Estimation and Inference of Semiparametric Models using Data from Several Sources

Moshe Buchinsky * Fanghua Li[†] Zhipeng Liao[‡]

This Version: July 2016

Abstract

This paper studies the estimation and inference of nonlinear econometric model when the economic variables are contained in different data sets. We construct a minimum distance (MD) estimator of the unknown structural parameter of interest when there are some common conditioning variables in different data sets. The MD estimator is show to be consistent and has asymptotic normal distribution. We provide the specific form of optimal weight for the MD estimation, and show that the optimal weighted MD estimator has the smallest asymptotic variance among all MD estimators. Consistent estimator of the variance-covariance matrix of the MD estimator, and hence inference procedure of the unknown parameter is also provided. The finite sample performances of the MD estimator and the inference procedure are investigated in simulation studies.

JEL Classification: C12, C14

Keywords: Conditional Moment Restrictions; Data Combination; Minimum Distance Estimation

1 Introduction

There are many cases in empirical micro studies where data needed to analyze a particular phenomenon is not always available in one data set. Typically, this hampers the possibility of meaningful empirical research. In fact, a common phenomenon is to make some simplifying assumptions, which would then permit the researcher to use information from more than one data source. For example, as Blundell et at. (2008) note, this is a crucial difficulty when faced by those studying households' consumption and saving behavior because of the lack of panel data on both household expenditures, income, and saving.

An important data for studying consumption is, for example, the Panel Study of Income Dynamics (PSID), a survey that provides longitudinal annual data for household that have been followed since 1968. The PSID collects data on a subset of consumption items, namely food at home and food away from home (with few gaps in some of the survey years) and income. However, the PSID is that it does not provide data on wealth.

^{*}Department of Economics, UC Los Angeles, 8373 Bunche Hall, Mail Stop: 147703, Los Angeles, CA 90095. Email: buchinsky@econ.ucla.edu

[†]Department of Economics, UC Los Angeles Mail Stop: 147703, Los Angeles, CA 90095.

[‡]Department of Economics, UC Los Angeles, 8379 Bunche Hall, Mail Stop: 147703, Los Angeles, CA 90095. Email: zhipeng.liao@econ.ucla.edu

In contrast, there are few data set that provide detailed data on income and wealth (e.g. Health and Retirement Survey (HRS), or the National Longitudinal Study (NLS)), these data sets provide no information on consumption.

Problems of similar nature exist in many other countries. For example, in the UK, the Family Expenditure Survey (FES) provides comprehensive data on household expenditures, but this is across-sectional data and thus the researcher does not get to observe households over time. In contrast, the British Household Panel Survey (BHPS) is a Panel data set that collects data on income or wealth, but it provides no information on consumption.¹ This is quite puzzling, given the vital need to study the consumption decisions jointly with the income and wealth processes.

Consequently, as is clearly and comprehensively explain in Blundell et al. (2008), studies the aimed at understanding consumption behavior, and testing alternative theories, have resorted to the limited data on food expenditure provided in the PSID. This includes, among others, Hall and Mishkin (1982), Zeldes (1989), Runkle (1991), and Shea (1995), Cochrane (1991), Hayashi, Altonji and Kotlikoff (1996), Cox, Ng and Waldkirch (2004), Martin (2003) and Hurst and Stafford (2004) for tests of many alternative theories. The main problem with all these studies is that they use consumption on limited number of goods (largely necessity goods), and thus putting into question the external validity of the results.

One way that has been used in the literature is to form synthetic panel data sets from repeated crosssection data sets in which consumption is reported (e.g. the CEX or the FES). This is done in, for example, Browning, Deaton and Irish (1985) and Attanasio and Weber (1993).

An alternative empirical approach that have been used occasionally in the literature involve imputation of consumption to the PSID households using information on consumption from the CEX. Specifically, Skinner (1987) proposes to impute total consumption in the PSID using the estimated coefficients of a regression of total consumption on a number of consumption items that are reported in both the PSID and the CEX. While this method seems appealing at first sight, it reduces any variation in total consumption, since it does not take into account the fact the there is considerable idiosyncratic elements that goes into the individual decision making. Ziliak (1998) and Browning and Leth-Petersen (2003) provided alternative method that are variants of that proposed by Skinner (1987).

However, this method has a major weakness, in that it ignored, by construction, the dynamics of the individuals' consumption. Avoiding direct control of Individual's heterogeneity has been shown to provide major obstacle when modeling individual's behavior in general.

The one paper in the literature that provides a method that is related in spirit to the method proposed here is the paper by Blundell et al. (2008).² The method is similar in nature to that of Skinner (1987), in that the authors impute consumption data for the households in the PSID using regression parameters estimated from the CEX data. The key difference, is that the authors in Blundell et al. (2006) is that they use "structural" regression of a standard demand function for food that depends not only on other consumption items, but also depend on prices and a set of demographic and socio-economic variable of the household. Assuming monotonicity of the demand for makes it possible to invert these function in order to obtain a structurally based formula non-durable consumption, which exists in the CEX, but is missing in the PSID. Nevertheless, the general problem with such an imputation method described above for Skinner's method still apply. If nothing else, the consumption data imputed in this fashion is likely to suffer from the well-known error-in-variable problem. Most importantly, it ignores the inherent individual heterogeneity of

¹Similar problems exist for many other countries, in particular countries in Europe (e.g. France and Spain) that collect detailed data on both consumption, income, and wealth, but the information never exists in a single data set.

 $^{^{2}}$ This method is used extensively in Blundell, Pistaferri and Preston (2008).

consumption.

In two recent papers Fan, Sherman and Shum (2014, 2016) address a special case of the problem addressed in our paper, namely the case of treatment effect. Under this scenario, the outcome variables and conditioning variable are observed in two separate data sets, so that the treatment effect parameters are not pointidentidfied. The authors provide sharp bounds on the counterfactual distributions and parameter of interest (see Fan, Sherman and Shum (2016)), and the corresponding inference (see Fan, Sherman and Shum (2014).

Our case is more general and encompass a more general situation in which some of the variables are available only in one data set, while others are available in a separate data sets. The key insight is that there are some variables that appear in both. Under relatively mild regularity conditions we provide a method that allow one to point-identified the structural parameters of interest in the main data of interest using the information provided in the other data set. The parameters of interest and the "imputation equation" are estimated simultaneously. We also provide the necessary theory for inference including cases in which the number of observation in "imputation" data set do not diverge to infinity at the same rate as that for the main data of interest.

The semiparametric estimator proposed in this paper is related to the two sample instrumental variable (2SIV) estimator studied in Klevmarken (1982), Angrist and Krueger (1992) and Arellano and Meghir (1992) and the two sample GMM estimator studied in Ridder and Moffitt (2007). The main difference is that these papers consider estimation of the structural parameters with finite many moment restrictions, while our paper studies the estimation problem with conditional moment conditions and hence infinite many unconditional moment restrictions.

Notation of this paper is standard. For any square matrix A, $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the smallest and largest eigenvalues of A respectively. Throughout this paper, we use C to denote a generic finite positive constant which is larger than 1. For any set of real vectors $\{a_l\}_{l\in I}$ where $I = \{l_1, \ldots, l_{d_I}\}$ is a index set with d_I distinct natural numbers, we define $(a_l)_{l\in I} = (a_{l_1}, \ldots, a_{l_{d_I}})$ and $(a_l)'_{l\in I} = (a_{l_1}, \ldots, a_{l_{d_I}})'$. The notation $\|\cdot\|$ denotes the Euclidean norm. A' refers to the transpose of any matrix A. I_k and $\mathbf{0}_l$ are used to denote $k \times k$ identity matrix and $l \times l$ zero matrices respectively. The symbolism $A \equiv B$ means that A is defined as B. the expression $a_n = o_p(b_n)$ signifies that $\Pr(|a_n/b_n| \ge \epsilon) \to 0$ for all $\epsilon > 0$ as n go to infinity; and $a_n = O_p(b_n)$ when $\Pr(|a_n/b_n| \ge M) \to 0$ as n and M go to infinity. As usual, $"\to_p"$ and $"\to_d"$ imply convergence in probability and convergence in distribution, respectively.

2 The Model and the Estimators

We are interested in estimating the following model

$$Y = g(X_1, X_2, \theta_0) + v$$
 (1)

where X_1 and X_2 are sets of regressors, $g(\cdot, \cdot, \cdot) : R^{d_{x_1}} \times R^{d_{x_2}} \times R^{d_{\theta}} \to R$ is a known function, θ_0 is the unknown parameter of interest and v is a unobservable residual term. Two data sets are available to estimate the unknown parameter: $\{(Y_i, X'_{2,i}, X'_{3,i})\}_{i \in I_1}$ and $\{(X_{1,i}, X'_{2,i}, X'_{3,i})\}_{i \in I_2}$, where I_1 and I_2 are two index sets with cardinalities n_1 and n_2 respectively.

The unknown parameter θ_0 could be conveniently estimated under the conditional moment restriction $E[v|X_1, X_2] = 0$, if we had the joint observations on (Y, X'_1, X'_2) . However, such straightforward method is not applicable here because Y and X_1 are contained in different data sets. On the other hand, the common

variables X_2 and X_3 contained in both data sets can be useful for identifying and estimating the unknown parameter θ_0 . For this purpose, we assume that

$$E[v|X_2, X_3] = 0. (2)$$

Using the expression in (1) and the conditional mean restriction in (2), we get

$$E[Y|X_2, X_3] = E[g(X_1, X_2, \theta_0)|X_2, X_3],$$
(3)

which is the key equation for the identification and estimation of θ_0 .

For ease of notations, we write $X = (X'_1, X'_2)'$ and $Z = (X'_2, X'_3)'$. Then the model can be written as

$$Y = g(X, \theta_0) + v \text{ with } E[v|Z] = 0.$$

$$\tag{4}$$

For any θ , we define the conditional expectation of $g(X, \theta)$ given Z as

$$\phi(Z,\theta) = E\left[g(X,\theta)|Z\right].$$

Then we can write

$$Y = \phi(Z, \theta_0) + \varepsilon + v = h_0(Z) + u_1$$

where $\varepsilon = g(X, \theta_0) - \phi(Z, \theta_0), u \equiv \varepsilon + v$ and $h_0(Z) = \phi(Z, \theta_0)$.

As the conditioning variable Z is available in both data sets, we have n $(n = n_1 + n_2)$ observations: $\{Z_i\}_{i \in I} = \{(X'_{2,i}, X'_{3,i})\}_{i \in I}$ where $I = I_1 \cup I_2$. Let $P_k(z) = [p_1(z), \ldots, p_k(z)]'$ be a k-dimensional vector of basis functions for any positive integer k. For any k and any n, we define $P_{n,k} = (P_k(Z_i))_{i \in I}$ which is an $n \times k$ matrix. Accordingly, we define $P_{n_1,k_1} = (P_{k_1}(Z_i))_{i \in I_1}$ and $P_{n_2,k_2} = (P_{k_2}(Z_i))_{i \in I_2}$ which are $n_1 \times k_1$ and $n_2 \times k_2$ matrices respectively.

The conditional mean function $h_0(Z) = E[Y|Z]$ can be estimated using the first data set by

$$\widehat{h}_{n_1}(z) = P_{k_1}(z)' (P'_{n_1,k_1} P_{n_1,k_1})^{-1} P'_{n_1,k_1} Y_{n_1}$$
(5)

where $Y_{n_1} = (Y_i)'_{i \in I_1}$. Using the second data set, we get the following estimator of the conditional mean function $\phi(Z, \theta)$ for any θ :

$$\widehat{\phi}_{n_2}(Z,\theta) = P_{k_2}(Z)' (P'_{n_2,k_2} P_{n_2,k_2})^{-1} P'_{n_2,k_2} g_{n_2}(\theta)$$
(6)

where $g_{n_2}(\theta) = (g(X_i, \theta))'_{i \in I_2}$. Using the estimators of $h_0(Z)$ and $\phi(Z, \theta)$, we can construct the estimator of θ_0 via the minimum distance (MD) estimation:

$$\widehat{\theta}_{n} = \arg\min_{\theta\in\Theta} n^{-1} \sum_{i\in I} \left[\widehat{w}_{n}(Z_{i}) \left| \widehat{h}_{n_{1}}(Z_{i}) - \widehat{\phi}_{n_{2}}(Z_{i},\theta) \right|^{2} \right]$$
(7)

where Θ denotes the parameter space containing θ_0 , and $\hat{w}_n(\cdot)$ is a non-negative weight function. One simple and straightforward choice of the weight function $\hat{w}_n(\cdot)$ is the identity function, i.e., $\hat{w}_n(z) = 1$ for any z. However, as we will show later in this paper, the identity weighted MD estimator may not have the smallest possible variance.

3 Asymptotic Properties of the MD Estimator

In this section, we establish the asymptotic properties of the MD estimator. For any positive integer k, Let $\xi_k = \sup_{z \in \mathbb{Z}} ||P_k(z)||$ and $Q_k = E[P_k(Z)P'_k(Z)]$, where \mathbb{Z} denotes the support of Z. We first state the sufficient conditions for consistency.

Assumption 1. (i) $\{(Y_i, Z_i)\}_{i \in I_1}$ and $\{(X_i, Z_i)\}_{i \in I_2}$ are independent with i.i.d. observations; (ii) Var[Y|Z] < C; (iii) $C^{-1} \leq \lambda_{\min}(Q_k) \leq \lambda_{\max}(Q_k) \leq C$ for all k; (iv) there exist $\beta_{h,k} \in \mathbb{R}^k$ and $r_h > 0$ such that

$$\sup_{z \in \mathcal{Z}} |h_0(z) - P_k(z)' \beta_{h,k}| = \sup_{z \in \mathcal{Z}} |h_0(z) - h_{0,k}(z)| = O(k^{-r_h});$$
(8)

(v) $\max_{j=1,2} \xi_{k_j}^2 \log(k_j) n_j^{-1} = o(1) \text{ and } k_1 n_1^{-1} + k_1^{-1} = o(1).$

Assumption 1 includes mild and standard conditions on nonparametric series estimation of conditional mean function (see, e.g. Andrews (1991), Newey (1997) and Chen (2007)).

Define

$$L_{n}(\theta) = n^{-1} \sum_{i \in I} \left(w_{n}(Z_{i}) \left| h_{0}(Z_{i}) - \phi(Z_{i}, \theta) \right|^{2} \text{ and } L_{n}^{*}(\theta) = E \left[w_{n}(Z) \left| h_{0}(Z) - \phi(Z, \theta) \right|^{2} \right] \right)$$

for any $\theta \in \Theta$, where $w_n(\cdot)$ is defined in Assumption 2(v) below.

Assumption 2. (i) $\sup_{\theta \in \Theta} E\left[\phi^2(Z,\theta)\right] \not\in C$; (ii) $n^{-1} \sum_{i \in I} |\widehat{\phi}_{n_2}(Z_i,\theta) - \phi(Z_i,\theta)|^2 = o_p(1)$ uniformly over θ ; (iii) for any $\varepsilon > 0$, there is $\eta_{\varepsilon} > 0$ such that

$$E\left[\left|h_{0}(Z)-\phi(Z,\theta)\right|^{2}\right] \Leftrightarrow \eta_{\varepsilon} \text{ for any } \theta \in \Theta \text{ with } \left|\left|\theta-\theta_{0}\right|\right| \geq \varepsilon;$$

(iv) $\sup_{\theta \in \Theta} |L_n(\theta) - L_n^*(\theta)| = o_p(1);$ (v) $\sup_{z \in \mathcal{Z}} |\widehat{w}_n(z) - w_n(z)| = O_p(\delta_{w,n})$ where $\delta_{w,n} = O(n_1^{-1/4} + n_2^{-1/4})$ and $w_n(\cdot)$ is a sequence of non-random functions with $C^{-1} \leq w_n(z) \leq C$ for any n and any $z \in \mathcal{Z}$.

Assumption 2(i) imposes uniform finite second moment condition on the function $\phi(Z, \theta)$. Assumption 2(ii) requires that the nonparametric estimator $\hat{\phi}_{n_2}(Z_i, \theta)$ of $\phi(Z_i, \theta)$ is consistent under the empirical L_2 norm uniformly over $\theta \in \Theta$. Assumption 2(iii) is the identification condition of θ_0 . Assumption 2(iv) is
a uniform law of large numbers of the function $w(Z_i) |h_0(Z_i) - \phi(Z_i, \theta)|^2$ indexed by θ . Assumption 2(v)
requires that $\hat{w}_n(\cdot)$ is approximated by a sequence of nonrandom function $w_n(\cdot)$ uniformly over z. For the
consistency of the MD estimator, it is sufficient to have $\delta_{w,n} = o(1)$ in Assumption 2(v). The rate condition $\delta_{w,n} = O(n_1^{-1/4} + n_2^{-1/4})$ is needed for deriving the asymptotic normality of the MD estimator. It is clear
that Assumption 2(v) holds trivially if $\hat{w}_n(\cdot)$ is the identity function.

Theorem 1. Under Assumptions 1 and 2, we have $\hat{\theta}_n = \theta_0 + o_p(1)$.

For ease of notations, we define

$$\begin{split} g_{\theta}(X,\theta) &= \frac{\partial g(X,\theta)}{\partial \theta}, \qquad g_{\theta\theta}(X,\theta) = \frac{\partial^2 g(X,\theta)}{\partial \theta \partial \theta'}, \\ \phi_{\theta}(Z,\theta) &= E\left[g_{\theta}(X,\theta)|Z\right], \quad \phi_{\theta\theta}(Z,\theta) = E\left[g_{\theta\theta}(X,\theta)|Z\right], \\ \widehat{\phi}_{\theta,n_2}(Z,\theta) &= \frac{\partial \widehat{\phi}_{n_2}(Z,\theta)}{\partial \theta}, \qquad \widehat{\phi}_{\theta\theta,n_2}(Z,\theta) = \frac{\partial^2 \widehat{\phi}_{n_2}(Z,\theta)}{\partial \theta \partial \theta'}. \end{split}$$

By the consistency of $\widehat{\theta}_n$, there exists a positive sequence $\delta_n = o(1)$ such that $\widehat{\theta}_n \in \mathcal{N}_{\delta_n}$ with probability approaching 1, where $\mathcal{N}_{\delta_n} = \{\theta \in \Theta : ||\theta - \theta_0|| \le \delta_n\}$. Define $H_{0,n} = E[w_n(Z)\phi_\theta(Z,\theta_0)\phi_\theta'(Z,\theta_0)]$. Let $\phi_{\theta_i}(z,\theta)$

denote the *j*-th component of $\phi_{\theta}(Z, \theta)$.

We next state the sufficient conditions for asymptotic normality of $\hat{\theta}_n$.

Assumption 3. The following conditions hold:

$$\begin{split} &(i) \sup_{\theta \in \mathcal{N}_n} n_2^{-1} \sum_{i \in I_2} \|g_{\theta\theta}(X_i, \theta)\|^2 = O_p(1); \\ &(ii) \lambda_{\min}(H_{0,n}) > C^{-1}; \\ &(iii) n^{-1} \sum_{i \notin I} ||\widehat{\phi}_{\theta,n_2}(Z_i, \theta_0) - \phi_{\theta}(Z_i, \theta_0)||^2 = o_p(n_2^{-1/2}); \\ &(iv) E \left[\|\phi_{\theta}(Z, \theta_0)\|^4 \right] & \leqslant \infty; \\ &(v) E \left[u^2 | Z \right] & \leftarrow C^{-1}, E \left[\varepsilon^2 | Z \right] & \leftarrow C^{-1} and E \left[u^4 + \varepsilon^4 | Z \right] & \Leftarrow C; \\ &(vi) \sup_{z \in \mathcal{Z}} |u_n(z)\phi_{\theta_j}(z, \theta_0) - P_k(z)\beta_{w\phi_j, n, k}| = o(1) where \begin{pmatrix} \varphi_{w\phi_j, n, k} \in \mathbb{R}^k & (j = 1, \dots, d_{\theta}); \\ \varphi_{w\phi_j, n, k} \in \mathbb{R}^k & (j = 1, \dots, d_{\theta}); \\ &(vi) \max_{j = 1, 2} (k_j n_j^{-1/2} + k_j^{-r_h} n_j^{1/2}) = o(1). \end{split}$$

Assumptions 3(i) holds when $||g_{\theta\theta}(x,\theta)||^2 < C$ for any x and any θ in the local neighborhood of θ_0 . The lower bound of the eigenvalue of $H_{0,n}$ in Assumptions 3(ii) ensures the local identification of θ_0 . Assumptions 3(iii) requires that the convergence rate of $\hat{\phi}_{\theta,n_2}(Z_i,\theta_0)$ under the empirical L_2 -norm is faster than $n_2^{-1/4}$. Assumptions 3(iv) imposes finite second moment on the derivative function $\phi_{\theta}(Z,\theta_0)$. Assumption 3(v) imposes moment conditions on the projection errors u and ε which are useful for deriving the asymptotic normality of the MD estimator. Assumption 3(vi) requires that the function $w_n(z)\phi_{\theta_j}(z,\theta_0)$ can be approximated by the basis functions. Assumption 3(vii) imposes restrictions on the number of basis functions and the smoothness of the unknown function h_0 .

Let
$$\sigma_u^2(Z) = E\left[u^2 \middle| Z\right], \begin{pmatrix} \sigma_\varepsilon^2(Z) = E\left[\varepsilon^2 \middle| Z\right] \text{ and } \phi_{w\theta,n} = (w_n(Z_i)\phi_\theta(Z_i,\theta))_{i\in I}. \text{ Define} \\ \sum_{n_1} \equiv \frac{\phi_{w\theta,n}P_{n,k_1}Q_{n_1,k_1}^{-1}Q_{n_1,k_1}Q_{n_1,k_1}P'_{n,k_1}\phi'_{w\theta,n}}{n^2n_1}$$

where $Q_{n_1,u} = n_1^{-1} \sum_{i \in I_1} \sigma_u^2(Z_i) P_{k_1}(Z_i) P'_{k_1}(Z_i)$, and

$$\Sigma_{n_2} \equiv \frac{\phi_{w\theta,n} P_{n,k_2} Q_{n_2,k_2}^{-1} Q_{n_2,\epsilon_2} Q_{n_2,k_2}^{-1} P_{n,k_2}' \phi_{w\theta,n}'}{n^2 n_2}$$

where $Q_{n_2,\varepsilon} = n_2^{-1} \sum_{i \in I_2} \sigma_{\varepsilon}^2(Z_i) P_{k_2}(Z_i) P'_{k_2}(Z_i).$

Theorem 2. Under Assumptions 1, 2 and 3, we have

$$\widehat{\theta}_n - \theta_0 = O_p(n_1^{-1/2} + n_2^{-1/2}) \tag{9}$$

and moreover

$$\gamma_n' (H_{0,n} (\Sigma_{n_1} + \Sigma_{n_2})^{-1} H_{0,n})^{1/2} (\widehat{\theta}_n - \theta_0) \to_d N(0, 1)$$
(10)

for any non-random sequence $\gamma_n \in R^{d_\theta}$ with $\gamma'_n \gamma_n = 1$.

Remark 1. The first result of Theorem 2, i.e., (9), implies that the convergence rate of the MD estimator is of the order $\max\{n_1^{-1/2}, n_2^{-1/2}\}$.

Remark 2. By the Cramer-Wold device and Theorem 2, we know that

$$(H_{0,n}(\Sigma_{n_1} + \Sigma_{n_2})^{-1} H_{0,n})^{1/2} (\widehat{\theta}_n - \theta_0) \to_d N(0_{d_\theta}, I_{d_\theta}),$$
(11)

which together with the continuous mapping theorem (CMT) implies that,

$$(\widehat{\theta}_n - \theta_0)' (H_{0,n}(\Sigma_{n_1} + \Sigma_{n_2})^{-1} H_{0,n}) (\widehat{\theta}_n - \theta_0) \to_d \chi^2(d_\theta).$$
(12)

Moreover, let ι_j^* be the $d_{\theta} \times 1$ selection vector whose j-th $(j = 1, \ldots, d_{\theta})$ component is 1 and rest components are 0. Define

$$\gamma_{j,n} = \frac{(H_{0,n}(\Sigma_{n_1} + \Sigma_{n_2})^{-1}H_{0,n})^{-1/2}}{(\iota_j^{*\prime}(H_{0,n}(\Sigma_{n_1} + \Sigma_{n_2})^{-1}H_{0,n})^{-1}\iota_j^{*})^{1/2}}\iota_j^{*}, \text{ for } j = 1, \dots, d_{\theta}.$$

It is clear that $\gamma'_{j,n}\gamma_{j,n} = 1$, and by Theorem 2, we have

$$\gamma_{j,n}'(H_{0,n}(\Sigma_{n_1} + \Sigma_{n_2})^{-1}H_{0,n})^{1/2}(\widehat{\theta}_n - \theta_0) = \frac{\widehat{\theta}_{j,n} - \theta_{j,0}}{(\iota_j^{*\prime}(H_{0,n}(\Sigma_{n_1} + \Sigma_{n_2})^{-1}H_{0,n})^{-1}\iota_j^*)^{1/2}} \to_d N(0,1)$$
(13)

where $\hat{\theta}_{j,n} = \iota_j^{*'} \hat{\theta}_n$ and $\theta_{j,0} = \iota_j^{*'} \theta_0$. Results in (12) and (13) can be used to conduct inference on $\theta_{j,0}$ and θ_0 if the consistent estimators of $H_{0,n}$, Σ_{n_1} and Σ_{n_2} are available.

Optimal Weighting 4

In this section, we compare the MD estimators through their finite sample variances. The comparison leads to an optimal weight matrix which gives MD estimator with smallest finite sample variance, as well as asymptotic variance, among all MD estimators. The following lemma simplifies the finite sample variancecovariance matrix which facilitates the comparison of the MD estimators.

Lemma 1. Under Assumptions 1(i), 1(iii), 1(v), 2(v) and 3(iv)-3(vi),

$$H_{0,n}^{-1}(\Sigma_{n_1} + \Sigma_{n_2})H_{0,n}^{-1} = V_{n,\theta}(1 + o_p(1)).$$

where $V_{n,\theta} = H_{0,n}^{-1} E\left[w_n^2(Z)\left(\eta_1^{-1}\sigma_u^2(Z) + n_2^{-1}\sigma_\varepsilon^2(Z)\right)\phi_\theta(Z,\theta_0)\phi_\theta'(Z,\theta_0)\right] H_{0,n}^{-1}$. If the sequence of the weight function is set to be

$$w_n^*(Z) = (n_1^{-1} + n_2^{-1})(n_1^{-1}\sigma_u^2(Z) + n_2^{-1}\sigma_\varepsilon^2(Z))^{-1},$$
(14)

then the finite sample variance of the MD estimator becomes

$$V_{n,\theta}^{*} = E\left[\left(\frac{\phi_{u}^{2}(Z)}{n_{1}} + \frac{\sigma_{\varepsilon}^{2}(Z)}{n_{2}}\right)^{-1} \phi_{\theta}(Z,\theta_{0})\phi_{\theta}'(Z,\theta_{0})\right]\right)^{-1}.$$
(15)

The next lemma shows that V_{θ}^* is the smallest asymptotic variance-covariance of the MD estimator.

Theorem 3. For any sequence of weight functions $w_n(Z)$, we have $V_{n,\theta} \ge V_{n,\theta}^*$ for any n_1 and any n_2 .

We call the MD estimator whose finite sample variance-covariance matrix equals V_{θ}^* optimal MD estimator. To ensure the optimal MD estimator is feasible, we have to: (i) show that $C^{-1} < w_n^*(z) < C$ for any $z \in \mathcal{Z}$ and any n_1, n_2 ; and (ii) construct an empirical weight function $\widehat{w}_n^*(z)$ such that $\sup_{z \in \mathcal{Z}} |\widehat{w}_n^*(z) - w_n^*(z)| = O_p(\delta_{w,n})$, where $\delta_{w,n} = O(n_1^{-1/4} + n_2^{-1/4})$. In the rest of this section, we show that $w_n^*(z)$ is bounded from above and from below. Construction of the empirical weight function $\widehat{w}_n^*(\cdot)$ is studied in the next section.

Lemma 2. Under Assumption 3(v), $C^{-1} < w_n^*(z) < C$ for any $z \in \mathbb{Z}$ and any n_1 , n_2 .

