n^3) O_p\left(\frac{l_n}{n} + l_n^{-2k}\right) = o_p(1)
\end{aligned}$$

Consequently by Markov's Inequality,

$$\begin{aligned}
||A_1||_E = O_p(n^{-1/2})&\Big(||n^{-1}\sum_{i=1}^{n}\mathbf{P}_n(X_i, V_i)||_E + n^{-1}\sum_{i=1}^{n}||\ddot{\mathbf{P}}_n^{(1)}(\tilde{V}_i)(\hat{V}_i - V_i)||_E\Big) \\
&+ O_p(n^{-1/2})\Big(||n^{-1}\sum_{i=1}^{n}\mathbf{P}_n(X_i, V_i)Z_{ji}||_E + n^{-1}\sum_{i=1}^{n}||\ddot{\mathbf{P}}_n^{(1)}(\tilde{V}_i)(\hat{V}_i - V_i)Z_{ji}||_E\Big) \\
= O_p(n^{-1/2})&
\end{aligned}$$

Consequently

$$||\hat{\Gamma}_n - \Gamma_n||_E = ||A_1||_E + ||A_2||_E = O_p\left(n^{-1/2} + \sqrt{\frac{l_n}{n}} + \sqrt{\frac{k_n}{n}} + l_n^{-k} + k_n^{-a}\right)$$

Proof of *(ii)*

$$\begin{aligned}
\sup_{X,V \in G_{XV}} |\mathbf{P}_n(X,V)'\hat{\Gamma}_n - h(X) - f(V)| &\leq \sup_{X,V \in G_{XV}} |\mathbf{P}_n(X,V)'\hat{\Gamma}_n - \mathbf{P}_n(X,V)'\Gamma_n| \\
&\quad + \sup_{X,V \in G_{XV}} |\mathbf{P}_n(X,V)'\Gamma_n - h(X) - f(V)| \\
&\leq ||\hat{\Gamma}_n - \Gamma_n||_E \sup_{X,V \in G_{XV}} ||\mathbf{P}_n(X,V)'||_E + O\big(k_n^{-a}\big) \\
&= O_p\left(\sqrt{\frac{l_n}{n}} + \sqrt{\frac{k_n}{n}} + l_n^{-k} + k_n^{-a}\right) O(k_n^{1/2}) + O(k^{-a}) \\
&= O_p\left(k_n^{1/2}\left[\sqrt{\frac{l_n}{n}} + \sqrt{\frac{k_n}{n}} + l_n^{-k} + k_n^{-a}\right]\right)
\end{aligned}$$

□

Proof of Theorem 5

$$\begin{aligned}
H[\ \tilde{h}(x_d)\ \ \tilde{h}^{(1)}(x_d)\ ]' &= H\big(\mathbf{X}_{dn}'\mathbf{K}_{4n}(x_d)\mathbf{X}_{dn}(x_d)\big)^{-1}H \cdot H^{-1}\mathbf{X}_{dn}'\mathbf{K}_{4n}(x_d)\tilde{Y}_{dn} \\
&= \big(n^{-1}H^{-1}\mathbf{X}_{dn}'\mathbf{K}_{4n}(x_d)\mathbf{X}_{dn}(x_d)H^{-1}\big)^{-1} n^{-1}H^{-1}\mathbf{X}_{dn}'\mathbf{K}_{4n}(x_d)\big(\tilde{Y}_{dn} - Y_{dn} + Y_{dn}\big) \\
&= \big(n^{-1}H^{-1}\mathbf{X}_{dn}'\mathbf{K}_{4n}(x_d)\mathbf{X}_{dn}(x_d)H^{-1} - \Sigma(x_d) + \Sigma(x_d)\big)^{-1} n^{-1}H^{-1}\mathbf{X}_{dn}'\mathbf{K}_{4n}(x_d)\big(\tilde{Y}_{dn} - Y_{dn}\big) \\
&\quad + \big(n^{-1}H^{-1}\mathbf{X}_{dn}'\mathbf{K}_{4n}(x_d)\mathbf{X}_{dn}(x_d)H^{-1} - \Sigma(x_d) + \Sigma(x_d)\big)^{-1} n^{-1}H^{-1}\mathbf{X}_{dn}'\mathbf{K}_{4n}(x_d)Y_{dn}
\end{aligned}$$

$$\equiv \big(A_1(x_d)+\Sigma(x_d)\big)^{-1}B_1(x_d)+\big(A_1(x_d)+\Sigma(x_d)\big)^{-1}B_2(x_d)$$

Note: In theorem 3.2 of Ozabaci et al. (2013) it is shown that,

$$A_1(x_b)=o_p(1) \text{ uniformly in } G_{X_d}$$

$$n^{1/2}h_4^{1/2}\Big[\big(A_1(x_d)+\Sigma(x_d)\big)^{-1}B_2(x_d)-[\ 1/2h_d^{(1)}(x_d)h_4^2\int u^2K_4(u)^2du\ \ 0\ ]'\Big]\xrightarrow{d} N(0,\Omega(x_d))$$

$$\sup_{x_d\in G_{X_d}}\big|\big|\big(A_1(x_d)+\Sigma(x_d)\big)^{-1}B_2(x_d)\big|\big|_E=O_p\left(\left[\frac{log(n)}{nh_4}\right]^{1/2}+h_4^2\right)$$