5 Estimation of the Variance and Optimal Weighting

The estimator of the variance-covariance matrix is constructed by its sample analog. Let $\hat{u}_i = Y_i - \hat{h}_{n_1}(Z_i)$ for any $i \in I_1$, and $\hat{\varepsilon}_i = g(Z_i, \hat{\theta}_n) - \hat{\phi}_{n_2}(Z_i, \hat{\theta}_n)$ for any $i \in I_2$. Define

$$\begin{split} \widehat{H}_{n} &= n^{-1} \sum_{i \in I} (\widehat{\psi}_{n}(Z_{i}) \widehat{\phi}_{\theta,n_{2}}(Z_{i}, \widehat{\theta}_{n}) \widehat{\phi}_{\theta,n_{2}}(Z_{i}, \widehat{\theta}_{n})', \\ \widehat{\Sigma}_{n_{1}} &= \frac{\widehat{\phi}_{w\theta,n} P_{n,k_{1}} Q_{n_{1},k_{1}}^{-1} \widehat{Q}_{n_{1},u} Q_{n_{1},k_{1}}^{-1} P_{n,k_{1}}' \widehat{\phi}_{w\theta,n}'}{n^{2} n_{1}}, \\ \widehat{\Sigma}_{n_{2}} &= \frac{\widehat{\phi}_{w\theta,n} P_{n,k_{2}} Q_{n_{2},k_{2}}^{-1} \widehat{Q}_{n_{2},\epsilon_{2}} Q_{n_{2},k_{2}}^{-1} P_{n,k_{2}}' \widehat{\phi}_{w\theta,n}'}{n^{2} n_{2}}, \end{split}$$

where $\widehat{\phi}_{w\theta,n} = (\widehat{w}_n(Z_i)\widehat{\phi}_{\theta,n_2}(Z_i,\widehat{\theta}_n))_{i\in I}, \widehat{Q}_{n_1,u} = n_1^{-1}\sum_{i\in I_1}\widehat{u}_i^2 P_{k_1}(Z_i)P'_{k_1}(Z_i)$ and $\widehat{Q}_{n_2,\varepsilon} = n_2^{-1}\sum_{i\in I_2}\widehat{\varepsilon}_i^2 P_{k_2}(Z_i)P'_{k_2}(Z_i)$. The variance estimator is defined as

$$\widehat{V}_n = \widehat{H}_n^{-1} (\widehat{\Sigma}_{n_1} + \widehat{\Sigma}_{n_1}) \widehat{H}_n^{-1}.$$
(16)

The following conditions are needed to show the consistency of \hat{V}_n and the empirical optimal weight function constructed later in this section.

Assumption 4. (i) $\sup_{\theta \in \mathcal{N}_n} n_2^{-1} \sum_{i \in I_2} \|g_{\theta}(X_i, \theta)\|^2 = O_p(1);$ (ii) there exist $\beta_{u,k} \in \mathbb{R}^k$ and $r_u > 0$ such that $\sup_{z \in \mathcal{Z}} |\sigma_u^2(z) - P_k(z)' \beta_{u,k}| \notin O(k^{-r_u});$ (17)

(iii) there exist $\beta_{\varepsilon,k} \in \mathbb{R}^k$ and $r_{\varepsilon} > 0$ such that

$$\sup_{z \in \mathcal{Z}} \left| \sigma_{\varepsilon}^{2}(z) - P_{k}(z)' \beta_{\varepsilon,k} \right| \stackrel{\text{def}}{=} O(k^{-r_{\varepsilon}}); \tag{18}$$

(*iv*) $\max_{j=1,2}(\xi_{k_j}k_j^{1/2}n_j^{-1/2} + \xi_{k_j}k_j^{-r_h}) = o(1); (v) E[||g_{\theta}(X,\theta_0)||^4] \le C.$

Assumption 4(i) requires that the sample average of $||g_{\theta}(X_i, \theta)||$ is stochastically bounded uniformly over the local neighborhood of θ_0 . Assumptions 4(ii) and 4(iii) implies that the conditional variances $\sigma_u^2(z)$ and $\sigma_{\varepsilon}^2(z)$ can be approximated by the basis functions $P_k(z)$. Assumption 4(iv) imposes restrictions on the numbers of basis functions and the smoothness of the conditional variance functions. Assumption 4(v) imposes finite fourth moment on $g_{\theta}(X, \theta_0)$.

Theorem 4. Suppose Assumptions 1, 2, 3, 4(i) and 4(iv) hold. If $(k_1 + k_2)\delta_{w,n}^2 = o(1)$, then we have

$$\widehat{H}_n^{-1}(\widehat{\Sigma}_{n_1} + \widehat{\Sigma}_{n_1})\widehat{H}_n^{-1} = H_{0,n}^{-1}(\Sigma_{n_1} + \Sigma_{n_2})H_{0,n}^{-1}(1 + o_p(1))$$
(19)

and moreover,

$$\gamma_n'(\widehat{H}_n(\widehat{\Sigma}_{n_1} + \widehat{\Sigma}_{n_1})^{-1}\widehat{H}_n)^{\frac{1}{2}}(\widehat{\theta}_n - \theta_0) \to_d N(0, 1),$$
(20)

for any non-random sequence $\gamma_n \in \mathbb{R}^{d_\theta}$ with $\gamma'_n \gamma_n = 1$.

Remark 3. By the consistency of the $\widehat{H}_n(\widehat{\Sigma}_{n_1} + \widehat{\Sigma}_{n_1})^{-1}\widehat{H}_n$ and CMT,

$$(\widehat{H}_n(\widehat{\Sigma}_{n_1} + \widehat{\Sigma}_{n_1})^{-1}\widehat{H})^{1/2}(\widehat{\theta}_n - \theta_0) \to_d N(0, I_{d_\theta}),$$

which together with the CMT implies that

$$W_n(\theta_0) = (\widehat{\theta}_n - \theta_0)' (\widehat{H}_n(\widehat{\Sigma}_{n_1} + \widehat{\Sigma}_{n_1})^{-1} \widehat{H}) (\widehat{\theta}_n - \theta_0) \to_d \chi^2(d_\theta).$$
(21)

Recall that ι_j^* is the $d_{\theta} \times 1$ selection vector whose *j*-th $(j = 1, \ldots, d_{\theta})$ component is 1 and rest components are 0. By the consistency of the $\widehat{H}_n(\widehat{\Sigma}_{n_1} + \widehat{\Sigma}_{n_1})^{-1}\widehat{H}_n$, we have

$$\iota_j^{*'} \hat{H}_n(\hat{\Sigma}_{n_1} + \hat{\Sigma}_{n_1})^{-1} \hat{H}_n \iota_j^* = \iota_j^{*'} H_0^{-1} (\Sigma_{n_1} + \Sigma_{n_2}) H_0^{-1} \iota_j^* (1 + o_p(1))$$

which together with (13) and the CMT implies that

$$t_{j,n}(\theta_{j,0}) = \frac{\hat{\theta}_{j,n} - \theta_{j,0}}{\sqrt{\iota_j^{*'} \hat{H}_n(\hat{\Sigma}_{n_1} + \hat{\Sigma}_{n_1})^{-1} \hat{H} \iota_j^*}} \to_d N(0,1).$$
(22)

The Student-t statistic in (22) and the Wald-statistic in (21) can be applied to conduct inference on $\theta_{j,0}$ for $j = 1, \ldots, d_{\theta}$ and joint inference on θ_0 respectively.

Remark 4. Theorem 4 can be applied to conduct inference on θ_0 using the identity weighted MD estimator $\hat{\theta}_{1,n}$ defined as

$$\widehat{\theta}_{1,n} = \arg\min_{\theta\in\Theta} n^{-1} \sum_{i\in I} (\widehat{h}_{n_1}(Z_i) - \widehat{\phi}_{n_2}(Z_i,\theta))^2.$$
(23)

As the identity weight function satisfies Assumption 2(v) and the condition $(k_1 + k_2)\delta_{w,n}^2 = o(1)$ holds trivially, under Assumptions 1, 2(i)-(iv) and 3, Theorem 2 implies that

$$\widehat{\theta}_{1,n} - \theta_0 = O_p(n_1^{-1/2} + n_2^{-1/2}).$$
(24)

The identity weighted MD estimator can be used to construct the empirical weight function which enables us to construct the optimal MD estimator.

Let $\widehat{u}_i = Y_i - \widehat{h}_{n_1}(Z_i)$ for any $i \in I_1$, and $\widetilde{\varepsilon}_i = g(Z_i, \widehat{\theta}_{1,n}) - \widehat{\phi}_{n_2}(Z_i, \widehat{\theta}_{1,n})$ for any $i \in I_2$. Define

$$\widehat{w}_{n}^{*}(z) = (n_{1}^{-1} + n_{2}^{-1})(n_{1}^{-1}\widehat{\sigma}_{n,u}^{2}(z) + n_{2}^{-1}\widehat{\sigma}_{n,\varepsilon}^{2}(z))^{-1},$$
(25)

where $\hat{\sigma}_{n,u}^2(z)$ and $\hat{\sigma}_{n,\varepsilon}^2(z)$ are the estimators of the conditional variances $\sigma_u^2(z)$ and $\sigma_{\varepsilon}^2(z)$:

$$\widehat{\sigma}_{n,u}^{2}(z) = n_{1}^{-1} P_{k_{1}}'(z) Q_{n_{1},k_{1}}^{-1} P_{n_{1},k_{1}}' \widehat{U}_{2,n_{1}} \text{ and } \widehat{\sigma}_{n,\varepsilon}^{2}(z) = n_{2}^{-1} P_{k_{2}}'(z) Q_{n_{2},k_{2}}^{-1} P_{n_{2},k_{2}}' \widehat{e}_{2,n_{2}}, \tag{26}$$

where $\widehat{U}_{2,n_1} = (\widehat{u}_i^2)'_{i \in I_1}$ and $\widehat{e}_{2,n_2} = (\widetilde{\varepsilon}_i^2)'_{i \in I_2}$. The optimal MD estimator is defined as

$$\widehat{\theta}_n^* = \arg\min_{\theta\in\Theta} n^{-1} \sum_{i\in I} \widehat{\psi}_n^*(Z_i) (\widehat{h}_{n_1}(Z_i) - \widehat{\phi}_{n_2}(Z_i, \theta))^2.$$
(27)

To show the optimality of $\hat{\theta}_n^*$, it is sufficient to show that $\hat{w}_n^*(Z_i)$ satisfies the high level conditions in Assumption 2(v). For this purpose, we first derive the convergence rates of $\hat{\sigma}_{n,u}^2(z)$ and $\hat{\sigma}_{n,\varepsilon}^2(z)$.

Lemma 3. Under Assumptions 1, 2(i)-(iv), 3 and 4, we have

$$\sup_{z\in\mathcal{Z}} \left| \widehat{\sigma}_{n,u}^2(z) - \sigma_u^2(z) \right| \notin O_p(\xi_{k_1}(k_1^{1/2}n_1^{-1/2} + k_1^{-r_u}) + \xi_{k_1}^2k_1^{-2r_h}),$$

and

$$\sup_{z \in \mathcal{Z}} \left| \widehat{\sigma}_{n,\varepsilon}^2(z) - \sigma_{\varepsilon}^2(z) \right| = O_p(\xi_{k_2}(k_2^{1/2}n_2^{-1/2} + k_2^{-r_{\varepsilon}}) + \xi_{k_2}^2(n_1^{-1} + k_2^{-2r_h}))$$

Remark 5. Under Assumption 4(iv) and

$$\xi_{k_1} k_1^{-r_u} + \xi_{k_2} k_2^{-r_\varepsilon} + \xi_{k_2}^2 n_1^{-1} = o(1),$$
(28)

Lemma 3 implies that

$$\sup_{z\in\mathcal{Z}} \left|\widehat{\sigma}_{n,u}^2(z) - \sigma_u^2(z)\right| \stackrel{}{\models} o_p(1) \text{ and } \sup_{z\in\mathcal{Z}} \left|\widehat{\sigma}_{n,\varepsilon}^2(z) - \sigma_{\varepsilon}^2(z)\right| \stackrel{}{\models} o_p(1), \tag{29}$$

which means that $\widehat{\sigma}_{n,u}^2(z)$ and $\widehat{\sigma}_{n,\varepsilon}^2(z)$ are consistent estimators of $\sigma_u^2(z)$ and $\sigma_{\varepsilon}^2(z)$ under the uniform metric.

Theorem 5. Under (28), Assumptions 1, 2(i)-(iv), 3 and 4, we have

$$\sup_{z\in\mathcal{Z}}|\widehat{w}_n^*(z)-w^*(z)|=O_p(\delta_{w,n})$$

where $\delta_{w,n} = \max_{j=1,2} (\xi_{k_j} k_j^{1/2} n_j^{-1/2} + \xi_{k_j}^2 k_j^{-2r_h}) + \xi_{k_1} k_1^{-r_u} + \xi_{k_2} k_2^{-r_\varepsilon} + \xi_{k_2}^2 n_1^{-1}.$

Remark 6. When the power series are used as the basis functions $P_k(z)$, we have $\xi_{k_j} \leq Ck_j$. Then the convergence rate of $\delta_{w,n}$ is simplified as

$$\delta_{w,n} = \max_{j=1,2} \left(k_j^{3/2} n_j^{-1/2} + k_j^{2-2r_h} \right) + k_1^{1-r_u} + k_2^{1-r_\varepsilon} + k_2^2 n_1^{-1}.$$

Hence in this case $\delta_{w,n} = o(1)$, if $\max_{j=1,2} k_j^3 n_j^{-1} + k_2^2 n_1^{-1} = o(1)$, $r_h > 1$, $r_u > 1$ and $r_{\varepsilon} > 1$. The condition $\delta_{w,n} = O(n_1^{-1/4} + n_2^{-1/4})$ hold when

$$\max_{j=1,2} k_j^6 n_j^{-1} + k_2^{8/3} n_1^{-1} = O(1) \text{ and } \max_{j=1,2} n_j^{1/4} k_j^{2-2r_h} + n_1^{1/4} k_1^{1-r_u} + n_2^{1/4} k_2^{1-r_\varepsilon} = O(1).$$
(30)

Moreover, $(k_1 + k_2)\delta_{w,n}^2 = o(1)$ holds under (30) and $k_2^2 n_1^{-1} = o(1)$.

Remark 7. When the splines or trigonometric functions are used as the basis functions $P_k(z)$, we have $\xi_{k_j} \leq Ck_j^{1/2}$. Then the convergence rate of $\delta_{w,n}$ is simplified as

$$\delta_{w,n} = \max_{j=1,2} (k_j n_j^{-1/2} + k_j^{1-2r_h}) + k_1^{1/2-r_u} + k_2^{1/2-r_\varepsilon} + k_2 n_1^{-1}.$$

Hence in this case $\delta_{w,n} = o(1)$, if $\max_{j=1,2} k_j^2 n_j^{-1} + k_2^2 n_1^{-1} = o(1)$, $r_h > 1/2$, $r_u > 1/2$ and $r_{\varepsilon} > 1/2$. The condition $\delta_{w,n} = O(n_1^{-1/4} + n_2^{-1/4})$ hold when

$$\max_{j=1,2} k_j^4 n_j^{-1} + k_2^{8/3} n_1^{-1} = O(1) \text{ and } \max_{j=1,2} n_j^{1/4} k_j^{1-2r_h} + n_1^{1/4} k_1^{1-r_u} + n_2^{1/4} k_2^{1-r_\varepsilon} = O(1).$$
(31)

Moreover, $(k_1 + k_2)\delta_{w,n}^2 = o(1)$ holds under (31) and $k_2^2 n_1^{-1} = o(1)$.

6 Monte Carlo Simulation

In this section, we study the finite sample performances of the MD estimator and the proposed inference method. The simulated data is from the following model

$$Y_i = g(X_i, \theta_0) + v_i, \tag{32}$$

where Y_i , X_i and v_i are scale random variables, $g(X_i, \theta_0)$ is a function specified in the following

$$g(X_i, \theta_0) = \begin{cases} \begin{pmatrix} X_i \theta_0 & \text{in Model 1} \\ \log(1 + X_i^2 \theta_0), & \text{in Model 2} \end{cases},$$
(33)

where $\theta_0 = 1$ is the unknown parameter.

To generate the simulated data, we first generate $(X_{1,i}^*, X_{2,i}^*, v_i)'$ from the joint normal distribution with mean zero and identity variance-covariance matrix. Let

$$Z_i = X_{2,i}^* (1 + X_{2,i}^{*2})^{-1/2} \text{ and } X_i = Z_i + X_{1,i}^* \log(Z_i^2).$$
(34)

We assume that (Y_i, Z_i) are observed together and (X_i, Z_i) are observed together. We generate the first data set $\{(Y_i, Z_i)\}_{i \in I_1}$ with sample size n_1 , and then independently generate the second data set $\{(X_i, Z_i)\}_{i \in I_2}$ with sample size n_2 . As both the magnitudes of n_1 , n_2 and their relative magnitude are important to the finite sample properties of the MD estimator, we consider two sampling schemes: equal sampling and unequal sampling separately. In the equal sampling scheme, we set $n_1 = n_2 = n_0$ where n_0 starts from 50 with increment 50 and ends at 1000. In the unequal sampling, we set $n_1 + n_2 = 1000$ where n_1 starts from 100 with increment 50 and ends at 900. For each combination of n_1 and n_2 , we generate 10000 simulated samples to evaluate the performances of the MD estimator and the proposed inference procedure.



In addition to the MD estimator, we study two alternative estimators based on data imputation. The first estimator (which is called the Y-imputed estimator in this section) is defined as

$$\widehat{\theta}_{X,n} = \arg\min_{\theta\in\Theta} n_1^{-1} \sum_{i\in I_1} (Y_i - g(\widehat{X}_i, \theta))^2$$
(35)

where $\hat{X}_i = n_2^{-1} P'_{k_2}(Z_i) Q_{n_2,k_2}^{-1} \sum_{i \in I_2} X_i P_{k_2}(Z_i)$ for any $i \in I_1$ is the predicted value of X_i in the first data set based on nonparametric regression. The second estimator (which is called the Y-imputed estimator in this section) is defined as

$$\widehat{\theta}_{Y,n} = \arg\min_{\theta\in\Theta} n_2^{-1} \sum_{i\in I_2} (\widehat{Y}_i - g(X_i, \theta))^2$$
(36)

where $\widehat{Y}_i = n_1^{-1} P'_{k_1}(Z_i) Q_{n_1,k_1}^{-1} \sum_{i \in I_1} Y_i P_{k_1}(Z_i)$ for any $i \in I_2$ is the predicted value of Y_i in the second data

set based on nonparametric regression. In the simulation studies, we set $k_1 = k_2 = 5$ and $P_{k_1}(Z) = P_{k_2}(Z) = (1, Z, Z^2, Z^3, Z^4)$. The minimization problem in the MD estimation and the nonlinear regressions (in (35) and (36)) are solved by grid search with $\Theta = [0, 2]$ and equally spaced grid points with grid length 0.001.



Figure 6.2. Properties of the MD and the Imputation Estimators $(n_1 + n_2 = 1000)$

The finite sample properties of the identity weighted MD estimator (the green dashed line), the optimal weighted MD estimator (the black solid line), the X-imputed estimator (the blue dotted line) and the Y-imputed estimator (the red dash-dotted line) are presented in Figures 6.1 and 6.2. In Figure 6.1, we see that the bias and variance of the two MD estimators converge to zero with the growth of both n_1 and n_2 . The optimal weighted MD estimator has smaller bias and smaller variance, and hence smaller RMSE than the identity weighted MD estimator. The improvement of the optimal MD estimator over the identity weighted MD estimator is clearly investigated in model 1. The X-imputed estimator has almost the same finite sample bias and finite sample variance as the identity weighted MD estimator in the linear model (i.e., model 1). But it has large and non-convergent finite sample bias in model 2, which indicates that

the X-imputed estimator may be inconsistent in general nonlinear models. The Y-imputed estimator has large and non-convergent finite sample bias in both model 1 and model 2, which shows that it may be an inconsistent estimator in general. The finite sample performances of the MD estimators and the two imputed estimators under unequal sampling scheme are presented in Figure 6.2. In this figure, we see that when n_1 (or n_2) is small, the finite sample bias and variance of the MD estimators are large regardless how big n_2 (or n_1) is. This means that the main part in the estimation error of the MD estimator is from the component estimated by the smaller sample, which is implied by Theorem 2.



Figure 6.3. Properties of the Confidence Intervals $(n_1 = n_2)$

The finite sample properties of the inference procedures based on the identity weighted MD estimator and the optimal weighted MD estimator are provided in Figures 6.3 and 6.4. In Figure 6.3, we see that the finite coverage probabilities of the confidence intervals based on the MD estimators converge to the nominal level 0.9 with both n_1 and n_2 increase to 1000. In model 1, the coverage probability of the confidence interval based on the optimal MD estimator is closer to the nominal level than that based on the identity weighted MD estimator in all sample sizes we considered. In model 2, the confidence interval based on the optimal MD estimator is slightly worse than that based on the identity weighted MD estimator when the sample sizes n_1 and n_2 are small, and the coverage probabilities of the two confidence intervals are identical and close to the nominal level when n_1 and n_2 are larger than 250. In both model 1 and model 2, the average length of the confidence interval of the optimal MD estimator is much smaller than that of the confidence interval of the identity weighted MD estimator, which is because the optimal MD estimator has smaller variance. The finite sample performances of the confidence intervals based on the MD estimator sunder unequal sampling scheme are presented in Figure 6.4. In this figure, we see that when n_1 (or n_2) is small, the coverage probabilities of the confidence intervals of the two MD estimators are away from the nominal level. The performance of the inference based on the identity weighted MD estimator is poor in model 1 when the sample size n_2 is small regardless of the size of the other sample n_1 . From figures 6.3 and 6.4, we also see that the average length of the confidence intervals of the optimally weighted MD estimators is smaller than the identity weighted MD estimator.



Figure 6.4. Properties of the Confidence Intervals $(n_1 + n_2 = 1000)$

7 Conclusion

This paper studies estimation and inference of nonlinear econometric models when the economic variables of the models are contained in different data sets in pratice. We provide a semiparametric MD estimator based on conditional moment restrictions with common conditioning variables which are contained in different data sets. The MD estimator is show to be consistent and has asymptotic normal distribution. We provide the specific form of optimal weight for the MD estimation, and show that the optimal weighted MD estimator has the smallest asymptotic variance among all MD estimators. Consistent estimator of the variance-covariance matrix of the MD estimator, and hence inference procedure of the unknown parameter is also provided. The finite sample performances of the MD estimator and the inference procedure are investigated in simulation studies.

References

- [1] Donald WK Andrews. Asymptotic normality of series estimators for nonparametric and semiparametric regression models. *Econometrica: Journal of the Econometric Society*, pages 307–345, 1991.
- [2] Joshua D Angrist and Alan B Krueger. The effect of age at school entry on educational attainment: an application of instrumental variables with moments from two samples. *Journal of the American statistical Association*, 87(418):328–336, 1992.
- [3] Manuel Arellano and Costas Meghir. Female labour supply and on-the-job search: an empirical model estimated using complementary data sets. *The Review of Economic Studies*, 59(3):537–559, 1992.
- [4] Orazio P Attanasio and Guglielmo Weber. Consumption growth, the interest rate and aggregation. The Review of Economic Studies, 60(3):631–649, 1993.
- [5] Richard Blundell, Luigi Pistaferri, and Ian Preston. Consumption inequality and partial insurance. American Economic Review, 98(5):1887–1921, 2008.
- [6] Martin Browning, Angus Deaton, and Margaret Irish. A profitable approach to labor supply and commodity demands over the life-cycle. *Econometrica: journal of the econometric society*, pages 503– 543, 1985.
- [7] Martin Browning and Søren Leth-Petersen. Imputing consumption from income and wealth information. The Economic Journal, 113(488), 2003.
- [8] Xiaohong Chen. Large sample sieve estimation of semi-nonparametric models. Handbook of econometrics, 6:5549–5632, 2007.
- John H Cochrane. A simple test of consumption insurance. Journal of political economy, 99(5):957–976, 1991.
- [10] Yanqin Fan, Robert Sherman, and Matthew Shum. Identifying treatment effects under data combination. *Econometrica*, 82(2):811–822, 2014.
- [11] Yanqin Fan, Robert Sherman, and Matthew Shum. Estimation and inference in an ecological inference model. Journal of Econometric Methods, 5(1):17–48, 2016.
- [12] Robert E Hall and Frederic S Mishkin. The sensitivity of consumption to transitory income: Estimates from panel data on households. *Econometrica*, 50(2):461–481, 1982.
- [13] Fumio Hayashi, Joseph Altonji, and Laurence Kotlikoff. Risk-sharing between and within families. Econometrica: Journal of the Econometric Society, pages 261–294, 1996.
- [14] Erik Hurst and Frank P Stafford. Home is where the equity is: Mortgage refinancing and household consumption. Journal of money, Credit, and Banking, 36(6):985–1014, 2004.
- [15] N Anders Klevmarken. Missing variables and two-stage least-squares estimation from more than one data set. Technical report, Research Institute of Industrial Economics, 1982.
- [16] Robert Martin. Consumption, durable goods, and transaction costs. 2003.

- [17] Whitney K Newey. Convergence rates and asymptotic normality for series estimators. Journal of econometrics, 79(1):147–168, 1997.
- [18] Geert Ridder and Robert Moffitt. The econometrics of data combination. Handbook of econometrics, 6:5469–5547, 2007.
- [19] David E Runkle. Liquidity constraints and the permanent-income hypothesis: Evidence from panel data. Journal of monetary Economics, 27(1):73–98, 1991.
- [20] John Shea. Myopia, liquidity constraints, and aggregate consumption: a simple test. *Journal of money*, credit and banking, 27(3):798–805, 1995.
- [21] Jonathan Skinner. A superior measure of consumption from the panel study of income dynamics. *Economics Letters*, 23(2):213–216, 1987.
- [22] Andreas Waldkirch, Serena Ng, and Donald Cox. Intergenerational linkages in consumption behavior. Journal of Human Resources, 39(2):355–381, 2004.
- [23] Stephen P Zeldes. Optimal consumption with stochastic income: Deviations from certainty equivalence. The Quarterly Journal of Economics, 104(2):275–298, 1989.
- [24] James P Ziliak. Does the choice of consumption measure matter? an application to the permanentincome hypothesis. Journal of monetary Economics, 41(1):201–216, 1998.

APPENDIX

A Proof of the Main Results in Section 3

Proof. [Proof of Theorem 1] Define the empirical criterion function of the MD estimation problem as

$$\widehat{L}_{n}(\theta) = n^{-1} \sum_{i \in I} \left[\widehat{w}_{n}(Z_{i}) \left| \widehat{h}_{n_{1}}(Z_{i}) - \widehat{\phi}_{n_{2}}(Z_{i}, \theta) \right|^{2} \right] \text{ for any } \theta \in \Theta.$$

$$(37)$$

By Assumptions 2(iii) and 2(v),

$$\inf_{\{\theta \in \Theta: \ ||\theta - \theta_0|| \ge \varepsilon\}} L_n^*(\theta) \ge \eta_{C,\varepsilon}$$
(38)

where $\eta_{C,\varepsilon} = C\eta_{\varepsilon} > 0$ is a fixed constant which only depends on ε . (38) implies that θ_0 is uniquely identified as the minimizer of $L_n^*(\theta)$. Hence, to prove the consistency of $\hat{\theta}_n$, it is sufficient to show that

$$\sup_{\theta \in \Theta} \left| \widehat{L}_n(\theta) - L_n^*(\theta) \right| = o_p(1).$$
(39)

Note that we can decompose $L_n(\theta)$ as

$$\widehat{L}_{n}(\theta) = n^{-1} \sum_{i \in I} \left(\widehat{v}_{n}(Z_{i})(|\widehat{h}_{n_{1}}(Z_{i}) - h_{0}(Z_{i})|^{2} + |\widehat{\phi}_{n_{2}}(Z_{i},\theta) - \phi(Z_{i},\theta)|^{2} + |h_{0}(Z_{i}) - \phi(Z_{i},\theta)|^{2}) - 2n^{-1} \sum_{i \in I} \left(\widehat{v}_{n}(Z_{i})(\widehat{h}_{n_{1}}(Z_{i}) - h_{0}(Z_{i}))(\widehat{\phi}_{n_{2}}(Z_{i},\theta) - \phi(Z_{i},\theta)) - 2n^{-1} \sum_{i \in I} \left(\widehat{v}_{n}(Z_{i})(\widehat{\phi}_{n_{2}}(Z_{i},\theta) - \phi(Z_{i},\theta))(h_{0}(Z_{i}) - \phi(Z_{i},\theta)) + 2n^{-1} \sum_{i \in I} \left(\widehat{v}_{n}(Z_{i})(\widehat{h}_{n_{1}}(Z_{i}) - h_{0}(Z_{i}))(h_{0}(Z_{i}) - \phi(Z_{i},\theta)) \right) \right) \right)$$

$$(40)$$

Using Assumption 1(i), one can use Rudelson's law of large numbers for matrices (see, e.g., Lemma 6.2 in Belloni, et. al. (2015)) to get

$$Q_{n,k_j} - Q_{k_j} = O_p(n^{-1/2}\xi_{k_j}(\log(k_j))^{1/2}) \text{ and } Q_{n_j,k_j} - Q_{k_j} = O_p(n_j^{-1/2}\xi_{k_j}(\log(k_j))^{1/2})$$
(41)

where $Q_{n,k_j} = n^{-1} \sum_{i \in I} P_{k_j}(Z_i) P'_{k_j}(Z_i)$, $Q_{n_j,k_j} = n_1^{-1} \sum_{i \in I_j} P_{k_j}(Z_i) P'_{k_j}(Z_i)$ and the convergence is under the operator norm of matrix. By (41), Assumptions 1(iii) and 1(v),

$$C^{-1} \le \lambda_{\min}(Q_{n,k_j}) \le \lambda_{\max}(Q_{n,k_j}) \le C \text{ and } C^{-1} \le \lambda_{\min}(Q_{n_j,k_j}) \le \lambda_{\max}(Q_{n_j,k_j}) \le C,$$
(42)

with probability approaching 1. Under Assumption 1 and (42), (A.2) in the proof of Theorem 1 in Newey (1997) implies that

$$\left\|\widehat{\beta}_{k_1,n_1} - \beta_{h,k_1}\right\|_{\ell}^2 = O_p(k_1 n_1^{-1} + k_1^{-2r_h}).$$
(43)

By the triangle inequality,

$$n^{-1} \sum_{i \in I} \left| \widehat{h}_{n_{1}}(Z_{i}) - h_{0}(Z_{i}) \right|^{2} \left(\sum_{i \in I} |\widehat{h}_{n_{1}}(Z_{i}) - h_{0,k_{1}}(Z_{i})|^{2} + 2n^{-1} \sum_{i \in I} |h_{0,k_{1}}(Z_{i}) - h_{0}(Z_{i})|^{2} \right) \right)^{2} \leq 2n^{-1} \sum_{i \in I} |\widehat{h}_{0,k_{1}}(Z_{i}) - h_{0}(Z_{i})|^{2} \\ \leq 2(\widehat{\beta}_{k_{1},n_{1}} - \beta_{h,k_{1}})'Q_{k_{1},n}(\widehat{\beta}_{k_{1},n_{1}} + 2\sup_{z \in \mathcal{Z}} |h_{0,k_{1}}(z) - h_{0}(z)|^{2} \\ \leq 2\lambda_{\max}(Q_{k_{1},n}) \left\| \widehat{\beta}_{k_{1},n_{1}} - \beta_{h,k_{1}} \right\|^{2} + 2\sup_{z \in \mathcal{Z}} |h_{0,k_{1}}(z) - h_{0}(z)|^{2} \\ = O_{p}(k_{1}n_{1}^{-1} + k_{1}^{-2r_{h}}) = o_{p}(1)$$

$$(44)$$

where the first equality is by Assumption 1(iv), (42) and (43), the second equality is by Assumption 1(v).

By the triangle inequality and Assumption 2(v),

$$\sup_{z \in \mathcal{Z}} |\widehat{w}_n(z)| \le \sup_{z \in \mathcal{Z}} |\widehat{w}_n(z) - w_n(z)| + \sup_{z \in \mathcal{Z}} |w_n(z)| < 2C$$

$$\tag{45}$$

with probability approaching 1. By (44) and (45),

$$n^{-1} \sum_{i \in I} \widehat{w}_n(Z_i) \left| \widehat{h}_{n_1}(Z_i) - h_0(Z_i) \right|^2 \le \sup_{z \in \mathcal{Z}} \left| \widehat{w}_n(z) \right| \sum_{i \in I} \left| \widehat{h}_{n_1}(Z_i) - h_0(Z_i) \right|^2 = o_p(1).$$
Assumption 2(ii)
$$(46)$$

By (45) and Assumption 2(ii),

$$\sup_{\theta \in \Theta} n^{-1} \sum_{i \in I} \widehat{w}_n(Z_i) \left| \widehat{\phi}_{n_2}(Z_i, \theta) - \phi(Z_i, \theta) \right|^2$$

$$\leq \sup_{z \in \mathcal{Z}} \left| \widehat{w}_n(z) \right| \sup_{\theta \in \Theta} n^{-1} \sum_{i \in I} \left| \widehat{\phi}_{n_2}(Z_i, \theta) - \phi(Z_i, \theta) \right|^2 = o_p(1).$$
(47)

Using (44), (45), Assumption 2(ii) and the Cauchy-Schwarz inequality, we get

$$\sup_{\theta \in \Theta} \left| n^{-1} \sum_{i \in I} \left| \widehat{\psi}_{n}(Z_{i})(\widehat{h}_{n_{1}}(Z_{i}) - h_{0}(Z_{i}))(\widehat{\phi}_{n_{2}}(Z_{i},\theta) - \phi(Z_{i},\theta)) \right| \\
\leq \sup_{z \in \mathcal{Z}} \left| \widehat{w}_{n}(z) \right| \sqrt{n^{-1} \sum_{i \in I} \left| \widehat{h}_{n_{1}}(Z_{i}) - h_{0}(Z_{i}) \right|^{2}} \sqrt{\sup_{\theta \in \Theta} n^{-1} \sum_{i \in I} \left| \widehat{\phi}_{n_{2}}(Z_{i},\theta) - \phi(Z_{i},\theta) \right|^{2}} = o_{p}(1). \quad (48)$$

By Assumption 1(ii), $E\left[h_0^2(Z)\right] < C$, which together with Assumption 2(i) implies that

$$\sup_{\theta \in \Theta} E\left[\left| h_0(Z) - \phi(Z, \theta) \right|^2 \right] \le 2E\left[h_0^2(Z) \right] + 2\sup_{\theta \in \Theta} E\left[\phi^2(Z, \theta) \right] \stackrel{\text{c}}{\leftarrow} C.$$
(49)

By (49), Assumptions 2.(iv) and 2.(v),

$$\sup_{\theta \in \Theta} n^{-1} \sum_{i \in I} \left(h_0(Z_i) - \phi(Z_i, \theta) \right)^2
\leq C \sup_{\theta \in \Theta} n^{-1} \sum_{i \in I} \left(\psi_n(Z_i) \left| h_0(Z_i) - \phi(Z_i, \theta) \right|^2 \right)
\leq (C + o_p(1)) \sup_{\theta \in \Theta} E \left[\psi_n(Z_i) \left| h_0(Z_i) - \phi(Z_i, \theta) \right|^2 \right]
\leq (C + o_p(1)) \sup_{\theta \in \Theta} E \left[\left| h_0(Z_i) - \phi(Z_i, \theta) \right|^2 \right] = O_p(1).$$
(50)

Using (45), (50), Assumptions 2(ii) and (iv), and the Cauchy-Schwarz inequality, we get

$$\sup_{\theta \in \Theta} \left| n^{-1} \sum_{i \in I} \left[\psi_n(Z_i)(\widehat{\phi}_{n_2}(Z_i, \theta) - \phi(Z_i, \theta))(h_0(Z_i) - \phi(Z_i, \theta)) \right] \\
\leq \sup_{z \in \mathcal{Z}} \left| \widehat{\psi}_n(z) \right| \sup_{\theta \in \Theta} \sqrt{n^{-1} \sum_{i \in I} \left| h_0(Z_i) - \phi(Z_i, \theta) \right|^2} \sqrt{n^{-1} \sum_{i \in I} \left| \widehat{\phi}_{n_2}(Z_i, \theta) - \phi(Z_i, \theta) \right|^2} \right] = o_p(1). \quad (51)$$

Similarly, using (45), (44), (50), Assumptions 2(iv) and the Cauchy-Schwarz inequality, we get

$$\sup_{\theta \in \Theta} \left| n^{-1} \sum_{i \in I} \left(\psi_n(Z_i)(\widehat{h}_{n_1}(Z_i) - h_0(Z_i))(h_0(Z_i) - \phi(Z_i, \theta)) \right| = o_p(1).$$
(52)
sults in (40), (46), (47), (48), (51) and (52), we get

Collecting the results in (40), (46), (47), (48), (51) and (52), we get

$$\sup_{\theta \in \Theta} L_n(\theta) = \sup_{\theta \in \Theta} n^{-1} \sum_{i \in I} \oint_{\infty} \left(h_0(Z_i) \left| h_0(Z_i) - \phi(Z_i, \theta) \right|^2 + o_p(1) \right).$$
(53)

By (50) and Assumption 2(v),

$$\sup_{\theta \in \Theta} \left| n^{-1} \sum_{i \in I} \left(\widehat{w}_n(Z_i) - w_n(Z_i) \right) \left| h_0(Z_i) - \phi(Z_i, \theta) \right|^2 \right| \\
\leq \sup_{z \in \mathcal{Z}} \left| \widehat{\psi}_n(z) - w_n(z) \right| \sup_{\theta \in \Theta} n^{-1} \sum_{i \in I} \left| h_0(Z_i) - \phi(Z_i, \theta) \right|^2 \left| = o_p(1) \quad (54)$$

which together with (53) and Assumption 2(iv),

$$\sup_{\theta \in \Theta} \left| \widehat{L}_n(\theta) - L_n^*(\theta) \right| = \sup_{\theta \in \Theta} \left| L_n(\theta) - L_n^*(\theta) \right| + o_p(1) = o_p(1).$$
(55)

This proves (39) and hence the claim of the theorem.

Lemma 4. By Assumptions 1(i), 1(iii), 1(v) and 3(i), we have

$$\sup_{\theta \in \mathcal{N}_{\delta_n}} n^{-1} \sum_{i \in I} \left\| \widehat{\phi}_{\theta\theta, n_2}(Z_i, \theta) \right\|_{\ell}^2 = O_p(1).$$

Proof. [Proof of Lemma 4] By definition,

$$\widehat{\phi}_{\theta\theta,n_2}(z,\theta) = n_2^{-1} P_{k_2}(z)' Q_{n_2,k_2}^{-1} \sum_{i \in I} \oint_{k_2} (Z_i) g_{\theta\theta}(X_i,\theta).$$

Let $g_{\theta_{j_1}\theta_{j_2}}(X_i, \theta)$ denote the (j_1, j_2) -th component of $g_{\theta\theta}(X_i, \theta)$, for any $j_1 = 1, \ldots, d_{\theta}$ and any $j_2 = 1, \ldots, d_{\theta}$. Let

$$\widehat{\phi}_{\theta_{j_1}\theta_{j_2},n_2}(z,\theta) = n_2^{-1} P'_{k_2}(z) Q_{n_2,k_2}^{-1} P'_{n_2,k_2} g_{\theta_{j_1}\theta_{j_2},n_2}(\theta),$$

where $g_{\theta_{j_1}\theta_{j_2},n_2}(\theta) = (g_{\theta_{j_1}\theta_{j_2}}(X_i,\theta))'_{i\in I_2}$. Then by definition,

$$n^{-1} \sum_{i \in I} \oint_{\theta_{j_1}\theta_{j_2}, n_2}^2 (Z_i, \theta) = g_{\theta_{j_1}\theta_{j_2}, n_2}(\theta)' P_{n_2, k_2} Q_{n_2, k_2}^{-1} Q_{n, k_2} Q_{n_2, k_2}^{-1} P'_{n_2, k_2} g_{\theta_{j_1}\theta_{j_2}, n_2}(\theta)$$

$$\leq \frac{\lambda_{\max}(Q_{n, k_2})}{\lambda_{\min}(Q_{n_1, k_2})} \frac{g_{\theta_{j_1}\theta_{j_2}, n_2}(\theta)' P_{n_2, k_2}(P'_{n_2, k_2} P_{n_2, k_2})^{-1} P'_{n_2, k_2} g_{\theta_{j_1}\theta_{j_2}, n_2}(\theta)}{n_2}$$

$$\leq \frac{\lambda_{\max}(Q_{n, k_2})}{\lambda_{\min}(Q_{n_1, k_2})} n_2^{-1} \sum_{i \in I_2} (g_{\theta_{j_1}\theta_{j_2}}(X_i, \theta))^2$$

which together with (42) (which holds under Assumptions 1(i), 1(iii) and 1(v)), and Assumption 3(i) implies that

$$\sup_{\theta \in \mathcal{N}_{\delta_n}} n^{-1} \sum_{i \in I} \widehat{\phi}_{\theta_{j_1}\theta_{j_2}, n_2}^2(Z_i, \theta) \le \frac{\lambda_{\max}(Q_{n, k_2})}{\lambda_{\min}(Q_{n_1, k_2})} \sup_{\theta \in \mathcal{N}_{\delta_n}} n_2^{-1} \sum_{i \in I_2} (g_{\theta_{j_1}\theta_{j_2}}(X_i, \theta))^2 = O_p(1)$$

for any $j_1 = 1, \ldots, d_{\theta}$ and any $j_2 = 1, \ldots, d_{\theta}$. This finishes the proof.

Lemma 5. By Assumptions 1(i), 1(iii), 1(iv), 1(v), 3(v) and 3(vii), we have

$$n^{-1} \sum_{i \in I} (\widehat{h}_{n_1}(Z_i) - \widehat{\phi}_{n_2}(Z_i, \theta_0))^2 = o_p(n_1^{-1/2} + n_2^{-1/2}).$$

Proof. [Proof of Lemma 5] By (44) (which holds under Assumptions 1(i), 1(iii), 1(iv) and 1(v)),

$$n^{-1} \sum_{i \in I} (\widehat{h}_{n_1}(Z_i) - \widehat{\phi}_{n_2}(Z_i, \theta_0))^2$$

$$\leq 2n^{-1} \sum_{i \in I} (\widehat{h}_{n_1}(Z_i) - h_0(Z_i))^2 + 2n^{-1} \sum_{i \in I} (\widehat{\phi}_{n_2}(Z_i, \theta_0) - h_0(Z_i))^2$$

$$= 2n^{-1} \sum_{i \in I} (\widehat{\phi}_{n_2}(Z_i, \theta_0) - h_0(Z_i))^2 + O_p(k_1 n_1^{-1} + k_1^{-2r_h}).$$
(56)

Let $\widehat{\beta}_{\phi,n_2} = (P'_{n_2,k_2}P_{n_2,k_2})^{-1}P'_{n_2,k_2}g_{n_2}(\theta_0)$, where $g_{n_2}(\theta_0) = (g(X_i,\theta_0))'_{i\in I_2}$. Then

$$\left\|\widehat{\beta}_{\phi,n_{2}}-\beta_{h,k_{2}}\right\|^{2} = (g_{n_{2}}(\theta_{0})-H_{n_{2},k_{2}})'P_{n_{2},k_{2}}(P_{n_{2},k_{2}}P_{n_{2},k_{2}})^{-2}P_{n_{2},k_{2}}'(g_{n_{2}}(\theta_{0})-H_{n_{2},k_{2}}) \left(\leq \frac{(g_{n_{2}}(\theta_{0})-H_{n_{2},k_{2}})'P_{n_{2},k_{2}}(P_{n_{2},k_{2}}P_{n_{2},k_{2}})^{-1}P_{n_{2},k_{2}}'(g_{n_{2}}(\theta_{0})-H_{n_{2},k_{2}})}{n_{2}\lambda_{\min}(Q_{n_{2},k_{2}})},$$
(57)

		1
		L
		L

where $H_{n_2,k_2} = (h_{0,k_2}(Z_i))'_{i \in I_2}$. By Assumptions 1(iv),

$$n_{2}^{-1}(H_{n_{2}} - H_{n_{2},k_{2}})'P_{n_{2},k_{2}}(P_{n_{2},k_{2}}'P_{n_{2},k_{2}})^{-1}P_{n_{2},k_{2}}'(H_{n_{2}} - H_{n_{2},k_{2}})$$

$$\leq n_{2}^{-1}(H_{n_{2}} - H_{n_{2},k_{2}})'(H_{n_{2}} - H_{n_{2},k_{2}}) = O(k_{2}^{-2r_{h}}),$$
(58)

where $H_{n_2} = (h_0(Z_i))'_{i \in I_2}$. By Assumptions 1(i), 1(iii) and 3(v),

$$E\left[n_{2}^{-1}(g_{n_{2}}(\theta_{0})-H_{n_{2}})'P_{n_{2},k_{2}}(P_{n_{2},k_{2}}'P_{n_{2},k_{2}})^{-1}P_{n_{2},k_{2}}'(g_{n_{2}}(\theta_{0})-H_{n_{2}})|\{Z_{i}\}_{i\in I_{2}}\right]\left(= n_{2}^{-1}tr\left(\left(P_{n_{2},k_{2}}'P_{n_{2},k_{2}}\right)^{-1}P_{n_{2},k_{2}}'E\left[(g_{n_{2}}(\theta_{0})-H_{n_{2}})(g_{n_{2}}(\theta_{0})-H_{n_{2}})'|\{Z_{i}\}_{i\in I_{2}}\right]R_{n_{2},k_{2}}\right) \\ \leq \sup_{z\in\mathcal{Z}}\sigma_{\varepsilon}^{2}(x)k_{2}n_{2}^{-1} = O(k_{2}n_{2}^{-1})$$

which together with the Markov inequality implies that

$$n_2^{-1}(g_{n_2}(\theta_0) - H_{n_2})' P_{n_2,k_2}(P'_{n_2,k_2}P_{n_2,k_2})^{-1} P'_{n_2,k_2}(g_{n_2}(\theta_0) - H_{n_2}) = O_p(k_2 n_2^{-1}).$$
(59)

Combining the results in (57), (58) and (59), and then applying (42), we get

$$\left\|\widehat{\beta}_{\phi,n_2} - \beta_{h,k_2}\right\|_{\ell}^2 = O_p(k_2 n_2^{-1} + k_2^{-2r_h}).$$
(60)

By (42) and (60)

$$n^{-1} \sum_{i \in I} \left(\widehat{\phi}_{n_2}(Z_i, \theta_0) - h_{0,k_2}(Z_i))^2 \\ = (\widehat{\beta}_{\phi,n_2} - \beta_{h,k_2})' Q_{n,k_2} (\widehat{\beta}_{\phi,n_2} - \beta_{h,k_2}) \\ \le \lambda_{\max}(Q_{n,k_2}) \left\| \widehat{\beta}_{\phi,n_2} - \beta_{h,k_2} \right\|_{\ell}^2 = O_p(k_2 n_2^{-1} + k_2^{-2r_h})$$
(61)

which together with (56) and Assumption 3(vii) proves the claim of the lemma.

Lemma 6. Under Assumptions 1(i), 2(v), 3(iv) and 3(vi), we have

$$n^{-1}\phi_{w\theta,n}P_{n,k_1}(P'_{n,k_1}P_{n,k_1})^{-1}P'_{n,k_1}\phi'_{w\theta,n} = E\left[w_n^2(Z)\phi_\theta(Z,\theta_0)\phi'_\theta(Z,\theta_0)\right] \not\models o_p(1).$$

Proof. [Proof of Lemma 6] For $j = 1, \ldots, d_{\theta}$, let $\phi_{w\theta_j,k_1,n}(z,\theta_0) = P'_k(z)\beta_{w\phi_j,k_1}, \phi_{w\theta_j,k_1,n} = (\phi_{w\theta_j,k_1,n}(Z_i,\theta_0))_{i\in I}$ and $\phi_{w\theta,k_1,n} = (\phi'_{w\theta_j,k_1,n})'_{j=1,\ldots,d_{\theta}}$. For ease of notations, we define $M_{k_1,n} = P_{n,k_1}(P'_{n,k_1}P_{n,k_1})^{-1}P'_{n,k_1}$.

By definition,

$$\phi_{w\theta,n}M_{k_{1},n}\phi_{w\theta,n}' = \phi_{w\theta,k_{1},n}M_{k_{1},n}\phi_{w\theta,k_{1},n}' + (\phi_{w\theta,n} - \phi_{w\theta,k_{1},n})M_{k_{1},n}(\phi_{w\theta,n} - \phi_{w\theta,k_{1},n})' \\
+ (\phi_{w\theta,n} - \phi_{w\theta,k_{1},n})M_{k_{1},n}\phi_{w\theta,k_{1},n}' + \phi_{w\theta,k_{1},n}M_{k_{1},n}(\phi_{w\theta,n} - \phi_{w\theta,k_{1},n})'.$$
(62)

For any $j = 1, \ldots, d_{\theta}$, let $\phi_{w\theta_j,n}$ denote the *j*-th row of $\phi_{w\theta,n}$. By the Cauchy-Schwarz inequality, for any

 $j_1 = 1, \ldots, d_{\theta}$ and any $j_2 = 1, \ldots, d_{\theta}$,

$$\left| n^{-1} (\phi_{w\theta_{j_{1}},n} - \phi_{w\theta_{j},k_{1},n}) M_{k_{1},n} (\phi_{w\theta_{j_{2}},n} - \phi_{w\theta_{j_{2}},k_{1},n})' \right|^{2} \\ \leq n^{-1} (\phi_{w\theta_{j_{1}},n} - \phi_{w\theta_{j},k_{1},n}) M_{k_{1},n} (\phi_{w\theta_{j_{1}},n} - \phi_{w\theta_{j},k_{1},n})' \\ \times n^{-1} (\phi_{w\theta_{j_{2}},n} - \phi_{w\theta_{j_{2}},k_{1},n}) M_{k_{1},n} (\phi_{w\theta_{j_{2}},n} - \phi_{w\theta_{j_{2}},k_{1},n})' \\ \leq n^{-1} \sum_{i \in I} \left| w_{n}(Z_{i}) \phi_{\theta_{j_{1}}}(Z_{i},\theta_{0}) - \phi_{w\theta_{j_{1}},k_{1}}(Z_{i},\theta_{0}) \right|^{2} \\ \times n^{-1} \sum_{i \in I} \left| w_{n}(Z_{i}) \phi_{\theta_{j_{2}}}(Z_{i},\theta_{0}) - \phi_{w\theta_{j_{2}},k_{1}}(Z_{i},\theta_{0}) \right|^{2} = o(1)$$
(63)

where the last equality is by Assumption 3(vi), and the fact that $M_{k_1,n}$ is an idempotent matrix. (63) then implies that

$$n^{-1} (\phi_{w\theta,n} - \phi_{w\theta,k_1,n}) M_{k_1,n} (\phi_{w\theta,n} - \phi_{w\theta,k_1,n}) = o(1).$$
(64)

For any $j_1 = 1, \ldots, d_{\theta}$ and any $j_2 = 1, \ldots, d_{\theta}$, by definition we can write

$$n^{-1}\phi_{w\theta_{j_{1}},k_{1,n}}M_{k_{1,n}}\phi'_{w\theta_{j_{2}},k_{1,n}}$$

$$= n^{-1}\beta'_{w\phi_{j_{1}},k}P'_{n,k_{1}}P_{n,k_{1}}(P'_{n,k_{1}}P_{n,k_{1}})^{-1}P'_{n,k_{1}}\phi'_{w\theta_{j_{2}},k_{1,n}}$$

$$= n^{-1}\sum_{i\in I}\phi_{w\theta_{j_{1}},k_{1,n}}(Z_{i},\theta_{0})\phi_{w\theta_{j_{2}},k_{1,n}}(Z_{i},\theta_{0})$$

$$= n^{-1}\sum_{i\in I}\phi_{u}^{2}(Z_{i})\phi_{\theta_{j_{1}}}(Z_{i},\theta_{0})\phi_{\theta_{j_{2}}}(Z_{i},\theta_{0})$$

$$+ n^{-1}\sum_{i\in I}\phi_{w\theta_{j_{1}},k_{1,n}}(Z_{i},\theta_{0}) - w_{n}(Z_{i})\phi_{\theta_{j_{1}}}(Z_{i},\theta_{0}))\phi_{w\theta_{j_{2}},k_{1,n}}(Z_{i},\theta_{0})$$

$$+ n^{-1}\sum_{i\in I}\phi_{u}(Z_{i})\phi_{\theta_{j_{1}}}(Z_{i},\theta_{0})(\phi_{w\theta_{j_{2}},k_{1}}(Z_{i},\theta_{0}) - w_{n}(Z_{i})\phi_{\theta_{j_{2}}}(Z_{i},\theta_{0})).$$
(65)

By Assumptions 1(i), 2(v), 3(iv) and the Markov inequality, we have

$$n^{-1} \sum_{i \in I} w_n^2(Z_i) \phi_{\theta_{j_1}}(Z_i, \theta_0) \phi_{\theta_{j_2}}(Z_i, \theta_0) - E\left[w_n^2(Z_i) \phi_{\theta_{j_1}}(Z_i, \theta_0) \phi_{\theta_{j_2}}(Z_i, \theta_0)\right] = O_p(n^{-1/2}), \tag{66}$$

where under Assumptions 2(v) and 3(iv)

$$\left| E\left[w_n^2(Z_i)\phi_{\theta_{j_1}}(Z_i,\theta_0)\phi_{\theta_{j_2}}(Z_i,\theta_0) \right] \right| < C.$$

$$\tag{67}$$

By Assumption 3(vi),

$$\left| \begin{array}{c} n^{-1} \sum_{i \in I} \left(\phi_{w\theta_{j_1},k_1,n}(Z_i,\theta_0) - w_n(Z_i)\phi_{\theta_{j_1}}(Z_i,\theta_0))(\phi_{w\theta_{j_2},k_1,n}(Z_i,\theta_0) - w_n(Z_i)\phi_{\theta_{j_2}}(Z_i,\theta_0)) \right| \\ \leq \left(\max_{j=1,\dots,d_\theta} \sup_{z \in \mathcal{Z}} |\phi_{w\theta_j,k_1,n}(z,\theta_0) - w_n(z)\phi_{\theta_j}(z,\theta_0)| \right)^2 = o(1), \end{array} \right|$$
(68)

which implies that

$$n^{-1} \sum_{i \in I} \left(\phi_{w\theta_{j},k_{1},n}(Z_{i},\theta_{0}) - w_{n}(Z_{i})\phi_{\theta_{j_{1}}}(Z_{i},\theta_{0}))\phi_{w\theta_{j_{2}},k_{1},n}(Z_{i},\theta_{0}) \right) \\ = n^{-1} \sum_{i \in I} \left(\phi_{w\theta_{j},k_{1},n}(Z_{i},\theta_{0}) - w_{n}(Z_{i})\phi_{\theta_{j_{1}}}(Z_{i},\theta_{0}))w_{n}(Z_{i})\phi_{\theta_{j_{2}}}(Z_{i},\theta_{0}) + o_{p}(1). \right)$$
(69)