Hence all that remains is to show that $B_1(x_d)=o_p(n^{-1/2}h_4^{-1/2})$ so consider,

$$\begin{aligned}
B_1(x_d)&=n^{-1}H^{-1}\mathbf{X}'_{dn}\mathbf{K}_{4n}(x_d)\big(\tilde{Y}_{dn}-Y_{dn}\big)\\
&=n^{-1}H^{-1}\mathbf{X}'_{dn}\mathbf{K}_{4n}(x_d)\big((\hat\beta_0-\beta_0)\mathbf{1}_n+\mathbf{Z}_n(\hat\beta_1-\beta_1)\big)\\
&\qquad+n^{-1}H^{-1}\mathbf{X}'_{dn}\mathbf{K}_{4n}(x_d)\big(\hat{\mathbf{h}}_{-dn}-\mathbf{h}_{-dn}+\hat{\mathbf{f}}_n-\mathbf{f}_n\big)\\
&=(\hat\beta_0-\beta_0)n^{-1}H^{-1}\mathbf{X}'_{dn}\mathbf{K}_{4n}(x_d)\mathbf{1}_n+\sum_{j=1}^{p}(\hat\beta_{1j}-\beta_{1j})n^{-1}H^{-1}\mathbf{X}'_{dn}\mathbf{K}_{4n}(x_d)\mathbf{Z}_{jn}\\
&\qquad+n^{-1}H^{-1}\mathbf{X}'_{dn}\mathbf{K}_{4n}(x_d)\big(\hat{\mathbf{h}}_{-dn}-\mathbf{h}_{-dn}+\hat{\mathbf{f}}_n-\mathbf{f}_n\big)\\
&=(\hat\beta_0-\beta_0)[\ B_{11}(x_d)\ \ B_{12}(x_d)\ ]'+\sum_{j=1}^{p}(\hat\beta_{1j}-\beta_{1j})[\ B_{21j}(x_d)\ \ B_{22j}(x_d)\ ]'+B_3(x_d)
\end{aligned}$$

Now we find a bound for each component of the following,

$$\begin{aligned}
E(|B_{11}(x_d)|)&=n^{-1}\sum_{i=1}^{n}\int h_4^{-1}|K_4[h_4^{-1}(X_{di}-x_d)]h_4^{-1}(X_{di}-x_d)|p(X_{di})dX_{di}\\
&=\sup_{X_d\in G_{X_d}}|p(X_d)|\int|K_4(\gamma)\gamma|d\gamma=O(1)
\end{aligned}$$

Next consider,

$$\begin{aligned}
E(|B_{12}(x_d)|)&=n^{-1}\sum_{i=1}^{n}\int|h_4^{-1}K_4[h_4^{-1}(X_{di}-x_d)]|p(X_{di})dX_{di}\\
&=\sup_{X_d\in G_{X_d}}|p(X_d)|\int|K_4(\gamma)|d\gamma=O(1)
\end{aligned}$$

Futhermore

$$\begin{aligned}
E(|B_{21}(x_d)|)&=n^{-1}\sum_{i=1}^{n}\int h_4^{-1}|K_4[h_4^{-1}(X_{di}-x_d)]h_4^{-1}(X_{di}-x_d)Z_{ji}|p(X_{di})dX_{di}\\
&=\sup_{X,V\in G_{XV}}E[|Z_j||X,V]\sup_{X_d\in G_{X_d}}|p(X_d)|\int|K_4(\gamma)\gamma|d\gamma=O(1)
\end{aligned}$$

Also,

$$E(|B_{22}(x_d)|)=n^{-1}\sum_{i=1}^{n}\int|h_4^{-1}K_4[h_4^{-1}(X_{di}-x_d)]Z_{ji}|p(X_{di})dX_{di}$$

$$= \sup_{X,V\in G_{XV}} E[|Z_j||X,V] \sup_{X_d\in G_{X_d}} |p(X_d)| \int |K_4(\gamma)|d\gamma = O(1)$$

Lastly note that in theorem 3.2 of Ozabaci et al. (2013) it is shown that $B_3(x_d) = o_p(n^{-1/2}h_4^{-1/2})$, thus by Markov's inequality,

$$\begin{aligned}
\sqrt{nh_4}B_1(x_d) &= \sqrt{nh_4}(\hat{\beta}_0 - \beta_0)[\ B_{11}(x_d) \ \ B_{12}(x_d)\ ]' + \sum_{j=1}^{p}(\hat{\beta}_{1j} - \beta_{1j})[\ B_{21j}(x_d) \ \ B_{22j}(x_d)\ ]' + B_3(x_d) \\
&= \sqrt{nh_4}O_p(n^{-1/2})[\ O_p(1) \ \ O_p(1)\ ] + \sqrt{nh_4}O_p(n^{-1/2})[\ O_p(1) \ \ O_p(1)\ ] + \sqrt{nh_4}o_p(n^{-1/2}h_4^{-1/2}) \\
&= o_p(1)
\end{aligned}$$

Which gives the result. □