By the Cauchy-Schwarz inequality,

$$\left| n^{-1} \sum_{i \in I} \left(\phi_{w\theta_{j},k_{1},n}(Z_{i},\theta_{0}) - w_{n}(Z_{i})\phi_{\theta_{j_{1}}}(Z_{i},\theta_{0}) \right) w_{n}(Z_{i})\phi_{\theta_{j_{2}}}(Z_{i},\theta_{0}) \right|^{2} \left(n^{-1} \sum_{i \in I} \left| \phi_{w\theta_{j},k_{1},n}(Z_{i},\theta_{0}) - w_{n}(Z_{i})\phi_{\theta_{j_{1}}}(Z_{i},\theta_{0}) \right|^{2} n^{-1} \sum_{i \in I} \left| \psi_{n}^{2}(Z_{i})\phi_{\theta_{j_{2}}}(Z_{i},\theta_{0}) - w_{n}(Z_{i})\phi_{\theta_{j_{1}}}(Z_{i},\theta_{0}) \right|^{2} n^{-1} \sum_{i \in I} \left| \psi_{n}^{2}(Z_{i})\phi_{\theta_{j_{2}}}(Z_{i},\theta_{0}) - w_{n}(Z_{i})\phi_{\theta_{j_{1}}}(Z_{i},\theta_{0}) \right|^{2} n^{-1} \sum_{i \in I} \left| \psi_{n}^{2}(Z_{i})\phi_{\theta_{j_{2}}}(Z_{i},\theta_{0}) - w_{n}(Z_{i})\phi_{\theta_{j_{1}}}(Z_{i},\theta_{0}) \right|^{2} n^{-1} \sum_{i \in I} \left| \psi_{n}^{2}(Z_{i})\phi_{\theta_{j_{2}}}(Z_{i},\theta_{0}) - w_{n}(Z_{i})\phi_{\theta_{j_{1}}}(Z_{i},\theta_{0}) \right|^{2} n^{-1} \sum_{i \in I} \left| \psi_{n}^{2}(Z_{i})\phi_{\theta_{j_{2}}}(Z_{i},\theta_{0}) - \psi_{n}(Z_{i})\phi_{\theta_{j_{1}}}(Z_{i},\theta_{0}) \right|^{2} n^{-1} \sum_{i \in I} \left| \psi_{n}^{2}(Z_{i})\phi_{\theta_{j_{2}}}(Z_{i},\theta_{0}) - \psi_{n}(Z_{i})\phi_{\theta_{j_{1}}}(Z_{i},\theta_{0}) \right|^{2} n^{-1} \sum_{i \in I} \left| \psi_{n}^{2}(Z_{i})\phi_{\theta_{j_{2}}}(Z_{i},\theta_{0}) - \psi_{n}(Z_{i})\phi_{\theta_{j_{1}}}(Z_{i},\theta_{0}) \right|^{2} n^{-1} \sum_{i \in I} \left| \psi_{n}^{2}(Z_{i})\phi_{\theta_{j_{2}}}(Z_{i},\theta_{0}) - \psi_{n}(Z_{i})\phi_{\theta_{j_{1}}}(Z_{i},\theta_{0}) \right|^{2} n^{-1} \sum_{i \in I} \left| \psi_{n}^{2}(Z_{i})\phi_{\theta_{j_{2}}}(Z_{i},\theta_{0}) - \psi_{n}(Z_{i})\phi_{\theta_{j_{1}}}(Z_{i},\theta_{0}) \right|^{2} n^{-1} \sum_{i \in I} \left| \psi_{n}^{2}(Z_{i},\theta_{0}) - \psi_{n}(Z_{i})\phi_{\theta_{j_{1}}}(Z_{i},\theta_{0}) \right|^{2} n^{-1} \sum_{i \in I} \left| \psi_{n}^{2}(Z_{i},\theta_{0}) - \psi_{n}(Z_{i})\phi_{\theta_{j_{1}}}(Z_{i},\theta_{0}) \right|^{2} n^{-1} \sum_{i \in I} \left| \psi_{n}^{2}(Z_{i},\theta_{0}) - \psi_{n}(Z_{i})\phi_{\theta_{j_{1}}}(Z_{i},\theta_{0}) \right|^{2} n^{-1} \sum_{i \in I} \left| \psi_{n}^{2}(Z_{i},\theta_{0}) - \psi_{n}(Z_{i},\theta_{0}) \right|^{2} n^{-1} \sum_{i \in I} \left| \psi_{n}^{2}(Z_{i},\theta_{0}) - \psi_{n}(Z_{i},\theta_{0}) \right|^{2} n^{-1} \sum_{i \in I} \left| \psi_{n}^{2}(Z_{i},\theta_{0}) - \psi_{n}(Z_{i},\theta_{0}) \right|^{2} n^{-1} \sum_{i \in I} \left| \psi_{n}^{2}(Z_{i},\theta_{0}) \right|^{2} n^{-1} \sum_{i$$

where the equality is by Assumption 3(vi), (66) and (67). Combining the results in (69) and (70), we get

$$n^{-1} \sum_{i \in I} (\phi_{w\theta_j, k_1, n}(Z_i, \theta_0) - w_n(Z_i)\phi_{\theta_{j_1}}(Z_i, \theta_0))\phi_{w\theta_{j_2}, k_1, n}(Z_i, \theta_0) = o_p(1).$$
(71)

Similarly, we can show that

$$n^{-1} \sum_{i \in I} \left(w_n(Z_i) \phi_{\theta_{j_1}}(Z_i, \theta_0) (\phi_{w\theta_{j_2}, k_1, n}(Z_i, \theta_0) - w_n(Z_i) \phi_{\theta_{j_2}}(Z_i, \theta_0)) = o_p(1). \right)$$
(72)

Collecting the results in (65), (66), (71) and (72), we have

$$n^{-1}\phi_{w\theta_{j_1},k_1,n}M_{k_1,n}\phi'_{w\theta_{j_2},k_1,n} = E\left[w_n^2(Z_i)\phi_{\theta_{j_1}}(Z_i,\theta_0)\phi_{\theta_{j_2}}(Z_i,\theta_0)\right] \not + o_p(1)$$
(73)

for any $j_1 = 1, \ldots, d_{\theta}$ and any $j_2 = 1, \ldots, d_{\theta}$, which implies that

$$n^{-1}\phi_{w\theta,k_{1},n}M_{k_{1},n}\phi_{w\theta,k_{1},n}' = E\left[w_{n}^{2}(Z_{i})\phi_{\theta}(Z_{i},\theta_{0})\phi_{\theta}'(Z_{i},\theta_{0})\right] \notin o_{p}(1).$$
(74)

By the Cauchy-Schwarz inequality, for any $j_1 = 1, \ldots, d_{\theta}$ and any $j_2 = 1, \ldots, d_{\theta}$,

$$\frac{\left|\frac{(\phi_{w\theta_{j_1},n} - \phi_{w\theta_{j_1},k_{1,n}})M_{k_1,n}\phi'_{w\theta_{j_2},k_{1,n}}}{n}\right|^2}{\left(\frac{(\phi_{w\theta_{j_1},n} - \phi_{w\theta_{j_1},k_{1,n}})M_{k_{1,n}}(\phi_{w\theta_{j_1},n})}{n}\right)^2}{n}\frac{\phi_{w\theta_{j_2},k_{1,n}}M_{k_{1,n}}\phi'_{w\theta_{j_2},k_{1,n}}}{n} = o_p(1) \quad (75)$$

where the equality is by (63), (67) and (73). (75) then implies that

$$n^{-1} \left(\phi_{w\theta,n} - \phi_{w\theta,k_1,n} \right) M_{k_1,n} \phi'_{w\theta,k_1,n} = o_p(1)$$
(76)

and similarly

$$n^{-1}\phi_{w\theta,k_1,n}M_{k_1,n}(\phi_{w\theta,n} - \phi_{w\theta,k_1,n})' = o_p(1).$$
(77)

Combining the results in (62), (64), (74), (76) and (77), we immediately get the claimed result. \Box

Proof. [Proof of Theorem 2] By the definition of $\hat{\theta}_n$, we have the following first order condition

$$n^{-1} \sum_{i \in I} \oint_{i \in I} (h_{n_1}(Z_i) - \hat{\phi}_{n_2}(Z_i, \hat{\theta}_n)) \hat{\phi}_{\theta, n_2}(Z_i, \hat{\theta}_n) = 0.$$
(78)

Applying the first order expansion to (78), we get

$$0 = n^{-1} \sum_{i \in I} \left(\hat{w}_n(Z_i)(\hat{h}_{n_1}(Z_i) - \hat{\phi}_{n_2}(Z_i, \theta_0)) \hat{\phi}_{\theta, n_2}(Z_i, \theta_0) + n^{-1} \sum_{i \in I} \left(\hat{w}_n(Z_i) \hat{\phi}_{\theta, n_2}(Z_i, \tilde{\theta}_n) \hat{\phi}_{\theta, n_2}(Z_i, \tilde{\theta}_n)'(\hat{\theta}_n - \theta_0) + n^{-1} \sum_{i \in I} \left(\hat{w}_n(Z_i)(\hat{h}_{n_1}(Z_i) - \hat{\phi}_{n_2}(Z_i, \tilde{\theta}_n)) \hat{\phi}_{\theta, n_2}(Z_i, \tilde{\theta}_n)(\hat{\theta}_n - \theta_0), \right) \right)$$

$$(79)$$

where $\tilde{\theta}_n$ is between $\hat{\theta}_n$ and θ_0 and it may differ across rows.

For any $j = 1, \ldots, d_{\theta}$, by the mean value expansion and the Cauchy-Schwarz inequality,

$$\left| \widehat{\phi}_{\theta_{j},n_{2}}(Z_{i},\widetilde{\theta}_{j,n}) - \widehat{\phi}_{\theta_{j},n_{2}}(Z_{i},\theta_{0}) \right| \leq \sup_{\theta \in \mathcal{N}_{\delta_{n}}} \left\| \widehat{\phi}_{\theta_{j}\theta,n_{2}}(Z_{i},\theta) \right\| \left\| \widetilde{\theta}_{j,n} - \theta_{0} \right\| \left(\sum_{i=1}^{n} \frac{1}{i} \sum_{i=1}^{n} \frac{1}{i} \sum_{j=1}^{n} \frac{1}{i} \sum_{i=1}^{n} \frac{1}{i} \sum_{j=1}^{n} \frac{1}{i} \sum$$

which together with the triangle inequality and Lemma 4 implies that

$$n^{-1} \sum_{i \in I} \left(\widehat{\phi}_{\theta_j, n_2}(Z_i, \widetilde{\theta}_{j, n}) - \widehat{\phi}_{\theta_j, n_2}(Z_i, \theta_0) \right)^2 \leq \sup_{\theta \in \mathcal{N}_{\delta_n}} n^{-1} \sum_{i \in I} \left\| \widehat{\psi}_{\theta_j \theta, n_2}(Z_i, \theta) \right\|^2 \left\| \widetilde{\theta}_{j, n} - \theta_0 \right\|_{\ell}^2 = o_p(1).$$

$$\tag{80}$$

By Assumption 3(iii) and (80),

$$n^{-1} \sum_{i \in I} (\widehat{\phi}_{\theta_{j}, n_{2}}(Z_{i}, \widetilde{\theta}_{j, n}) - \phi_{\theta_{j}}(Z_{i}, \theta_{0}))^{2} \leq 2n^{-1} \sum_{i \in I} (\widehat{\phi}_{\theta_{j}, n_{2}}(Z_{i}, \widetilde{\theta}_{j, n}) - \widehat{\phi}_{\theta_{j}, n_{2}}(Z_{i}, \theta_{0}))^{2} + 2n^{-1} \sum_{i \in I} (\widehat{\phi}_{\theta_{j}, n_{2}}(Z_{i}, \theta_{0}) - \phi_{\theta_{j}}(Z_{i}, \theta_{0}))^{2} = o_{p}(1).$$
(81)

By Assumption 3(iv) and the Markov inequality,

$$n^{-1} \sum_{i \in I} (\phi_{\theta_j}(Z_i, \theta_0))^2 = O_p(1), \tag{82}$$

which together with (81) implies that

$$n^{-1} \sum_{i \in I} (\widehat{\phi}_{\theta_j, n_2}(Z_i, \widetilde{\theta}_n))^2 = O_p(1)$$
(83)

for any $j = 1, \ldots, d_{\theta}$. For any $j_1 = 1, \ldots, d_{\theta}$ and any $j_2 = 1, \ldots, d_{\theta}$, we can use the triangle inequality and

the Cauchy-Schwarz inequality, Assumptions 2(v), (81), (82) and (83) to deduce that

$$\begin{vmatrix}
n^{-1} \sum_{i \in I} \left(\hat{w}_{n}(Z_{i}) \hat{\phi}_{\theta_{j_{1}},n_{2}}(Z_{i}, \tilde{\theta}_{j_{1},n}) \hat{\phi}_{\theta_{j_{2}},n_{2}}(Z_{i}, \tilde{\theta}_{j_{2},n}) - n^{-1} \sum_{i \in I} w_{n}(Z_{i}) \phi_{\theta_{j_{1}}}(Z_{i}, \theta_{0}) \phi_{\theta_{j_{2}}}(Z_{i}, \theta_{0}) \right| \\
\leq \begin{vmatrix}
n^{-1} \sum_{i \in I} \left(\hat{w}_{n}(Z_{i}) - w_{n}(Z_{i}) \right) \hat{\phi}_{\theta_{j_{1}},n_{2}}(Z_{i}, \tilde{\theta}_{j_{1},n}) \hat{\phi}_{\theta_{j_{2}},n_{2}}(Z_{i}, \tilde{\theta}_{j_{2},n}) \right| \\
+ \left| n^{-1} \sum_{i \in I} \left(w_{n}(Z_{i}) (\hat{\phi}_{\theta_{j_{1}},n_{2}}(Z_{i}, \tilde{\theta}_{j_{1},n}) - \phi_{\theta_{j_{1}}}(Z_{i}, \theta_{0})) \hat{\phi}_{\theta_{j_{2}},n_{2}}(Z_{i}, \theta_{0}) \right| \\
+ \left| n^{-1} \sum_{i \in I} \left(w_{n}(Z_{i}) \phi_{\theta_{j_{1}}}(Z_{i}, \theta_{0}) (\hat{\phi}_{\theta_{j_{2}},n_{2}}(Z_{i}, \tilde{\theta}_{j_{2},n}) - \phi_{\theta_{j_{2}}}(Z_{i}, \theta_{0})) \right| \\
= \left(n^{-1} \sum_{i \in I} \left(w_{n}(Z_{i}) \phi_{\theta_{j_{1}}}(Z_{i}, \theta_{0}) (\hat{\phi}_{\theta_{j_{2}},n_{2}}(Z_{i}, \tilde{\theta}_{j_{2},n}) - \phi_{\theta_{j_{2}}}(Z_{i}, \theta_{0})) \right| \\
= \left(n^{-1} \sum_{i \in I} \left(w_{n}(Z_{i}) \phi_{\theta_{j_{1}}}(Z_{i}, \theta_{0}) (\hat{\phi}_{\theta_{j_{2}},n_{2}}(Z_{i}, \tilde{\theta}_{j_{2},n}) - \phi_{\theta_{j_{2}}}(Z_{i}, \theta_{0})) \right| \\
= \left(n^{-1} \sum_{i \in I} \left(w_{n}(Z_{i}) \phi_{\theta_{j_{1}}}(Z_{i}, \theta_{0}) (\hat{\phi}_{\theta_{j_{2}},n_{2}}(Z_{i}, \tilde{\theta}_{j_{2},n}) - \phi_{\theta_{j_{2}}}(Z_{i}, \theta_{0})) \right|_{i \in I} \right)$$

$$(84)$$

Under Assumptions 1(i), 2(v) and 3(iv), we can the Markov inequality to deduce that

$$n^{-1} \sum_{i \in I} \left(w_n(Z_i) \phi_{\theta_{j_1}}(Z_i, \theta_0) \phi_{\theta_{j_2}}(Z_i, \theta_0) = E \left[w_n(Z_i) \phi_{\theta_{j_1}}(Z_i, \theta_0) \phi_{\theta_{j_2}}(Z_i, \theta_0) \right] + o_p(1)$$
(85)

for any $j_1 = 1, \ldots, d_{\theta}$ and any $j_2 = 1, \ldots, d_{\theta}$. Collecting the results in (84) and (85), we get

$$n^{-1} \sum_{i \in I} \widehat{\psi}_n(Z_i) \widehat{\phi}_{\theta, n_2}(Z_i, \widetilde{\theta}_n) \widehat{\phi}_{\theta, n_2}(Z_i, \widetilde{\theta}_n)' = H_0 + o_p(1).$$
(86)

By the second order Taylor expansion and the triangle inequality and the Cauchy-Schwarz inequality,

$$n^{-1} \sum_{i \in I} \left| \phi(Z_i, \widetilde{\theta}_{j,n}) - \phi(Z_i, \theta_0) \right|^2$$

$$\leq 2n^{-1} \sum_{i \in I} \left| \left(\phi_{\theta}(Z_i, \theta_0) \|^2 || \widetilde{\theta}_{j,n} - \theta_0 ||^2 + 2^{-1} \sup_{\theta \in \mathcal{N}_{\delta_n}} n^{-1} \sum_{i \in I} \left| \left(\phi_{\theta\theta}(Z_i, \theta) \|^2 || \widetilde{\theta}_{j,n} - \theta_0 ||^4 = o_p(1) \right) \right|^2$$

$$(87)$$

where the equality is by (82), Lemma 4 and $||\tilde{\theta}_{j,n} - \theta_0|| = o_p(1)$ for any $j = 1, \ldots, d_{\theta}$. (87) together with Assumptions 2(ii) then implies that

$$n^{-1} \sum_{i \in I} \left| \widehat{\phi}_{n_2}(Z_i, \widetilde{\theta}_{j,n}) - \phi(Z_i, \theta_0) \right|^2 \leq 2n^{-1} \sum_{i \in I} \left| \widehat{\phi}_{n_2}(Z_i, \widetilde{\theta}_{j,n}) - \phi(Z_i, \widetilde{\theta}_n) \right|^2 + 2n^{-1} \sum_{i \in I} \left| \phi(Z_i, \widetilde{\theta}_{j,n}) - \phi(Z_i, \theta_0) \right|^2 = o_p(1)$$

$$(88)$$

for any $j = 1, \ldots, d_{\theta}$. By the Cauchy-Schwarz inequality,

$$\left\| n^{-1} \sum_{i \in I} \left\{ \hat{w}_{n}(Z_{i})(\hat{h}_{n_{1}}(Z_{i}) - \hat{\phi}_{n_{2}}(Z_{i}, \widetilde{\theta}_{n})) \hat{\phi}_{\theta\theta, n_{2}}(Z_{i}, \widetilde{\theta}_{n}) \right\|_{z}^{2} \left\{ \sup_{i \in I} |\hat{w}_{n}(z)| \sqrt{\max_{j=1,...,d_{\theta}} n^{-1} \sum_{i \in I} (\hat{h}_{n_{1}}(Z_{i}) - \hat{\phi}_{n_{2}}(Z_{i}, \widetilde{\theta}_{j,n}))^{2}} \sqrt{\max_{j=1,...,d_{\theta}} n^{-1} \sum_{i \in I} \left\| \hat{\phi}_{\theta\theta, n_{2}}(Z_{i}, \widetilde{\theta}_{j,n}) \right\|_{z}^{2}} \right\} \\
\leq 2 \sup_{z} |\hat{w}_{n}(z)| \sqrt{\max_{j=1,...,d_{\theta}} n^{-1} \sum_{i \in I} \left\| \hat{\phi}_{\theta\theta, n_{2}}(Z_{i}, \widetilde{\theta}_{n}) \right\|_{z}^{2}} \left\{ \sqrt{n^{-1} \sum_{i \in I} \left| \hat{h}_{n_{1}}(Z_{i}) - h_{0}(Z_{i}) \right|_{z}^{2}} + \max_{j=1,...,d_{\theta}} n^{-1} \sum_{i \in I} \left| \hat{\phi}_{n_{2}}(Z_{i}, \widetilde{\theta}_{j,n}) - \phi(Z_{i}, \theta_{0}) \right|_{z}^{2}} = o_{p}(1) \quad (89)$$

where the last equality is by (44), (45), Lemma 4 and (88).

By definition,

$$n^{-1} \sum_{i \in I} \widehat{\psi}_{n}(Z_{i})(\widehat{h}_{n_{1}}(Z_{i}) - \widehat{\phi}_{n_{2}}(Z_{i},\theta_{0}))\widehat{\phi}_{\theta,n_{2}}(Z_{i},\theta_{0})$$

$$= n^{-1} \sum_{i \in I} \widehat{\psi}_{n}(Z_{i})(\widehat{h}_{n_{1}}(Z_{i}) - \widehat{\phi}_{n_{2}}(Z_{i},\theta_{0}))\phi_{\theta}(Z_{i},\theta_{0})$$

$$+ n^{-1} \sum_{i \in I} \widehat{\psi}_{n}(Z_{i}) - w_{n}(Z_{i}))(\widehat{h}_{n_{1}}(Z_{i}) - \widehat{\phi}_{n_{2}}(Z_{i},\theta_{0}))\widehat{\phi}_{\theta,n_{2}}(Z_{i},\theta_{0})$$

$$+ n^{-1} \sum_{i \in I} \widehat{\psi}_{n}(Z_{i})(\widehat{h}_{n_{1}}(Z_{i}) - \widehat{\phi}_{n_{2}}(Z_{i},\theta_{0}))(\widehat{\phi}_{\theta,n_{2}}(Z_{i},\theta_{0}) - \phi_{\theta}(Z_{i},\theta_{0})).$$
(90)

By Assumptions 3(iv) and 3(v), and the Markov inequality

$$n^{-1} \sum_{i \in I} \left\| \widehat{\phi}_{\theta, n_2}(Z_i, \theta_0) \|^2 \le 2n^{-1} \sum_{i \in I} \| \widehat{\phi}_{\theta, n_2}(Z_i, \theta_0) - \phi_{\theta}(Z_i, \theta_0) \|^2 + 2n^{-1} \sum_{i \in I} \| \phi_{\theta}(Z_i, \theta_0) \|^2 = O_p(1).$$
(91)

By the triangle inequality and the Cauchy-Schwarz inequality, (91), Lemma 5, Assumptions 2(v) and 3(vii),

$$\left\| n^{-1} \sum_{i \in I} \left(\widehat{w}_{n}(Z_{i}) - w_{n}(Z_{i}) \right) \left(\widehat{h}_{n_{1}}(Z_{i}) - \widehat{\phi}_{n_{2}}(Z_{i},\theta_{0}) \right) \widehat{\phi}_{\theta,n_{2}}(Z_{i},\theta_{0}) \right\|$$

$$\leq \sup_{z \in \mathcal{Z}} \left| \widehat{w}_{n}(z) - w_{n}(z) \right| \sqrt{n^{-1} \sum_{i \in I} \left(\widehat{h}_{n_{1}}(Z_{i}) - \widehat{\phi}_{n_{2}}(Z_{i},\theta_{0}) \right)^{2}} \sqrt{n^{-1} \sum_{i \in I} \left\| \widehat{\phi}_{\theta,n_{2}}(Z_{i},\theta_{0}) \right\|^{2}}$$

$$= o_{p}(n_{1}^{-1/2} + n_{2}^{-1/2}) \sqrt{n^{-1} \sum_{i \in I} \left\| \widehat{\phi}_{\theta,n_{2}}(Z_{i},\theta_{0}) \right\|^{2}} = o_{p}(n_{1}^{-1/2} + n_{2}^{-1/2}).$$

$$(92)$$

By the triangle inequality and the Cauchy-Schwarz inequality, Lemma 5, Assumptions 2(v), 3(iii) and 3(vii),

$$\left\| n^{-1} \sum_{i \in I} \left(w_n(Z_i)(\hat{h}_{n_1}(Z_i) - \hat{\phi}_{n_2}(Z_i, \theta_0))(\hat{\phi}_{\theta, n_2}(Z_i, \theta_0) - \phi_{\theta}(Z_i, \theta_0)) \right\| \\ \leq \left(\sup_{z \in \mathcal{Z}} |w_n(z)| \sqrt{n^{-1} \sum_{i \in I} (\hat{h}_{n_1}(Z_i) - \hat{\phi}_{n_2}(Z_i, \theta_0))^2} \sqrt{n^{-1} \sum_{i \in I} \left\| \hat{\phi}_{\theta, n_2}(Z_i, \theta_0) - \phi_{\theta}(Z_i, \theta_0) \right\|^2} \right) \\ = o_p(n_1^{-1/2} + n_2^{-1/2}).$$

$$(93)$$

Combining the results in (90), (92) and (93), we get

$$n^{-1} \sum_{i \in I} \left(\hat{b}_{n}(Z_{i})(\hat{h}_{n_{1}}(Z_{i}) - \hat{\phi}_{n_{2}}(Z_{i}, \theta_{0})) \hat{\phi}_{\theta, n_{2}}(Z_{i}, \theta_{0}) \right)$$

$$= n^{-1} \sum_{i \in I} \left(b_{n}(Z_{i})(\hat{h}_{n_{1}}(Z_{i}) - \hat{\phi}_{n_{2}}(Z_{i}, \theta_{0})) \phi_{\theta}(Z_{i}, \theta_{0}) + o_{p}(n_{1}^{-1/2} + n_{2}^{-1/2}) \right)$$

$$= n^{-1} \sum_{i \in I} \left(b_{n}(Z_{i})(\hat{h}_{n_{1}}(Z_{i}) - h_{0}(Z_{i})) \phi_{\theta}(Z_{i}, \theta_{0}) - n^{-1} \sum_{i \in I} \left(b_{n}(Z_{i})(\hat{\phi}_{n_{2}}(Z_{i}, \theta_{0}) - h_{0}(Z_{i})) \phi_{\theta}(Z_{i}, \theta_{0}) + o_{p}(n_{1}^{-1/2} + n_{2}^{-1/2}) \right) \right)$$
(94)

By the definition of $\hat{h}_{n_1}(Z_i)$, we can write

$$n^{-1} \sum_{i \in I} \left(\psi_n(Z_i)(\hat{h}_{n_1}(Z_i) - h_0(Z_i))\phi_\theta(Z_i, \theta_0) \right) \\ = \frac{\phi_{w\theta,n} P_{n,k_1}(P'_{n_1,k_1} P_{n_1,k_1})^{-1} P'_{n_1,k_1} U_{n_1}}{n} \\ + \frac{\phi_{w\theta,n} P_{n,k_1}(P'_{n_1,k_1} P_{n_1,k_1})^{-1} P'_{n_1,k_1} (H_{n_1} - H_{n_1,k_1})}{n} + \frac{\phi_{w\theta,n}(H_n - H_{n,k_1})}{n}.$$
(95)

where $H_n = (h_0(Z_i))'_{i \in I}, H_{n_1} = (h_0(Z_i))'_{i \in I_1}, U_{n_1} = (u_i)'_{i \in I_1}, H_{n,k_1} = (h_{0,k_1}(Z_i))'_{i \in I}, H_{n_1,k_1} = (h_{0,k_1}(Z_i))'_{i \in I_1}$ and $\phi_{w\theta,n} = (w_n(Z_i)\phi_\theta(Z_i,\theta_0))_{i \in I}$. By the Cauchy-Schwarz inequality,

$$\frac{\left|\phi_{w\theta,n}P_{n,k_{1}}(P_{n_{1},k_{1}}'P_{n_{1},k_{1}})^{-1}P_{n_{1},k_{1}}'(H_{n_{1}}-H_{n_{1},k_{1}})\right|^{2}}{n^{2}} \left(\sum_{\substack{\varphi_{w\theta,n}P_{n,k_{1}}P_{n,k_{1}}\phi_{w\theta,n}'(H_{n_{1}}-H_{n_{1},k_{1}})'P_{n_{1},k_{1}}(P_{n_{1},k_{1}}'P_{n_{1},k_{1}})^{-2}P_{n_{1},k_{1}}'(H_{n_{1}}-H_{n_{1},k_{1}})} \right) \\ \leq \frac{\lambda_{\max}(Q_{n,k_{1}})}{\lambda_{\min}(Q_{n_{1},k_{1}})} \frac{\phi_{w\theta,n}P_{n,k_{1}}(P_{n,k_{1}}'P_{n,k_{1}})^{-1}P_{n,k_{1}}'\phi_{w\theta,n}}{n}}{n} \\ \times \frac{(H_{n_{1}}-H_{n_{1},k_{1}})'P_{n_{1},k_{1}}(P_{n_{1},k_{1}}'P_{n_{1},k_{1}})^{-1}P_{n_{1},k_{1}}'(H_{n_{1}}-H_{n_{1},k_{1}})}{n_{1}}}{\lambda_{\min}(Q_{k_{1},n_{1}})} n^{-1}\sum_{i\in I} \left|\phi_{\theta}(Z_{i},\theta_{0})\|^{2} \times n_{1}^{-1}\sum_{i\in I_{1}}|h_{0,k_{1}}(Z_{i})-h_{0}(Z_{i})|^{2} = O_{p}(k_{1}^{-2r_{h}}), \quad (96)$$

where the last equality is by (42), (82), Assumptions 1(iv) and 2(v). By the triangle inequality,

$$\left\| n^{-1} \sum_{i \in I} \left(w_n(Z_i)(h_{0,k_1}(Z_i) - h_0(Z_i))\phi_{\theta}(Z_i, \theta_0) \right) \right\|$$

$$\leq \left(\sup_{z \in \mathcal{Z}} |w_n(z)| n^{-1} \sum_{i \in I} \left((h_{0,k_1}(Z_i) - h_0(Z_i))\phi_{\theta}(Z_i, \theta_0) \right) \right)$$

$$\leq C \frac{\sup_z |h_0(z) - h_{k_1}(z)|}{n} \sum_{i \in I} \left(\phi_{\theta}(Z_i, \theta_0) \right) = O_p(k_1^{-r_h}),$$

$$(97)$$

where the last equality is by (82), Assumptions 1(iv) and 2(v). Combining the results in (95), (96) and (97),

we get

$$n^{-1} \sum_{i \in I} \left(\psi_n(Z_i)(\hat{h}_{n_1}(Z_i) - h_0(Z_i)) \phi_\theta(Z_i, \theta_0) \right)$$

= $\frac{\phi_{w\theta,n} P_{n,k_1}(P'_{n_1,k_1} P_{n_1,k_1})^{-1}}{n} \sum_{i \in I_1} u_i P_{k_1}(Z_i) + O_p(k_1^{-r_h}).$ (98)

By the definition of $\widehat{\phi}_{n_2}(Z_i, \theta_0)$, we can write

$$n^{-1} \sum_{i \in I} \left(\psi_n(Z_i)(\widehat{\phi}_{n_2}(Z_i, \theta_0) - h_0(Z_i))\phi_\theta(Z_i, \theta_0) \right)$$

= $\frac{\phi_{w\theta,n} P_{n,k_2}(P'_{n_2,k_2} P_{n_2,k_2})^{-1}}{n} \sum_{i \in I_2} \varepsilon_i P_{k_2}(Z_i)$
+ $\frac{\phi_{w\theta,n} P_{n,k_2}(P'_{n_2,k_2} P_{n_2,k_2})^{-1}(H_{n_2} - H_{n_2,k_2})}{n} + \frac{\phi_{w\theta,n}(H_n - H_{n,k_2})}{n}$ (99)

where $H_{n_2,k_2} = (h_{0,k_2}(Z_i))'_{i \in I_2}$ and $H_{n,k_2} = (h_{0,k_2}(Z_i))_{i \in I}$. Using similar arguments in showing (96) and (97), we get

$$\frac{\phi_{w\theta,n}P_{n,k_2}(P'_{n_2,k_2}P_{n_2,k_2})^{-1}(H_{n_2}-H_{n_2,k_2})}{n} = O_p(k_2^{-r_h}) \text{ and } \frac{\phi_{w\theta,n}(H_n-H_{n,k_2})}{n} = O_p(k_2^{-r_h}),$$

which together with (99) implies that

$$n^{-1} \sum_{i \in I} \left(w_n(Z_i)(\widehat{\phi}_{n_2}(Z_i, \theta_0) - h_0(Z_i))\phi_\theta(Z_i, \theta_0) - \frac{\phi_{w\theta,n}P_{n,k_2}(P'_{n_2,k_2}P_{n_2,k_2})^{-1}}{n} \sum_{i \in I_2} \varepsilon_i P_{k_2}(Z_i) + O_p(k_2^{-r_h}). \right)$$
(100)

By Assumption 3(v), $C^{-1}Q_{n_1,k_1} \leq n_1 \sum_{i \notin I_1} \sigma_u^2(Z_i) P_{k_1}(Z_i) P_{k_1}(Z_i)' \leq CQ_{n_1,k_1}$, which together with (42) implies that

$$C^{-1} < \lambda_{\min}(Q_{n_1,u}) \le \lambda_{\max}(Q_{n_1,u}) < C,$$

$$(101)$$

with probability approaching 1. Similarly, we can show that

$$C^{-1} \le \lambda_{\min}(Q_{n_2,\varepsilon}) \le \lambda_{\max}(Q_{n_2,\varepsilon}) \le C$$
(102)

with probability approaching 1.

Under the i.i.d. assumption,

$$E\left[\left|\frac{\phi_{w\theta,n}P_{n,k_{1}}(P_{n_{1,k_{1}}}^{\prime}P_{n_{1,k_{1}}})^{-1}}{n}\sum_{i\in I_{1}}\left(\mu_{i}P_{k_{1}}(Z_{i})\right)^{2}\left\{Z_{i}\}_{i\in I}\right]\right]\left(\frac{\phi_{w\theta,n}P_{n,k_{1}}(P_{n_{1,k_{1}}}^{\prime}P_{n_{1,k_{1}}})^{-1}Q_{n_{1,u}}(P_{n_{1,k_{1}}}^{\prime}P_{n_{1,k_{1}}})^{-1}P_{n,k_{1}}^{\prime}\phi_{w\theta,n}}{n^{2}n_{1}^{-1}}\right]$$

$$\leq\frac{C\lambda_{\max}(Q_{n_{1,u}})}{\lambda_{\min}^{2}(Q_{k_{1,n_{1}}})}\frac{\phi_{w\theta,n}P_{n,k_{1}}P_{n,k_{1}}^{\prime}\phi_{w\theta,n}}{n^{2}n_{1}}$$

$$\leq\frac{\lambda_{\max}(Q_{n_{1,u}})\lambda_{\max}(Q_{n,k_{1}})}{n_{1}\lambda_{\min}^{2}(Q_{n_{1,k_{1}}})}\frac{\phi_{w\theta,n}P_{n,k_{1}}(P_{n,k_{1}}^{\prime}P_{n,k_{1}})^{-1}P_{n,k_{1}}^{\prime}\phi_{w\theta,n}}{n}$$

$$\leq\sup_{z\in\mathcal{Z}}\left|w_{n}^{2}(z)\right|\frac{\lambda_{\max}(Q_{n_{1,u}})\lambda_{\max}(Q_{n,k_{1}})}{n_{1}\lambda_{\min}^{2}(Q_{n_{1,k_{1}}})}n^{-1}\sum_{i\in I}\left\|\phi_{\theta}(Z_{i},\theta_{0})\right\|^{2}=O_{p}(n_{1}^{-1}),$$
(103)

where the last equality is by (101), (42), (82), Assumptions 2(v) and 3(iv). Combined with the Markov inequality, (103) implies that

$$\frac{\phi_{w\theta,n}P_{n,k_1}(P'_{n_1,k_1}P_{n_1,k_1})^{-1}}{n} \sum_{i \in I_1} \left(\mu_i P_{k_1}(Z_i) = O_p(n_1^{-1/2}). \right)$$
(104)

Similarly, we can show that

$$\frac{\phi_{w\theta,n}P_{n,k_2}(P'_{n_2,k_2}P_{n_2,k_2})^{-1}}{n} \sum_{i \in I_2} \oint_{i} P_{k_2}(Z_i) = O_p(n_2^{-1/2}).$$
(105)

By (86) and (89),

$$n^{-1}\sum_{i\in I} \oint_{i\in I} (\widehat{\psi}_n(Z_i)(\widehat{\phi}_{\theta,n_2}(Z_i,\widetilde{\theta}_n))\widehat{\phi}_{\theta,n_2}(Z_i,\widetilde{\theta}_n) + (\widehat{h}_{n_1}(Z_i) - \widehat{\phi}_{n_2}(Z_i,\widetilde{\theta}_n))\widehat{\phi}_{\theta\theta,n_2}(Z_i,\widetilde{\theta}_n)) = H_{0,n} + o_p(1), \quad (106)$$

which together with (79) implies that

$$[H_{0,n} + o_p(1)](\widehat{\theta}_n - \theta_0) = -n^{-1} \sum_{i \in I} \oint_{i \in I} \widehat{\psi}_n(Z_i)(\widehat{h}_{n_1}(Z_i) - \widehat{\phi}_{n_2}(Z_i, \theta_0))\widehat{\phi}_{\theta, n_2}(Z_i, \theta_0).$$
(107)

By (94), (98) and (100), and Assumption 3(vii),

$$n^{-1} \sum_{i \in I} \left(\hat{w}_{n}(Z_{i})(\hat{h}_{n_{1}}(Z_{i}) - \hat{\phi}_{n_{2}}(Z_{i}, \theta_{0})) \hat{\phi}_{\theta, n_{2}}(Z_{i}, \theta_{0}) \right)$$

$$= \frac{\phi_{w\theta, n} P_{n, k_{1}}(P'_{n_{1}, k_{1}} P_{n_{1}, k_{1}})^{-1}}{n} \sum_{i \in I_{1}} \left(\mu_{k_{1}}(Z_{i}) - \frac{\phi_{w\theta, n} P_{n, k_{2}}(P'_{n_{2}, k_{2}} P_{n_{2}, k_{2}})^{-1}}{n} \sum_{i \in I_{2}} \left(i P_{k_{2}}(Z_{i}) + o_{p}(n_{1}^{-1/2} + n_{2}^{-1/2}) \right)$$
(108)

which together with (104) and (105) implies that

$$\frac{1}{n}\sum_{i\in I} \left[\widehat{h}_{n_1}(Z_i) - \widehat{\phi}_{n_2}(Z_i, \theta_0)\right] \widehat{\phi}_{\theta, n_2}(Z_i, \theta_0) = O_p(n_1^{-1/2} + n_2^{-1/2}).$$
(109)

Using (107), (108) and (109), and then applying Assumption 3(ii), we get

$$(\widehat{\theta}_n - \theta_0) = \frac{-H_0^{-1} \phi_{w\theta,n} P_{n,k_1} Q_{n_1,k_1}^{-1}}{nn_1} \sum_{i \in I_1} u_i P_{k_1}(Z_i) + \frac{H_0^{-1} \phi_{w\theta,n} P_{n,k_2} Q_{n_2,k_2}^{-1}}{nn_2} \sum_{i \in I_2} \varepsilon_i P_{k_2}(Z_i) + o_p (n_1^{-1/2} + n_2^{-1/2}),$$
(110)

which together with Assumption 3(ii), (104) and (105) implies that $\hat{\theta}_n - \theta_0 = O_p(n_1^{-1/2} + n_2^{-1/2}).$

By Assumption 3(v), $Q_{n_1,u} \ge C^{-1}Q_{n_1,k_1}$ which implies that

$$n_{1}\Sigma_{n_{1}} = \frac{\phi_{w\theta,n}P_{n,k_{1}}Q_{n_{1},k_{1}}^{-1}Q_{n_{1},k_{1}}P_{n,k_{1}}'\phi_{w\theta,n}}{n^{2}}$$

$$\geq \frac{\phi_{w\theta,n}P_{n,k_{1}}Q_{n_{1},k_{1}}^{-1}P_{n,k_{1}}'\phi_{w\theta,n}}{Cn^{2}n_{1}^{-1}}$$

$$\geq \frac{\lambda_{\min}(Q_{n,k_{1}})}{C\lambda_{\max}(Q_{n,k_{1}})}\frac{\phi_{w\theta,n}P_{n,k_{1}}(P_{n,k_{1}}'P_{n,k_{1}})^{-1}P_{n,k_{1}}'\phi_{w\theta,n}}{n}$$

$$\geq \frac{\lambda_{\min}(Q_{n,k_{1}})}{C\lambda_{\max}(Q_{n,k_{1}})}H_{0,n} + o_{p}(1)$$
(111)

where the last equality is by (42), Assumption 2(v) and Lemma 6. Using (42), (111) and Assumption 3(ii), we have

$$\lambda_{\min}(n_1 \Sigma_{n_1}) \ge C^{-1} \tag{112}$$

with probability approaching 1. Similarly, we can show that

$$\lambda_{\min}(n_2 \Sigma_{n_2}) \ge C^{-1} \tag{113}$$

with probability approaching 1. By (112),

with probability approaching 1. Combining the results in (110), (114) and $\lambda_{\min}(H_0) > 0$ in Assumption 3(ii), we get

$$(H_{0,n}(\Sigma_{n_{1}} + \Sigma_{n_{2}})^{-1}H_{0,n})^{1/2}(\hat{\theta}_{n} - \theta_{0})$$

$$= \frac{-(H_{0,n}(\Sigma_{n_{1}} + \Sigma_{n_{2}})^{-1}H_{0,n})^{1/2}H_{0,n}^{-1}\phi_{w\theta,n}P_{n,k_{1}}Q_{n_{1},k_{1}}^{-1}}{nn_{1}}\sum_{i\in I_{1}}u_{i}P_{k_{1}}(Z_{i})$$

$$+ \frac{(H_{0,n}(\Sigma_{n_{1}} + \Sigma_{n_{2}})^{-1}H_{0,n})^{1/2}H_{0,n}^{-1}\phi_{w\theta,n}P_{n,k_{2}}Q_{n_{2},k_{2}}^{-1}}{nn_{2}}\sum_{i\in I_{2}}\varepsilon_{i}P_{k_{2}}(Z_{i}) + o_{p}(n_{1}^{-1/2} + n_{2}^{-1/2}). \quad (115)$$

Define

$$\omega_{i,n} = \begin{cases} \begin{pmatrix} -(H_{0,n}(\Sigma_{n_1} + \Sigma_{n_2})^{-1} H_{0,n})^{1/2} H_{0,n}^{-1} \phi_{w\theta,n} P_{n,k_1} Q_{n_1,k_1}^{-1} P_{k_1}(Z_i) u_i \\ \\ (H_{0,n}(\Sigma_{n_1} + \Sigma_{n_2})^{-1} H_{0,n})^{1/2} H_{0,n}^{-1} \phi_{w\theta,n} P_{n,k_2} Q_{n_2,k_2}^{-1} P_{k_2}(Z_i) \varepsilon_i \\ \\ \\ nn_2 \end{pmatrix}, & n_1 < i \le n \end{cases}$$

Then by (110), we can write

$$(H_{0,n}(\Sigma_{n_1} + \Sigma_{n_2})^{-1} H_{0,n})^{1/2} (\widehat{\theta}_n - \theta_0) = \sum_{i=1}^n (i_{i,n} + o_p (n_1^{-1/2} + n_2^{-1/2})).$$
(116)

Let $\mathcal{F}_{i,n}$ be the sigma field generated by $\{\omega_{1,n}, \ldots, \omega_{i,n}, \{Z_i\}_{i \in I}\}$ for $i = 1, \ldots, n$. Then under Assumption 1(i), $E[\gamma'_n \omega_{i,n} | \mathcal{F}_{i-1,n}] = 0$ which means that $\{\gamma'_n \omega_{i,n}\}_{i=1}^n$ is a martingale difference array. We next use the Martingale CLT to show the claim. There are two sufficient conditions to verify:

$$\sum_{i=1}^{n} E\left[\left(\gamma_{n}^{\prime}\omega_{i,n}\right)^{2}\middle|\mathcal{F}_{i,n}\right] \longleftrightarrow_{p} 1; \text{ and}$$

$$(117)$$

$$\sum_{i=1}^{n} \not E \left[\left(\gamma'_{n} \omega_{i,n} \right)^{2} I \left\{ \left| \gamma'_{n} \omega_{i,n} \right| > \varepsilon \right\} \right| \mathcal{F}_{i,n} \right] \stackrel{\mathsf{l}}{\longleftrightarrow}_{p} \ 0 \forall \varepsilon > 0.$$
(118)

For ease of notations, we define $D_n = (H_{0,n}(\Sigma_{n_1} + \Sigma_{n_2})^{-1}H_{0,n})^{1/2}H_{0,n}^{-1}$. By definition, we have

$$\sum_{i=1}^{n} E\left[\left(\gamma_{n}'\omega_{i,n}\right)^{2}\middle|\mathcal{F}_{i,n}\right] = \sum_{i=1}^{n} \gamma_{n}' E\left[\omega_{i,n}\omega_{i,n}'\middle|\mathcal{F}_{i,n}\right] \left(\sum_{n=1}^{n} \sum_{i=1}^{n} \gamma_{n}' E\left[\omega_{i,n}\omega_{i,n}'\middle|\mathcal{F}_{i,n}\right] \right) \left(\sum_{n=1}^{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{i=1}^{n} \gamma_{n}' E\left[\omega_{i,n}\omega_{i,n}'\middle|\mathcal{F}_{i,n}\right] \left(\sum_{n=1}^{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{j=1}^{n}$$

which proves (117). By the monotonicity of expectation,

$$\sum_{i=1}^{n} E\left[\left(\gamma_{n}^{\prime}\omega_{i,n}\right)^{2}I\left\{\left|\gamma_{n}^{\prime}\omega_{i,n}\right| > \varepsilon\right\}\right|\mathcal{F}_{i,n}\right]\left(\\ \leq \frac{1}{\varepsilon^{2}}\sum_{i=1}^{n} E\left[\left(\gamma_{n}^{\prime}\omega_{i,n}\right)^{4}\right|\mathcal{F}_{i,n}\right] \\ = \frac{1}{\varepsilon^{2}}\sum_{i\in I_{1}} E\left[\frac{\left|\gamma_{n}^{\prime}D_{n}\phi_{w\theta,n}P_{n,k_{1}}Q_{n_{1},k_{1}}^{-1}P_{k_{1}}(Z_{i})u_{i}\right|^{4}}{\left(\frac{n^{4}n^{4}}{\left(\frac{n^{4}n^{4}}{1}\right)^{4}}\right)}\left(\mathcal{F}_{i,n}\right) \\ + \frac{1}{\varepsilon^{2}}\sum_{i\in I_{2}} E\left[\frac{\left|\gamma_{n}^{\prime}D_{n}\phi_{w\theta,n}P_{n,k_{2}}Q_{n_{2},k_{2}}^{-1}P_{k_{2}}(Z_{i})\varepsilon_{i}\right|^{4}}{\left(\frac{\mathcal{F}_{i,n}}{1}\right)}\right].$$
(120)
A 3(iv),

By Assumptions 2(v) and 3(iv),

$$H_{0,n} = E\left[w_n(Z_i)\phi_\theta(Z_i,\theta_0)\phi'_\theta(Z_i,\theta_0)\right] \le C.$$
(121)

By (66) in the proof of Lemma 6,

$$n^{-1}\phi_{w\theta,n}\phi'_{w\theta,n} = E\left[w_n^2(Z_i)\phi_\theta(Z_i,\theta_0)\phi'_\theta(Z_i,\theta_0)\right] \notin o_p(1),$$
(122)

which together with (121), Assumptions 2(v) and 3(ii) implies that

$$C^{-1} < \lambda_{\max}(n^{-1}\phi_{w\theta,n}\phi'_{w\theta,n}) < C$$
(123)

with probability approaching 1. By (112),

$$\lambda_{\min}(n_1(\Sigma_{n_1} + \Sigma_{n_2})) \ge \lambda_{\min}(n_1\Sigma_{n_1}) > C^{-1}.$$
(124)

For any $\gamma \in \mathbb{R}^{d_{\theta}}$, we have

$$\frac{\gamma' D_n D'_n \gamma}{n_1} = \frac{\gamma' (H_{0,n} (\Sigma_{n_1} + \Sigma_{n_2})^{-1} H_{0,n})^{1/2} H_{0,n}^{-2} (H_0 (\Sigma_{n_1} + \Sigma_{n_2})^{-1} H_0)^{1/2} \gamma}{n_1} \\
\leq \frac{1}{\lambda_{\min}^2 (H_{0,n})} \frac{\gamma' H_{0,n} (\Sigma_{n_1} + \Sigma_{n_2})^{-1} H_{0,n} \gamma}{n_1} \\
\leq \frac{\gamma' H_{0,n}^2 \gamma}{\lambda_{\min}^2 (H_{0,n}) \lambda_{\min} (n_1 (\Sigma_{n_1} + \Sigma_{n_2}))} \\
\leq \frac{\lambda_{\max}^2 (H_{0,n})}{\lambda_{\min}^2 (H_{0,n}) \lambda_{\min} (n_1 (\Sigma_{n_1} + \Sigma_{n_2}))},$$
(125)

which combined with Assumption 3(ii), (121) and (124) implies that

$$\lambda_{\max}(n_1^{-1}D_n D_n') \le C \tag{126}$$

with probability approaching 1. By Assumption 4(v) and the Cauchy-Schwarz inequality,

$$\frac{1}{\varepsilon^{2}} \sum_{i \in I_{1}} E \left[\frac{\left| \gamma_{n}^{\prime} D_{n} \phi_{w\theta,n} P_{n,k_{1}} Q_{n_{1},k_{1}}^{-1} P_{k_{1}}(Z_{i}) u_{i} \right|^{4}}{n^{4} n_{1}^{4}} \right|^{\xi_{i,n}} \right] \\
\leq \frac{C}{\varepsilon^{2}} \sum_{i \in I_{1}} \frac{\left| \gamma_{n}^{\prime} D_{n} \phi_{w\theta,n} P_{n,k_{1}} Q_{n_{1},k_{1}}^{-1} P_{k_{1}}(Z_{i}) \right|^{4}}{n^{4} n_{1}^{4}} \left(\frac{\varepsilon^{2}}{\varepsilon^{2}} \sum_{i \in I_{1}} \frac{\left| \gamma_{n}^{\prime} D_{n} \phi_{w\theta,n} P_{n,k_{1}} Q_{n_{1},k_{1}}^{-1} P_{k_{1}}(Z_{i}) \right|^{2}}{\varepsilon^{2}} \left(\frac{\varepsilon^{2}}{\varepsilon^{2}} \sum_{i \in I_{1}} \frac{\left| \gamma_{n}^{\prime} D_{n} \phi_{w\theta,n} P_{n,k_{1}} Q_{n_{1},k_{1}}^{-1} P_{k_{1}}(Z_{i}) \right|^{4}}{\varepsilon^{2}} \right) \\ \leq \frac{C \xi_{k_{1}}^{2} \gamma_{n}^{\prime} D_{n} \phi_{w\theta,n} P_{n,k_{1}} Q_{n_{1},k_{1}}^{-2} P_{n,k_{1}}^{\prime} \phi_{w\theta,n} D_{n}^{\prime} \gamma_{n}}{\varepsilon^{2}} \sum_{i \in I_{1}} \frac{\left| \gamma_{n}^{\prime} D_{n} \phi_{w\theta,n} P_{n,k_{1}} Q_{n_{1},k_{1}}^{-1} P_{k_{1}}(Z_{i}) \right|^{2}}{\eta^{4} n^{4}} \left(\frac{\varepsilon^{2} \lambda_{\max}^{2}(Q_{n_{1},k_{1}})}{\varepsilon^{2} \lambda_{\min}^{2}(Q_{n_{1},k_{1}})} \frac{\xi_{k_{1}}^{2}}{n^{2} n^{4}} \frac{\gamma_{n}^{\prime} D_{n} \phi_{w\theta,n} \phi_{w\theta,n}^{\prime} D_{n}^{\prime} \gamma_{n}}{n} \sum_{i \in I_{1}} \left| \gamma_{n}^{\prime} D_{n} \phi_{w\theta,n} P_{n,k_{1}} Q_{n_{1},k_{1}}^{-1} P_{n,k_{1}}^{\prime} \phi_{w\theta,n} D_{n}^{\prime} \gamma_{n}}{n^{2} n^{2} n^{2} n^{2} n^{2} n^{4}} \right] \right) \\ \leq \frac{C \lambda_{\max}^{2}(Q_{n_{1},k_{1}})}{\varepsilon^{2} \lambda_{\min}^{2}(Q_{n_{1},k_{1}})} \frac{\xi_{k_{1}}^{2}}{n^{1}} \frac{\gamma_{n}^{\prime} D_{n} \phi_{w\theta,n} \phi_{w\theta,n}^{\prime} D_{n}^{\prime} \gamma_{n}}{nn_{1}} \frac{\gamma_{n}^{\prime} D_{n} \phi_{w\theta,n} P_{n,k_{1}} Q_{n,k_{1}}^{-1} \rho_{n,k_{1}}^{\prime} \phi_{w\theta,n} D_{n}^{\prime} \gamma_{n}}{n^{2} n^{2} n$$

where the last equality is by (42), (123), (126) and Assumptions 1(v). Similarly, we can show that

$$\frac{1}{\varepsilon^2} \sum_{i \in I_2} \not E \left[\frac{\left| \gamma'_n D_n \phi_{w\theta, n} P_{n, k_2} Q_{n_2, k_2}^{-1} P_{k_2}(Z_i) \varepsilon_i \right|^4}{\left(\begin{array}{c} M^4 n_2^4 \end{array} \right)^4} \right] \not = o_p(1), \quad (128)$$

which together with (120) and (127) proves (118). As a result, the asymptotic normality of $\hat{\theta}_n$ follows by the martingale CLT.

B Proof of the Main Results in Section 4

Lemma 7. Under Assumptions 1(i), 1(iii), 1(v), 2(v) and 3(iv)-3(vi), we have

$$n_1 \Sigma_{n_1} = E \left[u^2 w_n^2(Z) \phi_\theta(Z, \theta_0) \phi_\theta(Z, \theta_0)' \right] \notin o_p(1),$$

and $n_2 \Sigma_{n_2} = E \left[\varepsilon^2 w_n^2(Z) \phi_\theta(Z, \theta_0) \phi_\theta(Z, \theta_0)' \right] \notin o_p(1).$

Proof. [Proof of Lemma 7] For $j = 1, ..., d_{\theta}$, let $\phi_{w\theta_j,k_1,n}(z) = P_{k_1}(z)'Q_{n_1,k_1}^{-1}P'_{n,k_1}\phi'_{w\theta_j,n}$ where $\phi_{w\theta_j,n}$ denotes the *j*-th row of $\phi_{w\theta,n}$. We first show that

$$n^{-1} \sum_{i \in I} \left| \widetilde{\phi}_{w\theta_j, k_1, n}(Z_i) - w_n(Z_i) \phi_\theta(Z_i, \theta_0) \right|^2 = o_p(1)$$
(129)

for any $j = 1, \ldots, d_{\theta}$. Let $\widetilde{\beta}_{w\phi_j, k_1} = Q_{n_1, k_1}^{-1} P'_{n, k_1} \phi'_{w\theta_j, n}$. Then

$$n^{-1} \sum_{i \in I} \left| \widetilde{\phi}_{w\theta_{j},k_{1},n}(Z_{i}) - w_{n}(Z_{i})\phi_{\theta}(Z_{i},\theta_{0}) \right|^{2} \\ \leq 2n^{-1} \sum_{i \in I} \left| P_{k_{1}}'(Z_{i})\widetilde{\beta}_{w\phi_{j},k_{1}} - P_{k_{1}}'(Z_{i})\beta_{w\phi_{j},k_{1},n} \right|^{2} \\ + 2n^{-1} \sum_{i \in I} \left| P_{k_{1}}'(Z_{i})\beta_{w\phi_{j},k_{1},n} - w_{n}(Z_{i})\phi_{\theta}(Z_{i},\theta_{0}) \right|^{2} \\ = (\widetilde{\beta}_{w\phi_{j},k_{1}} - \beta_{w\phi_{j},k_{1},n})' Q_{n,k_{1}}(\widetilde{\beta}_{w\phi_{j},k_{1}} - \beta_{w\phi_{j},k_{1},n}) + o(1)$$
(130)

where the equality is by Assumption 3(vi). Moreover

$$\widetilde{\beta}_{w\phi_j,k_1} - \beta_{w\phi_j,k_1,n} = Q_{n_1,k_1}^{-1} P'_{n,k_1} (\phi_{w\theta_j,n} - \phi_{w\theta_j,n,k_1})'$$
(131)

where $\phi_{w\theta_j,n,k_1} = (P'_{k_1}(Z_i)\beta_{w\phi_j,k_1,n})_{i\in I}$, which implies that

$$\begin{aligned} &(\widetilde{\beta}_{w\phi_{j},k_{1}} - \beta_{w\phi_{j},k_{1},n})'Q_{n,k_{1}}(\widetilde{\beta}_{w\phi_{j},k_{1}} - \beta_{w\phi_{j},k_{1},n}) \\ &\leq \frac{\lambda_{\max}^{2}(Q_{n,k_{1}})}{n\lambda_{\min}^{2}(Q_{n_{1},k_{1}})}(\phi_{w\theta_{j},n} - \phi_{w\theta_{j},n,k_{1}})P_{n,k_{1}}(P_{n,k_{1}}'P_{n,k_{1}})^{-1}P_{n,k_{1}}'(\phi_{w\theta_{j},n} - \phi_{w\theta_{j},n,k_{1}})' \\ &\leq \frac{\lambda_{\max}^{2}(Q_{n,k_{1}})}{\lambda_{\min}^{2}(Q_{n_{1},k_{1}})}n^{-1}\sum_{i\in I}|P_{k_{1}}'(Z_{i})\beta_{w\phi_{j},k_{1},n} - w_{n}(Z_{i})\phi_{\theta}(Z_{i},\theta_{0})|^{2} = o_{p}(1) \end{aligned}$$
(132)

where the equality is by Assumption 3(vi) and (42). Combining the results in (130) and (132), we immediately get (129).

Recall that $\sigma_u^2(z) = E\left[\psi^2 \mid Z = z\right]$. By definition

$$n_1 \Sigma_{n_1} = n_1^{-1} \sum_{i \in I_1} \int_{u}^{2} (Z_i) \widetilde{\phi}_{w\theta,k_1}(Z_i) \widetilde{\phi}_{w\theta,k_1}(Z_i)'$$
(133)

where $\tilde{\phi}_{w\theta,k_1,n}(z) = (\tilde{\phi}_{w\theta_j,k_1,n}(z))'_{j=1,\dots,d_{\theta}}$. First note that

$$n_{1}^{-1} \sum_{i \in I_{1}} \oint_{u}^{2} (Z_{i}) \widetilde{\phi}_{w\theta,k_{1},n}(Z_{i}) \widetilde{\phi}_{w\theta,k_{1},n}(Z_{i})' - n_{1}^{-1} \sum_{i \in I_{1}} \sigma_{u}^{2}(Z_{i}) w_{n}^{2}(Z_{i}) \phi_{\theta}(Z_{i},\theta_{0}) \phi_{\theta}(Z_{i},\theta_{0})'$$

$$= n_{1}^{-1} \sum_{i \in I_{1}} \oint_{u}^{2} (Z_{i}) (\widetilde{\phi}_{w\theta,k_{1},n}(Z_{i}) - w_{n}(Z_{i}) \phi_{\theta}(Z_{i},\theta_{0})) w_{n}(Z_{i}) \phi_{\theta}(Z_{i},\theta_{0})'$$

$$+ n_{1}^{-1} \sum_{i \in I_{1}} \oint_{u}^{2} (Z_{i}) w_{n}(Z_{i}) \phi_{\theta}(Z_{i},\theta_{0}) (\widetilde{\phi}_{w\theta,k_{1},n}(Z_{i}) - w_{n}(Z_{i}) \phi_{\theta}(Z_{i},\theta_{0}))'$$

$$+ n_{1}^{-1} \sum_{i \in I_{1}} \oint_{u}^{2} (Z_{i}) (\widetilde{\phi}_{w\theta,k_{1},n}(Z_{i}) - w_{n}(Z_{i}) \phi_{\theta}(Z_{i},\theta_{0})) (\widetilde{\phi}_{w\theta,k_{1},n}(Z_{i}) - w_{n}(Z_{i}) \phi_{\theta}(Z_{i},\theta_{0}))'. \quad (134)$$

For any any $j_1 = 1, \ldots, d_{\theta}$ and any $j_2 = 1, \ldots, d_{\theta}$, by the Cauchy-Schwarz inequality

$$\begin{cases}
\left| n_{1}^{-1} \sum_{i \in I_{1}} \left(\tilde{\phi}_{u}^{2}(Z_{i})(\tilde{\phi}_{w\theta_{j_{1}},k_{1},n}(Z_{i}) - w_{n}(Z_{i})\phi_{\theta_{j_{1}}}(Z_{i},\theta_{0}))(\tilde{\phi}_{w\theta_{j_{2}},k_{1},n}(Z_{i}) - w_{n}(Z_{i})\phi_{\theta_{j_{2}}}(Z_{i},\theta_{0})) \right|^{2} \\
\leq n_{1}^{-1} \sum_{i \in I_{1}} \left(\tilde{\phi}_{u}^{2}(Z_{i})(\tilde{\phi}_{w\theta_{j_{2}},k_{1},n}(Z_{i}) - w_{n}(Z_{i})\phi_{\theta_{j_{2}}}(Z_{i},\theta_{0}))^{2} \\
\leq Cn_{1}^{-1} \sum_{i \in I_{1}} \left(\tilde{\phi}_{w\theta_{j_{1}},k_{1},n}(Z_{i}) - w_{n}(Z_{i})\phi_{\theta_{j_{1}}}(Z_{i},\theta_{0}))^{2} \\
\times n_{1}^{-1} \sum_{i \in I_{1}} \left(\tilde{\phi}_{w\theta_{j_{2}},k_{1},n}(Z_{i}) - w_{n}(Z_{i})\phi_{\theta_{j_{2}}}(Z_{i},\theta_{0}))^{2} \\
\leq Cn_{1}^{-1} \sum_{i \in I_{1}} \left(\tilde{\phi}_{w\theta_{j_{2}},k_{1},n}(Z_{i}) - w_{n}(Z_{i})\phi_{\theta_{j_{2}}}(Z_{i},\theta_{0}))^{2} \\
\times n_{1}^{-1} \sum_{i \in I_{1}} \left(\tilde{\phi}_{w\theta_{j_{2}},k_{1},n}(Z_{i}) - w_{n}(Z_{i})\phi_{\theta_{j_{2}}}(Z_{i},\theta_{0}))^{2} \\
= o_{p}(1)
\end{cases}$$
(135)

where the second inequality is by Assumption 3(v), the equality is by (129). (135) then implies that

$$n_1^{-1} \sum_{i \in I_1} \oint_u^2 (Z_i) (\widetilde{\phi}_{w\theta, k_1, n}(Z_i) - w_n(Z_i)\phi_\theta(Z_i, \theta_0)) (\widetilde{\phi}_{w\theta, k_1, n}(Z_i) - w_n(Z_i)\phi_\theta(Z_i, \theta_0))' = o_p(1).$$
(136)

For any any $j_1 = 1, \ldots, d_{\theta}$ and any $j_2 = 1, \ldots, d_{\theta}$, by the Cauchy-Schwarz inequality

$$\begin{vmatrix}
n_{1}^{-1} \sum_{i \in I_{1}} \phi_{u}^{2}(Z_{i}) w_{n}(Z_{i}) \phi_{\theta_{j_{1}}}(Z_{i},\theta_{0}) (\widetilde{\phi}_{w\theta_{j_{2}},k_{1},n}(Z_{i}) - w_{n}(Z_{i}) \phi_{\theta_{j_{2}}}(Z_{i},\theta_{0})) \\
\leq \left(n_{1}^{-1} \sum_{i \in I_{1}} \phi_{u}^{2}(Z_{i}) w_{n}^{2}(Z_{i}) \phi_{\theta_{j_{1}}}^{2}(Z_{i},\theta_{0}) \times n_{1}^{-1} \sum_{i \in I_{1}} \sigma_{u}^{2}(Z_{i}) (\widetilde{\phi}_{w\theta_{j_{2}},k_{1},n}(Z_{i}) - u_{n}(Z_{i}) \phi_{\theta_{j_{2}}}(Z_{i},\theta_{0}))^{2} \\
\leq Cn_{1}^{-1} \sum_{i \in I_{1}} \phi_{\theta_{j_{1}}}^{2}(Z_{i},\theta_{0}) \times n_{1}^{-1} \sum_{i \in I_{1}} (\widetilde{\phi}_{w\theta_{j_{2}},k_{1},n}(Z_{i}) - w_{n}(Z_{i}) \phi_{\theta_{j_{2}}}(Z_{i},\theta_{0}))^{2} = o_{p}(1)$$
(137)

where the second inequality is by Assumptions 2(v) and 3(v), the equality is by (129) and (82). (137) implies that

$$n_1^{-1} \sum_{i \in I_1} \oint_u^2 (Z_i) (\widetilde{\phi}_{w\theta, k_1, n}(Z_i) - w_n(Z_i) \phi_\theta(Z_i, \theta_0)) w_n(Z_i) \phi_\theta(Z_i, \theta_0)' = o_p(1)$$
(138)

and

$$n_1^{-1} \sum_{i \in I_1} \oint_u^2 (Z_i) w_n(Z_i) \phi_\theta(Z_i, \theta_0) (\widetilde{\phi}_{w\theta, k_1, n}(Z_i) - w_n(Z_i) \phi_\theta(Z_i, \theta_0))' = o_p(1).$$
(139)

Combining the results in (134), (136), (138) and (139), we have

$$n_1^{-1} \sum_{i \in I_1} \sigma_u^2(Z_i) \widetilde{\phi}_{w\theta,k_1,n}(Z_i) \widetilde{\phi}_{w\theta,k_1,n}(Z_i)' - n_1^{-1} \sum_{i \in I_1} \oint_{u}^2 (Z_i) w_n^2(Z_i) \phi_\theta(Z_i,\theta_0) \phi_\theta(Z_i,\theta_0)' = o_p(1).$$
(140)

For any $j_1 = 1, \ldots, d_{\theta}$ and any $j_2 = 1, \ldots, d_{\theta}$,

where the first inequality is by Assumptions 2(v) and 3(v), the second inequality is by the Hölder inequality, and the last inequality is by Assumption 3(iv). The i.i.d. assumption together with (141) and the Markov inequality implies that

$$n_1^{-1} \sum_{i \in I_1} \oint_u^2 (Z_i) w_n^2(Z_i) \phi_{\theta_{j_1}}(Z_i, \theta_0) \phi_{\theta_{j_2}}(Z_i, \theta_0) - E\left[\sigma_u^2(Z) w_n^2(Z) \phi_{\theta_{j_1}}(Z, \theta_0) \phi_{\theta_{j_2}}(Z, \theta_0)\right] = o_p(1).$$
(142)

Collecting the results in (140) and (142), we have

$$n_1 \Sigma_{n_1} = E \left[u^2 w_n^2(Z) \phi_\theta(Z, \theta_0) \phi_\theta(Z, \theta_0)' \right] + o_p(1)$$
(143)

which proves the first claim of the lemma. The proof of the second claim of the lemma is similar and hence omitted. $\hfill \Box$

Proof. [Proof of Lemma 1] By Lemma 7,

$$\Sigma_{n_1} + \Sigma_{n_2} = E\left[w_n^2(Z)\left(\frac{\sigma_u^2(Z)}{n_1} + \frac{\sigma_\varepsilon^2(Z)}{n_2}\right)\phi_\theta(Z,\theta_0)\phi_\theta'(Z,\theta_0)\right] \left(+ o_p(n_1^{-1} + n_2^{-1}).$$
(144)

By Assumptions 2(v), 3(ii) and 3(v),

$$E\left[w_n^2(Z)\left(\frac{\sigma_u^2(Z)}{n_1} + \frac{\sigma_\varepsilon^2(Z)}{n_2}\right) \phi_\theta(Z,\theta_0)\phi_\theta'(Z,\theta_0)\right] \rightleftharpoons C^{-1}(n_1^{-1} + n_2^{-1})$$
(145)

which together with (144) implies that

$$\Sigma_{n_1} + \Sigma_{n_2} = E\left[w_n^2(Z)\left(\frac{\sigma_u^2(Z)}{n_1} + \frac{\sigma_\varepsilon^2(Z)}{n_2}\right)\phi_\theta(Z,\theta_0)\phi_\theta'(Z,\theta_0)\right]\left(1 + o_p(1)\right).$$
(146)

The claim of the lemma then follows by combining (146) and Assumption 3(ii).

Proof. [Proof of Theorem 3] Let $w_n^*(Z) = (n_1^{-1} + n_2^{-1})(n_1^{-1}\sigma_u^2(Z) + n_2^{-1}\sigma_{\varepsilon}^2(Z))^{-1}$ and $H_{0,n}^* = E[w_n^*(Z)\phi_{\theta}(Z,\theta_0)\phi_{\theta}'(Z,\theta_0)]$. Then by definition

$$V_{n,\theta}^* = (n_1^{-1} + n_2^{-1})(H_{0,n}^*)^{-1}.$$
(147)

For any $w_n(\cdot)$, define

$$A_n(Z, w_n) = H_{0,n}^{-1} \phi_\theta(Z, \theta_0) w_n(Z) - H_{0,n}^{*-1} \phi_\theta(Z, \theta_0) w_n^*(Z).$$
(148)

Then we have

$$V_{n,\theta} - V_{n,\theta}^* = E\left[A_n(Z, w_n)\Omega_n(Z)A_n(Z, w_n)'\right] \ge 0,$$
(149)

for any n_1 and any n_2 , where $\Omega_n(Z) = n_1^{-1}\sigma_u^2(Z) + n_2^{-1}\sigma_\varepsilon^2(Z)$ and the inequality is by the fact that $\Omega_n(Z) \in (0,\infty)$ and $A_n(Z,w_n)\Omega_n(Z)A_n(Z,w_n)'$ is a positive semidefinite matrix almost surely. \Box

Proof. [Proof of Lemma 2] By definition $w_n^*(z)^{-1} = (n_1^{-1}\sigma_u^2(z) + n_2^{-1}\sigma_\varepsilon^2(z))(n_1^{-1} + n_2^{-1})^{-1}$. Then for any n_1 , n_2 ,

$$\min\left\{ \left(\inf_{z \in \mathcal{Z}} \sigma_u^2(z), \inf_{z \in \mathcal{Z}} \sigma_\varepsilon^2(z) \right\} \le w_n^*(z)^{-1} \le \max\left\{ \sup_{z \in \mathcal{Z}} \sigma_u^2(z), \sup_{z \in \mathcal{Z}} \sigma_\varepsilon^2(z) \right\}$$

which together with Assumption 3(v) proves the claim of the lemma.

C Proof of the Main Results in Section 5

Lemma 8. Under Assumptions 1, 2, 3 and 4(i), we have

$$n_{2}^{-1} \sum_{i \in I_{2}} \left| \widehat{\phi}_{n_{2}}(Z_{i}, \widehat{\theta}_{n}) - \phi(Z_{i}, \theta_{0}) \right|^{2} = O_{p}(n_{1}^{-1} + k_{2}n_{2}^{-1} + k_{2}^{-2r_{h}})$$
(150)

and moreover,

$$\sup_{z\in\mathcal{Z}} \left| \widehat{\phi}_{n_2}(z,\widehat{\theta}_n) - \phi(z,\theta_0) \right| = O_p(\xi_{k_2} n_1^{-1/2} + \xi_{k_2} k_2^{1/2} n_2^{-1/2} + \xi_{k_2} k_2^{-r_h}).$$
(151)

Proof. [Proof of Lemma 8] By definition,

$$\begin{aligned} \widehat{\phi}_{n_2}(z,\widehat{\theta}_n) &= P'_{k_2}(z)(P'_{k_2,n_2}P_{k_2,n_2})^{-1}P'_{k_2,n_2}g_{n_2}(\widehat{\theta}_n);\\ \widehat{\phi}_{n_2}(z,\theta_0) &= P'_{k_2}(z)(P'_{k_2,n_2}P_{k_2,n_2})^{-1}P'_{k_2,n_2}g_{n_2}(\theta_0), \end{aligned}$$

where $g_{n_2}(\widehat{\theta}_n) = (g(Z_i, \widehat{\theta}_n))'_{i \in I_2}$ and $g_{n_2}(\theta_0) = (g(Z_i, \theta_0))'_{i \in I_2}$. By the triangle inequality, for any z,

$$\begin{aligned} \left| \widehat{\phi}_{n_2}(z, \widehat{\theta}_n) - \phi(z, \theta_0) \right| &\leq \left| \widehat{\phi}_{n_2}(z, \widehat{\theta}_n) - \widehat{\phi}_{n_2}(z, \theta_0) \right| \\ &+ \left| \widehat{\phi}_{n_2}(z, \theta_0) - h_{0, k_2}(z) \right| + \left| h_{0, k_2}(z) - h_0(z) \right|. \end{aligned}$$
(152)

By the mean value expansion and the Cauchy-Schwarz inequality,

$$\left|g(X_i,\widehat{\theta}_n) - g(X_i,\theta_0)\right| \le \sup_{\theta \in \mathcal{N}_n} \left\|g_\theta(X_i,\theta)\right\| \left\|\widehat{\theta}_n - \theta_0\right\|$$
(153)

which together with Assumption 4(i) and (9) in Theorem 2 implies that

$$n_{2}^{-1} \sum_{i \in I_{2}} \left| g(Z_{i}, \widehat{\theta}_{n}) - g(Z_{i}, \theta_{0}) \right|^{2} \leq \sup_{\theta \in \mathcal{N}_{n}} n_{2}^{-1} \sum_{i \in I_{2}} \left| g_{\theta}(X_{i}, \theta) \|^{2} \left\| \widehat{\theta}_{n} - \theta_{0} \right\|^{2} = O_{p}(n_{1}^{-1} + n_{2}^{-1}).$$
(154)

Under Assumptions 1(i), 1(iii), 1(iv) and 3(v), we can use similar arguments in proving (61) to show that

$$n_2^{-1} \sum_{i \in I_2} \left(\widehat{\phi}_{n_2}(Z_i, \theta_0) - h_{0,k_2}(Z_i))^2 = O_p(k_2 n_2^{-1} + k_2^{-2r_h}).$$
(155)

Using (154), we get

$$n_{2}^{-1} \sum_{i \in I_{2}} \left| \widehat{\phi}_{n_{2}}(Z_{i}, \widehat{\theta}_{n}) - \widehat{\phi}_{n_{2}}(Z_{i}, \theta_{0}) \right|^{2} \left(\sum_{\substack{g_{n_{2}}(\widehat{\theta}_{n}) - g_{n_{2}}(\theta_{0}) \\ -g_{n_{2}}(\widehat{\theta}_{n}) - g_{n_{2}}(\theta_{0}) \right|^{2} P_{k_{2},n_{2}}(P_{k_{2},n_{2}})^{-1} P_{k_{2},n_{2}}' [g_{n_{2}}(\widehat{\theta}_{n}) - g_{n_{2}}(\theta_{0})] }{n_{2}} \\ \leq n_{2}^{-1} \sum_{i \in I_{2}} \left| g(Z_{i}, \widehat{\theta}_{n}) - g(Z_{i}, \theta_{0}) \right|^{2} = O_{p}(n_{1}^{-1} + n_{2}^{-1}),$$
(156)

which together with (152), (155) and Assumption 1(iv) implies that

$$n_{2}^{-1} \sum_{i \in I_{2}} \left| \widehat{\phi}_{n_{2}}(Z_{i}, \widehat{\theta}_{n}) - \phi(Z_{i}, \theta_{0}) \right|_{\ell}^{2} = O_{p}(n_{1}^{-1} + k_{2}n_{2}^{-1} + k_{2}^{-2r_{h}}).$$
(157)

This proves the first claim of the lemma.

By (60) and the Cauchy-Schwarz inequality,

$$\sup_{z \in \mathcal{Z}} \left| \widehat{\phi}_{n_2}(z, \theta_0) - h_{0, k_2}(z) \right| = O_p(\xi_{k_2} k_2^{1/2} n_2^{-1/2} + \xi_{k_2} k_2^{-r_h}).$$
(158)

Using (154), we get

$$\sup_{z \in \mathcal{Z}} \left| \widehat{\phi}_{n_{2}}(z, \widehat{\theta}_{n}) - \widehat{\phi}_{n_{2}}(z, \theta_{0}) \right| \\
\leq \xi_{k_{2}} \left\| \left[g_{n_{2}}(\widehat{\theta}_{n}) - g_{n_{2}}(\theta_{0}) \right]' P_{k_{2}, n_{2}} (P'_{k_{2}, n_{2}} P_{k_{2}, n_{2}})^{-1} \right\| \\
\leq \xi_{k_{2}} \left\| n_{2}^{-1} \sum_{i \in I_{2}} \left(g(Z_{i}, \widehat{\theta}_{n}) - g(Z_{i}, \theta_{0}) \right)^{2} \right)^{1/2} \\
= O_{p}(\xi_{k_{2}} n_{1}^{-1/2} + \xi_{k_{2}} n_{2}^{-1/2}) \tag{159}$$

which together with (152), (158) and Assumption 1(iv) implies that

$$\sup_{z\in\mathcal{Z}} \left| \widehat{\phi}_{n_2}(z,\widehat{\theta}_n) - \phi(z,\theta_0) \right| = O_p(\xi_{k_2} n_1^{-1/2} + \xi_{k_2} k_2^{1/2} n_2^{-1/2} + \xi_{k_2} k_2^{-r_h}).$$
(160)

This proves the second claim of the lemma.

Lemma 9. Under the conditions of Theorem 4, we have

$$\begin{array}{l} (i) \ \widehat{H}_{n} = H_{0,n} + o_{p}(1); \\ (ii) \ n^{-1}(\widehat{\phi}_{w\theta,n} - \phi_{w\theta,n})'P_{n,k_{1}} = o_{p}(1); \\ (iii) \ n^{-1}(\widehat{\phi}_{w\theta,n} - \phi_{w\theta,n})'P_{n,k_{2}} = o_{p}(1); \\ (iv) \ \sup_{\{\gamma_{k} \in R^{k}: \ \gamma_{k}'\gamma_{k} = 1\}} \left[\gamma_{k}'(\widehat{Q}_{n_{1},u} - Q_{n_{1},u})\gamma_{k} \right] = o_{p}(1); \\ (v) \ \sup_{\{\gamma_{k} \in R^{k}: \ \gamma_{k}'\gamma_{k} = 1\}} \left[\gamma_{k}'(\widehat{Q}_{n_{2},\varepsilon} - Q_{n_{2},\varepsilon})\gamma_{k} \right] = o_{p}(1).$$

Proof. [Proof of Lemma 9] (i) The proof of the first claim follows by the consistency of $\hat{\theta}_n$ and similar arguments in deriving (86). Hence it is omitted.

(ii) By definition,

$$n^{-1}(\widehat{\phi}_{w\theta,n} - \phi_{w\theta,n})'P_{n,k_{1}} = n^{-1} \sum_{i \in I} \left(\widehat{w}_{n}(Z_{i}) - w_{n}(Z_{i})) \widehat{\phi}_{\theta,n_{2}}(Z_{i},\widehat{\theta}_{n}) P_{k_{1}}(Z_{i}) + n^{-1} \sum_{i \in I} \left(w_{n}(Z_{i})(\widehat{\phi}_{\theta,n_{2}}(Z_{i},\widehat{\theta}_{n}) - \phi_{\theta}(Z_{i},\theta_{0})) P_{k_{1}}(Z_{i}). \right) \right)$$
(161)

By (9) in Theorem 2, Lemma 4 and similar arguments in showing (80),

$$n^{-1} \sum_{i \in I} \left\| \widehat{\phi}_{\theta, n_{2}}(Z_{i}, \widehat{\theta}_{n}) - \widehat{\phi}_{\theta, n_{2}}(Z_{i}, \theta_{0}) \right\|^{2} \left\| \widehat{\phi}_{\theta, n_{2}}(Z_{i}, \theta) \right\|^{2} \left\| \widehat{\theta}_{n} - \theta_{0} \right\|^{2} = O_{p}(n_{1}^{-1} + n_{2}^{-1}),$$
(162)
sumption 3(iii) implies that

which together with Assumption 3(iii) implies that

By (42), (82), (163), Assumption 4(iv), and the Cauchy-Schwarz inequality,

$$\left\| n^{-1} \sum_{i \in I} \left(\widehat{w}_{n}(Z_{i}) - w_{n}(Z_{i}) \right) \widehat{\phi}_{\theta, n_{2}}(Z_{i}, \widehat{\theta}_{n}) P_{k_{1}}(Z_{i}) \right\|^{2}$$

$$\leq \sup_{z \in \mathcal{Z}} \left| \widehat{w}_{n}(z) - w_{n}(z) \right|^{2} \quad n^{-1} \sum_{i \in I} \left\| \widehat{\phi}_{\theta, n_{2}}(Z_{i}, \widehat{\theta}_{n}) \right\|^{2}$$

$$= o_{p}(k_{1}\delta_{w, n}^{2}) = o_{p}(1),$$

$$(164)$$

where the $n^{-1} \sum_{i \in I} \left\| \widehat{\phi}_{\theta, n_2}(Z_i, \widehat{\theta}_n) \right\|_{-1}^2 = O_p(1)$ is used in the first equality which is by (82) and (163). By

Assumptions 2(v) and 3(vii), (163) and the Cauchy-Schwarz inequality,

$$\left\| n^{-1} \sum_{i \in I} \left(w_n(Z_i)(\widehat{\phi}_{\theta, n_2}(Z_i, \widehat{\theta}_n) - \phi_{\theta}(Z_i, \theta_0)) P_{k_1}(Z_i) \right) \right\|_{Z \in \mathcal{Z}}^2$$

$$\leq \left(\sup_{z \in \mathcal{Z}} |w_n(z)| n^{-1} \sum_{i \in I} \left\| \widehat{\phi}_{\theta, n_2}(Z_i, \widehat{\theta}_n) - \phi_{\theta}(Z_i, \theta_0) \right\|_{Q}^2 n^{\left(-1 \sum_{i \in I} \|P_{k_1}(Z_i)\|^2 = o_p(1). \right)}$$

$$(165)$$

Collecting the results in (161), (164) and (165), we immediately prove the second claim of the lemma.

- (iii) The proof of this result is similar to the arguments in the proof in (ii) and hence is omitted.
- (iv) By definition,

$$\widehat{Q}_{n_{1},u} - Q_{n_{1},u} = n_{1}^{-1} \sum_{i \in I_{1}} (Y_{i} - \widehat{h}_{n_{1}}(Z_{i}))^{2} P_{k_{1}}(Z_{i}) P'_{k_{1}}(Z_{i}) - Q_{n_{1},u}
= n_{1}^{-1} \sum_{i \in I_{1}} (u_{i}^{2} - \sigma_{u}^{2}(Z_{i})) P_{k_{1}}(Z_{i}) P'_{k_{1}}(Z_{i})
+ n_{1}^{-1} \sum_{i \in I_{1}} (\widehat{h}_{n_{1}}(Z_{i}) - h_{0}(Z_{i}))^{2} P_{k_{1}}(Z_{i}) P'_{k_{1}}(Z_{i})
- 2n_{1}^{-1} \sum_{i \in I_{1}} (\widehat{h}_{n_{1}}(Z_{i}) - h_{0}(Z_{i})) u_{i} P_{k_{1}}(Z_{i}) P'_{k_{1}}(Z_{i}).$$
(166)

By (42), Assumptions 1(i), 3(v),

$$E\left[\left\| n_{1}^{-1} \sum_{i \in I_{1}} \left(\psi_{i}^{2} - \sigma_{u}^{2}(Z_{i}) \right) P_{k_{1}}(Z_{i}) P_{k_{1}}'(Z_{i}) \right\|^{2} \left\{ Z_{i} \right\}_{i \in I_{1}} \right] \left(\sum_{i \in I_{1}} E\left[\left(\psi_{i}^{2} - \sigma_{u}^{2}(Z_{i}) \right)^{2} | Z_{i} \right] |P_{k_{1}}'(Z_{i}) P_{k_{1}}(Z_{i}) \right]^{2} \left(Z_{i} \right)^{2} \left($$

which together with the Markov inequality and Assumption 4(iv) implies that

$$n_1^{-1} \sum_{i \in I_1} \left(u_i^2 - \sigma_u^2(Z_i) \right) P_{k_1}(Z_i) P'_{k_1}(Z_i) = o_p(1).$$
(168)

Using Assumption 1(iv), (43), we get

$$\sup_{z \in \mathcal{Z}} \left| \widehat{h}_{n_{1}}(z) - h_{0}(z) \right| \leq \sup_{z \in \mathcal{Z}} \left| \widehat{h}_{n_{1}}(z) - h_{k_{1}}(z) \right| + \sup_{z \in \mathcal{Z}} \left| h_{0,k_{1}}(z) - h_{0}(z) \right| \\
\leq \sup_{z \in \mathcal{Z}} \left| (\widehat{\beta}_{k_{1},n_{1}} - \beta_{h,k_{1}})' P_{k_{1}}(z) \right| + \sup_{z \in \mathcal{Z}} \left| h_{0,k_{1}}(z) - h_{0}(z) \right| \\
\leq \left\| \widehat{\beta}_{k_{1},n_{1}} - \beta_{h,k_{1}} \right\| \xi_{k_{1}} + \sup_{z \in \mathcal{Z}} \left| h_{0,k_{1}}(z) - h_{0}(z) \right| \\
= O_{p}(\xi_{k_{1}}k_{1}^{1/2}n_{1}^{-1/2} + \xi_{k_{1}}k_{1}^{-r_{h}}),$$
(169)

where the first inequality is by the triangle inequality, the second inequality is by the Cauchy-Schwarz

inequality. For any $\gamma_k \in \mathbb{R}^k$ with $\gamma'_k \gamma_k = 1$, we have

$$\sup_{\{\gamma_{k}\in R^{k}: \gamma_{k}'\gamma_{k}=1\}} n_{1}^{-1} \sum_{i\in I_{1}} \left(\widehat{h}_{n_{1}}(Z_{i}) - h_{0}(Z_{i}) \right)^{2} \gamma_{k}' P_{k_{1}}(Z_{i}) P_{k_{1}}'(Z_{i}) \gamma_{k} \\
\leq \sup_{z} \left| \widehat{h}_{n_{1}}(z) - h_{0}(z) \right|^{2} \sup_{\{\gamma_{k}\in R^{k}: \gamma_{k}'\gamma_{k}=1\}} \frac{1}{n_{1}} \sum_{i\in I_{1}} \gamma_{k}' P_{k_{1}}(Z_{i}) P_{k_{1}}'(Z_{i}) \gamma_{k} \\
\leq \sup_{z} \left| \widehat{h}_{n_{1}}(z) - h_{0}(z) \right|^{2} \max_{\max}(Q_{k_{1},n_{1}}) \\
= O_{p}(\xi_{k_{1}}^{2}k_{1}n_{1}^{-1} + \xi_{k_{1}}^{2}k_{1}^{2r_{h}}) = o_{p}(1)$$
(170)

where the first equality is by (42) and (169), the last equality is by Assumption 4(iv). By the triangle inequality,

$$\sup_{\substack{\{\gamma_{k}\in R^{k}: \gamma_{k}'\gamma_{k}=1\}}} \left\| n_{1}^{-1} \sum_{i\in I_{1}} (h_{0,k_{1}}(Z_{i}) - h_{0}(Z_{i})) u_{i}\gamma_{k}' P_{k_{1}}(Z_{i}) P_{k_{1}}'(Z_{i})\gamma_{k} \right\| \\
\leq \left\| n_{1}^{-1} \sum_{i\in I_{1}} (h_{0,k_{1}}(Z_{i}) - h_{0}(Z_{i})) u_{i} P_{k_{1}}(Z_{i}) P_{k_{1}}'(Z_{i}) \right\|. \tag{171}$$

$$(171)$$

$$(171)$$

By Assumptions 1(i), 1(ix), 1(iv) and 3(v),

$$E \left[\left\| n_{1}^{-1} \sum_{i \in I_{1}} \left(h_{0,k_{1}}(Z_{i}) - h_{0}(Z_{i}) \right) u_{i} P_{k_{1}}(Z_{i}) P_{k_{1}}'(Z_{i}) \right\|^{2} \right]$$

$$= n_{1}^{-1} E \left[\sum_{i \notin I_{1}} (h_{0,k_{1}}(Z_{i}) - h_{0}(Z_{i}))^{2} u_{i}^{2} \left| P_{k_{1}}'(Z_{i}) P_{k_{1}}(Z_{i}) \right|^{2} \right]$$

$$\leq C n_{1}^{-1} \sup_{z \in \mathcal{Z}} \left| h_{0,k_{1}}(z) - h_{0}(z) \right|^{2} \xi_{k_{1}}^{2} E \left[P_{k_{1}}'(Z_{i}) P_{k_{1}}(Z_{i}) \right] \stackrel{P}{\leftarrow} O_{p}(\xi_{k_{1}}^{2} k_{1}^{1-2r_{h}} n_{1}^{-1})$$
(172)

which together with (171), Assumption 4(iv), and the Markov inequality implies that

$$\sup_{\{\gamma_k \in R^k: \, \gamma'_k \gamma_k = 1\}} \left\| n_1^{-1} \sum_{i \in I_1} (h_{0,k_1}(Z_i) - h_0(Z_i)) u_i \gamma'_k P_{k_1}(Z_i) P'_{k_1}(Z_i) \gamma_k \right\|_{(173)} = o_p(1).$$
(173)
-Schwarz inequality,

By the Cauchy-Schwarz inequality,

$$\sup_{\{\gamma_{k}\in R^{k}: \gamma_{k}'\gamma_{k}=1\}} \left\| n_{1}^{-1} \sum_{i\in I_{1}} \left(\widehat{h}_{n_{1}}(Z_{i}) - h_{0,k_{1}}(Z_{i})) u_{i}\gamma_{k}' P_{k_{1}}(Z_{i}) P_{k_{1}}'(Z_{i})\gamma_{k} \right\|^{2} \\
\leq \left\| \widehat{\beta}_{k_{1},n_{1}} - \beta_{h,k_{1}} \right\|^{2} \left(\sup_{\substack{\{\gamma_{k}\in R^{k}: \gamma_{k}'\gamma_{k}=1\} \\ j=1 \\ j=1$$

By Assumptions 1(i), 1(iii) and 3(v),

$$E\left[\sum_{j=1}^{k_{1}}\left(n_{1}^{-1}\sum_{i\in I_{1}}\left(j(Z_{i})u_{i}P_{k_{1}}(Z_{i})P_{k_{1}}'(Z_{i})||^{2}\right)\right]\right)$$

$$=\sum_{j=1}^{k_{1}}n_{1}^{-1}E\left[p_{j}^{2}(Z_{i})u_{i}^{2}\left|P_{k_{1}}'(Z_{i})P_{k_{1}}(Z_{i})\right|^{2}\right]$$

$$\leq Cn_{1}^{-1}\xi_{k_{1}}^{4}\sum_{j=1}^{k_{1}}E\left[p_{j}^{2}(Z_{i})\right] \in O(\xi_{k_{1}}^{4}k_{1}n_{1}^{-1}), \qquad (175)$$

which together with (43), (174), Assumption 4(iv), and the Markov inequality implies that

$$\sup_{\{\gamma_k \in R^k: \gamma'_k \gamma_k = 1\}} \left\| n_1^{-1} \sum_{i \in I_1} \left\| \hat{h}_{n_1}(Z_i) - h_{0,k_1}(Z_i) \right\|_{i \gamma'_k} P_{k_1}(Z_i) P'_{k_1}(Z_i) \gamma_k \right\|_{i = O_p(\xi_{k_1}^2 k_1^{1/2} n_1^{-1/2} \left\| k_1^{1/2} n_1^{-1/2} + k_1^{-r_h} \right\|_{i = O_p(1)} \right\|_{i = O_p(1)}$$
(176)

Combining the results in (173), (176) and the triangle inequality, we get

$$\sup_{\{\gamma_k \in R^k: \ \gamma'_k \gamma_k = 1\}} \left\| n_1^{-1} \sum_{i \in I_1} \left\| \widehat{h}_{n_1}(Z_i) - h_0(Z_i) \right\|_{u_i \gamma'_k} P_{k_1}(Z_i) P'_{k_1}(Z_i) \gamma_k \right\|_{\ell} = o_p(1).$$
(177)

Combining the results in (166), (168), (170) and (177), we prove the fourth claim of the lemma.

(v) By definition,

$$\widehat{Q}_{n_{2},\varepsilon} - Q_{n_{2},\varepsilon} = n_{2}^{-1} \sum_{i \in I_{2}} \left(\widehat{q}_{i}^{2} - \sigma_{\varepsilon}^{2}(Z_{i}) \right) P_{k_{2}}(Z_{i}) P_{k_{2}}(Z_{i}) + 2n_{2}^{-1} \sum_{i \in I_{2}} \left[\left((X_{i}, \widehat{\theta}_{n}) - \widehat{\phi}(Z_{i}, \widehat{\theta}_{n}) - g(X_{i}, \theta_{0}) + \phi(Z_{i}, \theta_{0}) \right] \left((P_{k_{2}}(Z_{i}) P_{k_{2}}(Z_{i}) + n_{2}^{-1} \sum_{i \in I_{2}} \left[\left| g(X_{i}, \widehat{\theta}_{n}) - \widehat{\phi}(Z_{i}, \widehat{\theta}_{n}) - g(X_{i}, \theta_{0}) + \phi(Z_{i}, \theta_{0}) \right|^{2} \right] P_{k_{2}}(Z_{i}) P_{k_{2}}(Z_{i}). \quad (178)$$
similar arguments in showing (168), we get

Using similar arguments in showing (168), we get

$$n_2^{-1} \sum_{i \in I_2} \left(\boldsymbol{\varepsilon}_i^2 - E[\boldsymbol{\varepsilon}_i^2 | Z_i] \right) P_{k_2}(Z_i) P_{k_2}'(Z_i) = O_p(\xi_{k_2} k_2^{1/2} n_2^{-1/2}) = o_p(1).$$
(179)

By the Cauchy-Schwarz inequality,

$$\sup_{\{\gamma_k \in R^k: \ \gamma'_k \gamma_k = 1\}} n_2^{-1} \sum_{i \in I_2} \left(g(X_i, \widehat{\theta}_n) - g(X_i, \theta_0))^2 \gamma'_k P_{k_2}(Z_i) P'_{k_2}(Z_i) \gamma_k \right)$$

$$\leq \xi_{k_2}^2 n_2^{-1} \sum_{i \in I_2} \left(g(X_i, \widehat{\theta}_n) - g(X_i, \theta_0))^2 = O_p(\xi_{k_2}^2 n_1^{-1} + \xi_{k_2}^2 n_2^{-1}) = o_p(1) \right)$$
(180)

where the first equality is by (154), the second equality is by Assumption 4(iv). Similarly

$$\sup_{\{\gamma_k \in R^k: \ \gamma'_k \gamma_k = 1\}} n_2^{-1} \sum_{i \in I_2} \left(\widehat{\phi}(Z_i, \widehat{\theta}_n) - \phi(Z_i, \theta_0))^2 \gamma'_k P_{k_2}(Z_i) P'_{k_2}(Z_i) \gamma_k \right)$$

$$\leq \xi_{k_2}^2 n_2^{-1} \sum_{i \in I_2} \left(\widehat{\phi}(Z_i, \widehat{\theta}_n) - \phi(Z_i, \theta_0))^2 = O_p(\xi_{k_2}^2 k_2 n_2^{-1} + \xi_{k_2}^2 k_2^{-2r_h}) = o_p(1)$$
(181)

where the first equality is by (150), the second equality is by Assumption 4(iv). Collecting the results in (180) and (181), we have

$$\sup_{\{\gamma_{k}\in R^{k}: \gamma_{k}'\gamma_{k}=1\}} n_{2}^{-1} \sum_{i\in I_{2}} \left\{ \left| g(X_{i},\widehat{\theta}_{n}) - \widehat{\phi}(Z_{i},\widehat{\theta}_{n}) - g(X_{i},\theta_{0}) - \phi(Z_{i},\theta_{0}) \right|^{2} \right\} \left\{ \gamma_{k}'P_{k_{2}}(Z_{i})P_{k_{2}}'(Z_{i})P_{k_{2}}'(Z_{i})\gamma_{k} \right\} \\
\leq \sup_{\{\gamma_{k}\in R^{k}: \gamma_{k}'\gamma_{k}=1\}} 2n_{2}^{-1} \sum_{i\in I_{2}} \left[\left| g(X_{i},\widehat{\theta}_{n}) - g(X_{i},\theta_{0}) \right|^{2} \right] \gamma_{k}'P_{k_{2}}(Z_{i})P_{k_{2}}'(Z_{i})\gamma_{k} \\
+ \sup_{\{\gamma_{k}\in R^{k}: \gamma_{k}'\gamma_{k}=1\}} 2n_{2}^{-1} \sum_{i\in I_{2}} \left[\left| \widehat{\phi}(Z_{i},\widehat{\theta}_{n}) - \phi(Z_{i},\theta_{0}) \right|^{2} \right] \left\{ \gamma_{k}'P_{k_{2}}(Z_{i})P_{k_{2}}'(Z_{i})\gamma_{k} = o_{p}(1). \right\}$$
(182)

Next, note that by the Cauchy-Schwarz inequality and the triangle inequality,

$$\begin{vmatrix}
n_{2}^{-1} \sum_{i \in I_{2}} \left(g(X_{i}, \widehat{\theta}_{n}) - \widehat{\phi}(Z_{i}, \widehat{\theta}_{n}) - g(X_{i}, \theta_{0}) - \phi(Z_{i}, \theta_{0}))\varepsilon_{i}\gamma_{k}'P_{k_{2}}(Z_{i})P_{k_{2}}'(Z_{i})\gamma_{k} \\
\leq \begin{vmatrix}
n_{2}^{-1} \sum_{i \in I_{2}} \left(g(X_{i}, \widehat{\theta}_{n}) - g(X_{i}, \theta_{0}))\varepsilon_{i}\gamma_{k}'P_{k_{2}}(Z_{i})P_{k_{2}}'(Z_{i})\gamma_{k} \\
+ \begin{vmatrix}
n_{2}^{-1} \sum_{i \in I_{2}} \left(\widehat{\phi}(Z_{i}, \widehat{\theta}_{n}) - \phi(Z_{i}, \theta_{0}))\varepsilon_{i}\gamma_{k}'P_{k_{2}}(Z_{i})P_{k_{2}}'(Z_{i})\gamma_{k} \\
\end{vmatrix} \end{vmatrix} \begin{vmatrix}
(183)$$

By the definition of γ_k , we can use the Cauchy-Schwarz inequality to show that

$$\sup_{\{\gamma_{k} \in R^{k}: \gamma_{k}' \gamma_{k} = 1\}} n_{2}^{-1} \sum_{i \in I_{2}} \left\{ \stackrel{2}{\epsilon} \left| \gamma_{k}' P_{k_{2}}(Z_{i}) P_{k_{2}}'(Z_{i}) \gamma_{k} \right|^{2} \right\}$$

$$\leq \sup_{\{\gamma_{k} \in R^{k}: \gamma_{k}' \gamma_{k} = 1\}} \xi_{k_{2}}^{2} \gamma_{k}' \quad n_{2}^{-1} \sum_{i \in I_{2}} \left\{ \stackrel{2}{\epsilon} P_{k_{2}}(Z_{i}) P_{k_{2}}'(Z_{i}) \right\} \left\{ \gamma_{k} \left\{ \xi_{k_{2}}^{2} \lambda_{\max} \quad n_{2}^{-1} \sum_{i \in I_{2}} \left\{ \stackrel{2}{\epsilon} P_{k_{2}}(Z_{i}) P_{k_{2}}'(Z_{i}) \right\} \right\} \right\} \left\{ (184)$$

By Assumptions 1(i), 1(iii), 3(v),

$$E\left[\left\|n_{1}\sum_{i\in I_{1}}\left(\varepsilon_{i}^{2}-\sigma_{\varepsilon}^{2}(Z_{i})\right)P_{k_{1}}(Z_{i})P_{k_{1}}(Z_{i})'\right\|^{2}\right]$$

$$= n_{1}^{-1}E\left[\left(\varepsilon_{i}^{2}-\sigma_{\varepsilon}^{2}(Z_{i})\right)^{2}|P_{k_{1}}(Z_{i})'P_{k_{1}}(Z_{i})|^{2}\right]$$

$$\leq Cn_{1}^{-1}\xi_{k_{1}}^{2}E\left[P_{k_{1}}(Z_{i})'P_{k_{1}}(Z_{i})\right] \leq Ck_{1}\xi_{k_{1}}^{2}n_{1}^{-1}$$
(185)

which together with the Markov inequality and Assumption 4(iv) implies that

$$n_1 \sum_{i \in I_1} \left\{ i P_{k_1}(Z_i) P_{k_1}(Z_i)' - Q_{n_1,\varepsilon} = o_p(1). \right.$$
(186)

By (102) and (186), we have

$$\lambda_{\max} \quad n_2^{-1} \sum_{i \in I_2} \left(\sum_{i=1}^{2} P_{k_2}(Z_i) P_{k_2}'(Z_i) \right) \left(\leq C \right)$$

$$(187)$$

with probability approaching 1, which together with (154), (184) and Assumption 4(iv), implies that

$$n_{2}^{-1} \sum_{i \in I_{2}} \left(g(X_{i}, \widehat{\theta}_{n}) - g(X_{i}, \theta_{0}))^{2} \times n_{2}^{-1} \sum_{i \in I_{2}} \varepsilon_{i}^{2} \left| \gamma_{k}' P_{k_{2}}(Z_{i}) P_{k_{2}}'(Z_{i}) \gamma_{k} \right|^{2} \right)^{1/2}$$

= $O_{p}(\xi_{k_{2}} n_{1}^{-1/2} + \xi_{k_{2}} n_{2}^{-1/2}) = o_{p}(1).$ (188)

Similarly, by (150), (184), (187) and Assumption 4(iv)

$$n_{2}^{-1} \sum_{i \in I_{2}} \left(\widehat{\phi}(Z_{i}, \widehat{\theta}_{n}) - \phi(Z_{i}, \theta_{0}))^{2} \times n_{2}^{-1} \sum_{i \in I_{2}} \varepsilon_{i}^{2} \left| \gamma_{k}' P_{k_{2}}(Z_{i}) P_{k_{2}}'(Z_{i}) \gamma_{k} \right|^{2} \right)^{1/2} = O_{p}(\xi_{k_{2}} k_{2}^{-1/2} + \xi_{k_{2}} k_{2}^{-r_{h}}) = o_{p}(1),$$
(189)

which together with (183), (188), (189) and the Cauchy-Schwarz inequality implies that

$$n_{2}^{-1} \sum_{i \in I_{2}} \left(g(X_{i}, \widehat{\theta}_{n}) - \widehat{\phi}(Z_{i}, \widehat{\theta}_{n}) - g(X_{i}, \theta_{0}) - \phi(Z_{i}, \theta_{0})) \varepsilon_{i} \gamma_{k}' P_{k_{2}}(Z_{i}) P_{k_{2}}'(Z_{i}) \gamma_{k} = o_{p}(1)$$
(190)

uniformly over γ_k with $\gamma'_k \gamma_k = 1$. Collecting the results in (178), (179), (182) and (190), we prove the last claim of the theorem.

Proof. [Proof of Theorem 4] By definition,

$$\widehat{\Sigma}_{n_{1}} - \Sigma_{n_{1}} = \frac{\widehat{\phi}_{w\theta,n} P_{n,k_{1}} - \phi_{w\theta,n} P_{n,k_{1}}}{n} \frac{Q_{n_{1},k_{1}}^{-1} \widehat{Q}_{n_{1},u} Q_{n_{1},k_{1}}^{-1}}{n_{1}} \frac{P_{n,k_{1}}' \widehat{\phi}_{w\theta,n}'}{n} \\
+ \frac{\phi_{w\theta,n} P_{n,k_{1}}}{n} Q_{n_{1},k_{1}}^{-1} \frac{\widehat{Q}_{n_{1},u} - Q_{n_{1},u}}{n_{1}} Q_{n_{1},k_{1}}^{-1} \frac{P_{n,k_{1}}' \widehat{\phi}_{w\theta,n}'}{n} \\
+ \frac{\phi_{w\theta,n} P_{n,k_{1}}}{n} Q_{n_{1},k_{1}}^{-1} \frac{Q_{n_{1},u}}{n_{1}} Q_{n_{1},k_{1}}^{-1} \frac{P_{n,k_{1}}' \widehat{\phi}_{w\theta,n}'}{n}.$$
(191)

For any $j = 1, \ldots, d_{\theta}$,

$$\frac{\phi_{w\theta_{j},n}P_{n,k_{1}}}{n} \frac{P'_{n,k_{1}}\phi'_{w\theta_{j},n}}{n} \leq \lambda_{\max}(Q_{n,k_{1}})n^{-1}\phi'_{w\theta_{j},n}\phi_{w\theta_{j},n} \\ \leq \lambda_{\max}(Q_{n,k_{1}})n^{-1}\sum_{i\in I} \left(\psi_{n}^{2}(Z_{i})\phi_{\theta_{j}}^{2}(Z_{i},\theta_{0}) \right) \\ = \lambda_{\max}(Q_{n,k_{1}})E[w_{n}^{2}(Z_{i})\phi_{\theta_{j}}^{2}(Z_{i},\theta_{0})] + o_{p}(1) = O_{p}(1)$$
(192)

where the first equality is by (42) and (66), the second equality is by (42) and (67). Moreover,

$$\frac{\widehat{\phi}_{w\theta_{j,n}}P_{n,k_{1}}}{n}\frac{P_{n,k_{1}}'\widehat{\phi}_{w\theta_{j,n}}'}{n} \leq 2\frac{(\widehat{\phi}_{w\theta_{j,n}} - \phi_{w\theta_{j,n}})P_{n,k_{1}}}{n}\frac{P_{n,k_{1}}'(\widehat{\phi}_{w\theta_{j,n}} - \phi_{w\theta_{j,n}})}{n} + 2\frac{\phi_{w\theta_{j,n}}P_{n,k_{1}}}{n}\frac{P_{n,k_{1}}'\phi_{w\theta_{j,n}}'}{n} = O_{p}(1)$$
(193)

where the equality is by Lemma 9(ii) and (192). By Lemma 9(iv) and (101), we know that

$$C^{-1} \le \lambda_{\min}(\widehat{Q}_{n_1,u}) \le \lambda_{\max}(\widehat{Q}_{n_1,u}) \le C$$
(194)

with probability approaching 1. For any $j_1 = 1, \ldots, d_{\theta}$ and any $j_2 = 1, \ldots, d_{\theta}$, by the Cauchy-Schwarz inequality

$$\begin{vmatrix} \widehat{\phi}_{w\theta_{j_{1},n}}P_{n,k_{1}} - \phi_{w\theta_{j_{1},n}}P_{n,k_{1}}} \frac{Q_{n_{1},k_{1}}^{-1}\widehat{Q}_{n_{1},u}}{n} Q_{n_{1},k_{1}}^{-1}} \frac{P_{n,k_{1}}^{\prime}\widehat{\phi}_{w\theta_{j_{2},n}}}{n} \\ \leq n_{1}^{-1} \left\| n^{-1}(\widehat{\phi}_{w\theta_{j_{1},n}} - \phi_{w\theta_{j_{1},n}})P_{n,k_{1}}) \right\| \sqrt{\frac{\phi_{w\theta_{j_{2},n}}P_{n,k_{1}}}{n}} \left(Q_{n_{1},k_{1}}^{-1}\widehat{Q}_{n_{1},u}} Q_{n_{1},k_{1}}^{-1} \right)^{2} \frac{P_{n,k_{1}}^{\prime}\widehat{\phi}_{w\theta_{j_{2},n}}}{n} \\ \leq \frac{\lambda_{\max}(\widehat{Q}_{n_{1},u})}{n_{1}\lambda_{\min}^{2}\left(Q_{k_{1},n_{1}}\right)} \left\| n^{-1}(\widehat{\phi}_{w\theta_{j_{1},n}} - \phi_{w\theta_{j_{1},n}})P_{n,k_{1}}) \right\| \sqrt{\frac{\phi_{w\theta_{j_{2},n}}P_{n,k_{1}}}{n}} \frac{P_{n,k_{1}}^{\prime}\widehat{\phi}_{w\theta_{j_{2},n}}}{n} = o_{p}(n_{1}^{-1})$$
(195)

where the equality is by (42), (194), Lemma 9(ii) and (193). Similarly for any $j_1 = 1, \ldots, d_{\theta}$ and any $j_2 = 1, \ldots, d_{\theta}$,

$$\begin{vmatrix} \frac{\phi_{w\theta_{j_{1}},n}P_{n,k_{1}}}{n}Q_{n_{1},k_{1}}^{-1}\frac{\widehat{Q}_{n_{1},u}-Q_{n_{1},u}}{n_{1}}Q_{n_{1},k_{1}}^{-1}\frac{P_{n,k_{1}}^{\prime}\widehat{\phi}_{w\theta_{j_{2}},n}}{n} \\ \leq \frac{\sup_{\{\gamma_{k}\in R^{k}:\,\gamma_{k}^{\prime}\gamma_{k}=1\}}\left[\frac{\gamma_{k}^{\prime}(\widehat{Q}_{n_{1},u}-Q_{n_{1},u})\gamma_{k}}{n_{1}}\right]}{n_{1}}\sqrt{\frac{\phi_{w\theta_{j_{1}},n}}{n}}\sqrt{\frac{\phi_{w\theta_{j_{1}},n}}{n}}Q_{n_{1},k_{1}}^{-2}\frac{P_{n,k_{1}}^{\prime}\phi_{w\theta_{j_{1}},n}^{\prime}}{n}}{n}} \\ \leq \frac{\sup_{\{\gamma_{k}\in R^{k}:\,\gamma_{k}^{\prime}\gamma_{k}=1\}}\left[\gamma_{k}^{\prime}(\widehat{Q}_{n_{1},u}-Q_{n_{1},u})\gamma_{k}\right]}{n_{1}\lambda_{\min}^{2}(Q_{n_{1},k_{1}})}\left(\frac{1}{2}\right) \\ \times \sqrt{\frac{\phi_{w\theta_{j_{1}},n}P_{n,k_{1}}}{n}}\frac{P_{n,k_{1}}^{\prime}\phi_{w\theta_{j_{1}},n}^{\prime}}{n}}{\sqrt{\frac{\phi_{w\theta_{j_{2}},n}P_{n,k_{1}}}{n}}}\sqrt{\frac{\phi_{w\theta_{j_{2}},n}P_{n,k_{1}}}{n}}{n}} = o_{p}(n_{1}^{-1})$$
(196)

where the equality is by (42), Lemma 9(iv), (192) and (193). Similarly for any $j_1 = 1, \ldots, d_{\theta}$ and any

$$j_{2} = 1, \dots, d_{\theta},$$

$$\left|\frac{\phi_{w\theta_{j_{1},n}}P_{n,k_{1}}}{n}Q_{n_{1},k_{1}}^{-1}\frac{Q_{n_{1},u}}{n_{1}}Q_{n_{1},k_{1}}^{-1}\frac{P_{n,k_{1}}'\hat{\phi}_{w\theta_{j_{2},n}}' - P_{n,k_{1}}'\phi_{w\theta_{j_{2},n}}'}{n}\right|$$

$$\leq \sqrt{\left|\frac{\phi_{w\theta_{j_{1},n}}P_{n,k_{1}}}{n}Q_{n_{1},k_{1}}^{-1}\frac{Q_{n_{1},u}}{n_{1}}Q_{n_{1},k_{1}}^{-2}\frac{Q_{n_{1},u}}{n_{1}}Q_{n_{1},k_{1}}^{-1}\frac{P_{n,k_{1}}'\phi_{w\theta_{j_{1},n}}'}{n}}{n}\left|\left|n^{-1}P_{k_{1},n}'(\hat{\phi}_{\theta_{j_{2},n}} - \phi_{\theta_{j_{2},n}})\right|\right| \left(\frac{\lambda_{\max}(Q_{n_{1},u})}{n_{1}\lambda_{\min}^{2}(Q_{n_{1},k_{1}})}\sqrt{\left|\frac{\phi_{w\theta_{j_{1},n}}P_{n,k_{1}}}{n}\frac{P_{n,k_{1}}'\phi_{w\theta_{j_{1},n}}'}{n}\right|} \left|\left|n^{-1}P_{k_{1,n}}'(\hat{\phi}_{\theta_{j_{2},n}} - \phi_{\theta_{j_{2},n}})\right|\right| = o_{p}(n_{1}^{-1})$$

$$(197)$$

where the equality is by (42), (101), Lemma 9(ii) and (192). Collecting the results in (191), (195), (196) and (197), we get

$$\widehat{\Sigma}_{n_1} - \Sigma_{n_1} = o_p(n_1^{-1}).$$
(198)

Using similar arguments in proving (198), we can show that

$$\widehat{\Sigma}_{n_2} - \Sigma_{n_2} = o_p(n_2^{-1}). \tag{199}$$

By (112), (113), (198) and (199),

$$\widehat{\Sigma}_{n_1} + \widehat{\Sigma}_{n_2} = (\Sigma_{n_1} + \Sigma_{n_2})(1 + o_p(1)).$$
(200)

By (199), we deduce that

$$H_0^{-1}(\Sigma_{n_1} + \Sigma_{n_2})H_0^{-1}\hat{H}_n(\hat{\Sigma}_{n_1} + \hat{\Sigma}_{n_2})^{-1}\hat{H}_n$$

$$= H_0^{-1}(\Sigma_{n_1} + \Sigma_{n_2})H_0^{-1}\hat{H}_n(\Sigma_{n_1} + \Sigma_{n_2})^{-1}\hat{H}_n(1 + o_p(1))$$

$$= H_0^{-1}(\Sigma_{n_1} + \Sigma_{n_2})H_0^{-1}(\hat{H}_n - H_0)(\Sigma_{n_1} + \Sigma_{n_2})^{-1}\hat{H}_n(1 + o_p(1))$$

$$+ H_0^{-1}(\hat{H}_n - H_0)(1 + o_p(1)) + I_{d_{\theta}}(1 + o_p(1))$$

$$= I_{d_{\theta}} + H_0^{-1}(\Sigma_{n_1} + \Sigma_{n_2})H_0^{-1}(\hat{H}_n - H_0)(\Sigma_{n_1} + \Sigma_{n_2})^{-1}H_0(1 + o_p(1)) + o_p(1)$$
(201)

where the last equality is by Lemma 9(i). Using (42), (101), (102) and Lemma 6, we have

$$\lambda_{\max}(n_1 \Sigma_{n_1}) \le C \text{ and } \lambda_{\max}(n_2 \Sigma_{n_2}) \le C$$
(202)

with probability approaching 1. For any $\gamma_1, \gamma_2 \in \mathbb{R}^{d_\theta}$ with $\gamma'_1 \gamma_1 = 1$ and $\gamma'_2 \gamma_2 = 1$, by the Cauchy-Schwarz inequality we have

$$\begin{aligned} \left| \gamma_{1}^{\prime} \Sigma_{n_{1}} H_{0}^{-1} (\widehat{H}_{n} - H_{0}) (\Sigma_{n_{1}} + \Sigma_{n_{2}})^{-1} H_{0} \gamma_{2} \right| \\ &\leq \left(\gamma_{1}^{\prime} (n_{1} \Sigma_{n_{1}}) H_{0}^{-1} (\widehat{H}_{n} - H_{0})^{2} H_{0}^{-1} (n_{1} \Sigma_{n_{1}}) \right) \left(1 \right)^{1/2} (\gamma_{2}^{\prime} H_{0} (n_{1} \Sigma_{n_{1}} + n_{1} \Sigma_{n_{2}})^{-2} H_{0} \gamma_{2})^{1/2} \\ &\leq \frac{C \lambda_{\max} (n_{1} \Sigma_{n_{1}}) || \widehat{H}_{n} - H_{0} ||}{\lambda_{\min} (H_{0}) \lambda_{\min} (n_{1} \Sigma_{n_{1}} + n_{1} \Sigma_{n_{2}})} = o_{p}(1) \end{aligned}$$

$$(203)$$

where the equality is by Assumptions 3(ii), Lemma 9(i), (112) and (202). Similarly, we can show that

$$\left|\gamma_{1}'\Sigma_{n_{2}}H_{0}^{-1}(\widehat{H}_{n}-H_{0})(\Sigma_{n_{1}}+\Sigma_{n_{2}})^{-1}H_{0}\gamma_{2}\right|=o_{p}(1).$$
(204)

Let γ'_1 be any row of H_0^{-1} and γ_2 be any column of H_0 . Then we can use (203) and (204) to deduce that

$$H_0^{-1}(\Sigma_{n_1} + \Sigma_{n_2})H_0^{-1}(\widehat{H}_n - H_0)(\Sigma_{n_1} + \Sigma_{n_2})^{-1}H_0 = o_p(1)$$
(205)

which together with (201) implies that

$$\widehat{H}_n(\widehat{\Sigma}_{n_1} + \widehat{\Sigma}_{n_2})^{-1}\widehat{H}_n = (H_0(\Sigma_{n_1} + \Sigma_{n_2})^{-1}H_0)(I_{d_\theta} + o_p(1)).$$
(206)

This shows (19). Using (19) and Theorem 2, and then applying CMT, we immediately prove the claim of the theorem. $\hfill \Box$

Proof. [Proof of Lemma 3] By definition,

$$\widehat{u}_i = Y_i - \widehat{h}_{n_1}(Z_i) = u_i + h_0(Z_i) - \widehat{h}_{n_1}(Z_i)$$
(207)

which implies that

$$\widehat{u}_i^2 = u_i^2 + (h_0(Z_i) - \widehat{h}_{n_1}(Z_i))^2 + 2u_i(h_0(Z_i) - \widehat{h}_{n_1}(Z_i)).$$
(208)

Hence,

$$\widehat{\sigma}_{n,u}^{2}(z) = n_{1}^{-1} P_{k_{1}}'(z) Q_{n_{1},k_{1}}^{-1} \sum_{i \in I_{1}} \mu_{i}^{2} P_{k_{1}}(Z_{i}) + n_{1}^{-1} P_{k_{1}}'(z) Q_{n_{1},k_{1}}^{-1} \sum_{i \in I_{1}} (h_{0}(Z_{i}) - \widehat{h}_{n_{1}}(Z_{i}))^{2} P_{k_{1}}(Z_{i}) + 2n_{1}^{-1} P_{k_{1}}'(z) Q_{n_{1},k_{1}}^{-1} \sum_{i \in I_{1}} \mu_{i}(h_{0}(Z_{i}) - \widehat{h}_{n_{1}}(Z_{i})) P_{k_{1}}(Z_{i}).$$

$$(209)$$

Let $\widehat{\beta}_{u,n_1} = n_1^{-1} Q_{k,n_1}^{-1} \sum_{i \in I_1} u_i^2 P_k(Z_i)$. By Assumptions 1(i), 1(iii), 1(v), 3(v) and 4(i), we can use similar arguments in showing (43) to deduce that

$$\left\|\widehat{\beta}_{u,n_{1}} - \beta_{u,k}\right\|^{2} = O_{p}(k_{1}n_{1}^{-1} + k_{1}^{-2r_{u}}).$$
(210)

By (210) and Assumptions 4(ii),

$$\sup_{z \in \mathcal{Z}} \left| P_{k_1}'(z) \widehat{\beta}_{u,n_1} - \sigma_u^2(z) \right| = O_p(\xi_{k_1} k_1^{1/2} n_1^{-1/2} + \xi_{k_1} k_1^{-r_u}).$$
(211)

By the triangle inequality, the Cauchy-Schwarz inequality, (42) and (44),

$$\begin{vmatrix} n_{1}^{-1}P_{k_{1}}'(z)Q_{n_{1},k_{1}}^{-1}\sum_{i\in I_{1}} \left(h_{0}(Z_{i})-\hat{h}_{n_{1}}(Z_{i})\right)^{2}P_{k}(Z_{i}) \\ \leq \left(\sup_{z\in\mathcal{Z}}(P_{k_{1}}'(z)Q_{n_{1},k_{1}}^{-1}P_{k_{1}}(z))^{2}n_{1}^{-1}\sum_{i\in I_{1}} \left(h_{0}(Z_{i})-\hat{h}_{n}\left(Z_{i}\right)\right)^{2} \\ = O_{p}(\xi_{k_{1}}^{2}k_{1}n_{1}^{-1}+\xi_{k_{1}}^{2}k_{1}^{-2r_{h}}). \end{aligned}$$

$$(212)$$

By definition,

$$n_{1}^{-1}P_{k_{1}}'(z)Q_{n_{1},k_{1}}^{-1}\sum_{i\in I_{1}}\mu_{i}(h_{0}(Z_{i})-\hat{h}_{n_{1}}(Z_{i}))P_{k}(Z_{i})$$

$$=n_{1}^{-1}P_{k_{1}}'(z)Q_{n_{1},k_{1}}^{-1}\sum_{i\in I_{1}}\mu_{i}(h_{0}(Z_{i})-h_{0,k_{1}}(Z_{i}))P_{k}(Z_{i})$$

$$+n_{1}^{-1}P_{k_{1}}'(z)Q_{n_{1},k_{1}}^{-1}\sum_{i\in I_{1}}\mu_{i}P_{k}(Z_{i})P_{k}'(Z_{i})(\hat{\beta}_{k_{1},n_{1}}-\beta_{h,k_{1}}).$$
(213)

By Assumptions 1(i), 1(iii), 1(iv) and 3(v),

$$E\left[\left\| n_{1}^{-1} \sum_{i \in I_{1}} \left(h_{0}(Z_{i}) - h_{0,k_{1}}(Z_{i}) \right) P_{k_{1}}(Z_{i}) \right\|^{2} \right]$$

= $n_{1}^{-1} E\left[u^{2}(h_{0}(Z) - h_{0,k_{1}}(Z))^{2} P_{k_{1}}(Z)' P_{k_{1}}(Z) \right]$
 $\leq C n_{1}^{-1} \sup_{z \in \mathcal{Z}} |h_{0}(z) - h_{0,k_{1}}(z)|^{2} E\left[P_{k_{1}}(Z)' P_{k_{1}}(Z) \right] = O(k_{1}^{1-2r_{h}} n_{1}^{-1})$ (214)

which together with the Markov inequality implies that

$$n_1^{-1} \sum_{i \in I_1} \left(h_0(Z_i) - h_{0,k_1}(Z_i)) P_{k_1}(Z_i) = O_p(k_1^{1/2 - r_h} n_1^{-1/2}).$$
(215)

By (42), (44), (215) and the Cauchy-Schwarz inequality,

$$\sup_{z \in \mathcal{Z}} \left| n_{1}^{-1} P_{k_{1}}'(z) Q_{n_{1},k_{1}}^{-1} \sum_{i \in I_{1}} \left(\mu_{i}(h_{0}(Z_{i}) - h_{0,k_{1}}(Z_{i})) P_{k}(Z_{i}) \right) \right| \\
\leq \sup_{z \in \mathcal{Z}} \left(P_{k_{1}}'(z) Q_{n_{1},k_{1}}^{-2} P_{k_{1}}(z) \right)^{1/2} \left\| n_{1}^{-1} \sum_{i \in I_{1}} \left(\mu_{i}(h_{0}(Z_{i}) - h_{0,k_{1}}(Z_{i})) P_{k_{1}}(Z_{i}) \right) \right\| \\
= O_{p}(\xi_{k_{1}} n_{1}^{-1/2} k_{1}^{1/2 - r_{h}}).$$
(216)

By Assumptions 1(i), 1(iii) and 3(vi),

$$E\left[\left\| \left(n_{1}^{-1} \sum_{i \in I_{1}} u_{i} P_{k}(Z_{i}) P_{k}'(Z_{i}) \right\|^{2} \right] = \frac{E\left[u^{2} (P_{k_{1}}(Z)' P_{k_{1}}(Z))^{2} \right]}{\left(n_{1} \right)} \le \frac{C\xi_{k_{1}}^{2} E\left[P_{k_{1}}(Z)' P_{k_{1}}(Z) \right]}{n_{1}} = C\xi_{k_{1}}^{2} k_{1} n_{1}^{-1}$$

which together with the Markov inequality implies that

$$\left\| n_{1}^{-1} \sum_{i \in I_{1}} \left\| u_{i} P_{k}(Z_{i}) P_{k}'(Z_{i}) \right\| = O_{p}(\xi_{k_{1}} k_{1}^{1/2} n_{1}^{-1/2}).$$

$$(217)$$

By (42), (43), (217) and the Cauchy-Schwarz inequality,

$$\begin{vmatrix} n_{1}^{-1} P_{k_{1}}'(z) Q_{n_{1},k_{1}}^{-1} \sum_{i \in I_{1}} \left(\iota_{i} P_{k}(Z_{i}) P_{k}'(Z_{i}) (\widehat{\beta}_{k_{1},n_{1}} - \beta_{h,k_{1}}) \right) \\ \leq \sup_{z \in \mathbb{Z}} (P_{k_{1}}'(z) Q_{n_{1},k_{1}}^{-2} P_{k_{1}}(z))^{1/2} \\ = O_{p}(\xi_{k_{1}}^{2} k_{1}^{1/2} n_{1}^{-1/2} (k_{1}^{1/2} n_{1}^{-1/2} + \left(k_{1}^{-r_{h}}) \right). \end{aligned}$$

$$(218)$$

Combining the results in (213), (216) and (218), we get

$$\sup_{z \in \mathcal{Z}} \left| n_1^{-1} P'_{k_1}(z) Q_{n_1, k_1}^{-1} \sum_{i \in I_1} \left(\mu_i(h_0(Z_i) - \hat{h}_{n_1}(Z_i)) P_k(Z_i) \right) \right|_{\mathcal{L}} = O_p(\xi_{k_1}^2 k_1 n_1^{-1} + \xi_{k_1}^2 k_1^{-2r_h}), \quad (219)$$
her with (209), (211) and (212) implies that

which together with (209), (211) and (212) implies that

$$\sup_{z \in \mathcal{Z}} \left| \widehat{\sigma}_{n,u}^2(z) - \sigma_u^2(z) \right| = O_p(\xi_{k_1}(k_1^{1/2}n_1^{-1/2} + k_1^{-r_u}) + \xi_{k_1}^2(k_1n_1^{-1} + k_1^{-2r_h})).$$
(220)

This together with Assumptions 4(iv) proves the first claim of the lemma.

By definition,

$$\widetilde{\varepsilon}_i = g(X_i, \widehat{\theta}_{1,n}) - \widehat{\phi}_{n_2}(Z_i, \widehat{\theta}_{1,n}) = \varepsilon_i + (g(X_i, \widehat{\theta}_{1,n}) - g(X_i, \theta_0)) - (\widehat{\phi}_{n_2}(Z_i, \widehat{\theta}_{1,n}) - \phi(Z_i, \theta_0)).$$

Hence,

$$\widehat{\sigma}_{n,\varepsilon}^{2}(z) = n_{2}^{-1} P_{k_{2}}'(z) Q_{n_{2},k_{2}}^{-1} \sum_{i \in I_{2}} \left\{ \widehat{P}_{k_{2}}(Z_{i}) + n_{2}^{-1} P_{k_{2}}'(z) Q_{n_{2},k_{2}}^{-1} \sum_{i \in I_{2}} (g(X_{i},\widehat{\theta}_{1,n}) - g(X_{i},\theta_{0}))^{2} P_{k_{2}}(Z_{i}) \\
+ n_{2}^{-1} P_{k_{2}}'(z) Q_{n_{2},k_{2}}^{-1} \sum_{i \in I_{2}} \left\{ \widehat{\phi}_{n_{2}}(Z_{i},\widehat{\theta}_{1,n}) - \phi(Z_{i},\theta_{0}) \right)^{2} P_{k_{2}}(Z_{i}) \\
+ 2n_{2}^{-1} P_{k_{2}}'(z) Q_{n_{2},k_{2}}^{-1} \sum_{i \in I_{2}} \left\{ i(g(X_{i},\widehat{\theta}_{1,n}) - g(X_{i},\theta_{0})) P_{k_{2}}(Z_{i}) \\
- 2n_{2}^{-1} P_{k_{2}}'(z) Q_{n_{2},k_{2}}^{-1} \sum_{i \in I_{2}} \left\{ i(\widehat{\phi}_{n_{2}}(Z_{i},\widehat{\theta}_{1,n}) - \phi(Z_{i},\theta_{0})) P_{k_{2}}(Z_{i}) \\
- 2n_{2}^{-1} P_{k_{2}}'(z) Q_{n_{2},k_{2}}^{-1} \sum_{i \in I_{2}} \left\{ g(X_{i},\widehat{\theta}_{1,n}) - g(X_{i},\theta_{0}) \right) (\widehat{\phi}_{n_{2}}(Z_{i},\widehat{\theta}_{1,n}) - \phi(Z_{i},\theta_{0})) P_{k_{2}}(Z_{i}) \\
- 2n_{2}^{-1} P_{k_{2}}'(z) Q_{n_{2},k_{2}}^{-1} \sum_{i \in I_{2}} \left\{ g(X_{i},\widehat{\theta}_{1,n}) - g(X_{i},\theta_{0}) \right) (\widehat{\phi}_{n_{2}}(Z_{i},\widehat{\theta}_{1,n}) - \phi(Z_{i},\theta_{0})) P_{k_{2}}(Z_{i}). \\$$
(221)

Using similar arguments in showing (211), we can show that

$$\sup_{z \in \mathcal{Z}} \left| P_{k_2}'(z) \widehat{\beta}_{\varepsilon, n_2} - \sigma_{\varepsilon}^2(z) \right| = O_p(\xi_{k_2} k_2^{1/2} n_2^{-1/2} + \xi_{k_2} k_2^{-r_{\varepsilon}}).$$
(222)

By the Cauchy-Schwarz inequality, (42) and (154)

.

$$\begin{vmatrix} n_{2}^{-1}P_{k_{2}}'(z)Q_{n_{2},k_{2}}^{-1}\sum_{i\in I_{2}} \left(g(X_{i},\widehat{\theta}_{1,n}) - g(X_{i},\theta_{0}))^{2}P_{k_{2}}(Z_{i}) \right| \\ \leq \sup_{z\in\mathcal{Z}} (P_{k_{2}}'(z)Q_{n_{2},k_{2}}^{-1}P_{k_{2}}(z))n_{2}^{-1}\sum_{i\in I_{2}} \left(g(X_{i},\widehat{\theta}_{1,n}) - g(X_{i}, \theta_{0}))^{2} \right) \\ = O_{p}(\xi_{k_{2}}^{2}n_{1}^{-1} + \xi_{k_{2}}^{2}n_{2}^{-1}).$$

$$(223)$$

Similarly, by the Cauchy-Schwarz inequality, (42) and (157)

$$\begin{vmatrix} n_{2}^{-1} P_{k_{2}}'(z) Q_{n_{2},k_{2}}^{-1} \sum_{i \in I_{2}} \left(\widehat{\phi}_{n_{2}}(Z_{i},\widehat{\theta}_{1,n}) - \phi(Z_{i},\theta_{0}) \right)^{2} P_{k_{2}}(Z_{i}) \\ \left(\sup_{z \in \mathcal{Z}} (P_{k_{2}}'(z) Q_{n_{2},k_{2}}^{-1} P_{k_{2}}(z)) n_{2}^{-1} \sum_{i \in I_{2}} \left(\widehat{\phi}_{n_{2}}(Z_{i},\widehat{\theta}_{1,n}) - \phi(Z_{i},\theta_{0}) \right)^{2} \\ = O_{p}(\xi_{k_{2}}^{2} n_{1}^{-1} + \xi_{k_{2}}^{2} k_{2} n_{2}^{-1} + \xi_{k_{2}}^{2} k_{2}^{-2r_{h}}). \end{aligned}$$
(224)

By the Cauchy-Schwarz inequality, (42), (154) and (157)

$$\left| n_{2}^{-1} P_{k_{2}}'(z) Q_{n_{2},k_{2}}^{-1} \sum_{i \in I_{2}} \left(g(Z_{i},\widehat{\theta}_{1,n}) - g(Z_{i},\theta_{0}))(\widehat{\phi}_{n_{2}}(Z_{i},\widehat{\theta}_{1,n}) - \phi(Z_{i},\theta_{0})) P_{k_{2}}(Z_{i}) \right)^{2} \\ \leq \sup_{z \in \mathcal{Z}} (P_{k_{2}}'(z) Q_{n_{2},k_{2}}^{-1} P_{k_{2}}(z))^{2} n_{2}^{-1} \sum_{i \in I_{2}} \left(g(Z_{i},\widehat{\theta}_{1,n}) - g(Z_{i},\theta_{0}))^{2} \right)^{2} \\ \times n_{2}^{-1} \sum_{i \in I_{2}} \left(\widehat{\phi}_{n_{2}}(Z_{i},\widehat{\theta}_{1,n}) - \phi(Z_{i},\theta_{0}))^{2} \right)^{2} \\ = O_{p}(\xi_{k_{2}}^{4}(n_{1}^{-1} + n_{2}^{-1})(n_{1}^{-1} + k_{2}n_{2}^{-1} + k_{2}^{-2r_{h}})).$$
(225)

By the second order expansion,

$$g(X_i,\widehat{\theta}_{1,n}) - g(X_i,\theta_0) = g_{\theta}(X_i,\theta_0)'(\widehat{\theta}_{1,n} - \theta_0) + 2^{-1}(\widehat{\theta}_{1,n} - \theta_0)'g_{\theta\theta}(X_i,\widetilde{\theta}_{i,n})(\widehat{\theta}_{1,n} - \theta_0)$$
(226)

where $\hat{\theta}_{i,n}$ is between $\hat{\theta}_n$ and θ_0 . By Assumption 1(i), 3(v) and 4(v)

$$E\left[\left\|n_{2}^{-1}\sum_{i\in I_{2}}\left(P_{k_{2}}(Z_{i})g_{\theta}(X_{i},\theta_{0})'\right\|^{2}\right] = n_{2}^{-1}E\left[\varepsilon_{i}^{2}\left\|P_{k_{2}}(Z_{i})g_{\theta}(X_{i},\theta_{0})'\right\|^{2}\right] \\ \leq n_{2}^{-1}\xi_{k_{2}}^{2}E\left[\left|q_{i}^{2}\right|\left|g_{\theta}(X_{i},\theta_{0})\right|^{2}\right] \\ \leq n_{2}^{-1}\xi_{k_{2}}^{2}\sqrt{E\left[\varepsilon_{i}^{4}\right]E\left[\left|g_{\theta}(X_{i},\theta_{0})\right|^{4}\right]} = O(\xi_{k_{2}}^{2}n_{2}^{-1}) \quad (227)$$

which together with the Markov inequality implies that

$$n_2^{-1} \sum_{i \in I_2} \left(i P_{k_2}(Z_i) g_{\theta}(X_i, \theta_0)' = O_p(\xi_{k_2} n_2^{-1/2}). \right)$$
(228)

By (24), (42), the triangle inequality and the Cauchy-Schwarz inequality,

.

$$\sup_{z \in \mathcal{Z}} \left| n_{2}^{-1} P_{k_{2}}'(z) Q_{n_{2},k_{2}}^{-1} \sum_{i \in I_{2}} \left\{ i P_{k_{2}}(Z_{i}) g_{\theta}(X_{i},\theta_{0})'(\widehat{\theta}_{1,n}-\theta_{0}) \right| \\
\leq \sup_{z \in \mathcal{Z}} (P_{k_{2}}'(z) Q_{n_{2},k_{2}}^{-2} P_{k_{2}}(z))^{1/2} \left\| n_{2}^{-1} \sum_{i \in I_{2}} \left\{ i P_{k_{2}}(Z_{i}) g_{\theta}(X_{i},\theta_{0})' \right\| \left\| \widehat{\theta}_{1,n} - \theta_{0} \right\| \\
= O_{p}(\xi_{k_{2}}^{2} n_{2}^{-1/2} (n_{1}^{-1/2} + n_{2}^{-1/2})) \left\| \left\| e_{k_{2}}^{-1} \sum_{i \in I_{2}} \left\{ i P_{k_{2}}(Z_{i}) g_{\theta}(X_{i},\theta_{0})' \right\| \right\| \right\| \right\} \left\| e_{k_{2}}^{-1} \sum_{i \in I_{2}} \left\{ i P_{k_{2}}(Z_{i}) g_{\theta}(X_{i},\theta_{0})' \right\| \right\| e_{k_{2}}^{-1} \left\| e_{k_{2}}^{-1} e_{k_{2}$$

By Assumptions 1(i), 3(i) and 3(v), and the Cauchy-Schwarz inequality,

$$\max_{i \in I_2} \left\| n_2^{-1} \sum_{i \in I_2} |\varepsilon_i| g_{\theta\theta}(X_i, \tilde{\theta}_{i,n}) \right\|_{\ell}^2 \le n_2^{-1} \sum_{i \in I_2} \left\| \varepsilon_i^2 \times \sup_{\theta \in \mathcal{N}_n} n_2^{-1} \sum_{i \in I_2} \|g_{\theta\theta}(X_i, \theta)\|^2 = O_p(1).$$
(230)

By (24), the triangle inequality and the Cauchy-Schwarz inequality,

$$\sup_{z \in \mathcal{Z}} \left| n_{2}^{-1} P_{k_{2}}'(z) Q_{n_{2},k_{2}}^{-1} \sum_{i \in I_{2}} \left\{ P_{k_{2}}(Z_{i})(\widehat{\theta}_{1,n} - \theta_{0})' g_{\theta\theta}(X_{i},\widetilde{\theta}_{i,n})(\widehat{\theta}_{1,n} - \theta_{0}) \right| \\
\leq \sup_{z \in \mathcal{Z}} \left(P_{k_{2}}'(z) Q_{n_{2},k_{2}}^{-1} P_{k_{2}}(z)) \max_{i \in I_{2}} \left\| n_{2}^{-1} \sum_{i \in I_{2}} |\varepsilon_{i}| g_{\theta\theta}(X_{i},\widetilde{\theta}_{i,n}) \right\| \left\| \widehat{\theta}_{1,n} - \theta_{0} \right\|_{\ell}^{2} \left(\left(231 \right) \right) \right) \\
= O_{p}(\xi_{k_{2}}^{2}(n_{1}^{-1} + n_{2}^{-1})).$$

Combining the results in (226), (229) and (221), we get

$$\sup_{z \in \mathcal{Z}} \left| n_2^{-1} P'_{k_2}(z) Q_{n_2, k_2}^{-1} \sum_{i \in I_2} \left| f_i(g(X_i, \hat{\theta}_{1,n}) - g(X_i, \theta_0)) P_{k_2}(Z_i) \right| = O_p(\xi_{k_2}^2(n_1^{-1} + n_2^{-1})).$$
(232)
(55), Assumptions 1(i), 1(iv) and 3(v),

By (42), (155), Assumptions 1(i), 1(iv) and 3(v),

$$E\left[\left\|n_{2}^{-1}\sum_{i\in I_{2}}\left(\hat{\phi}_{n_{2}}(Z_{i},\theta_{0})-\phi(Z_{i},\theta_{0})\right)P_{k_{2}}(Z_{i})\right\|^{2}\right|\{Z_{i}\}_{i\in I_{2}}\right]$$

$$=n_{2}^{-2}\sum_{i\in I_{2}}\left(p_{\varepsilon}^{2}(Z_{i})(\hat{\phi}_{n_{2}}(Z_{i},\theta_{0})-\phi(Z_{i},\theta_{0}))^{2}P_{k_{2}}(Z_{i})'R_{k_{2}}(Z_{i})\right)$$

$$\leq C\xi_{k_{2}}^{2}n_{2}^{-2}\sum_{i\in I_{2}}\left(p_{n_{2}}(Z_{i},\theta_{0})-\phi(Z_{i},\theta_{0}))^{2}=O_{p}(\xi_{k_{2}}^{2}n_{2}^{-1}(k_{2}n_{2}^{-1}+k_{2}^{-2r_{h}})),$$
(233)

which together with (42), the Markov inequality, the triangle inequality and the Cauchy-Schwarz inequality implies that

$$\sup_{z \in \mathcal{Z}} \left| n_{2}^{-1} P_{k_{2}}'(z) Q_{n_{2},k_{2}}^{-1} \sum_{i \in I_{2}} \left\{ i(\widehat{\phi}_{n_{2}}(Z_{i},\theta_{0}) - \phi(Z_{i},\theta_{0})) P_{k_{2}}(Z_{i}) \right| \\
\leq \sup_{z \in \mathcal{Z}} (P_{k_{2}}'(z) Q_{n_{2},k_{2}}^{-2} P_{k_{2}}(z))^{1/2} \left\| n_{2}^{-1} \sum_{i \in I_{2}} \left\{ i(\widehat{\phi}_{n_{2}}(Z_{i},\theta_{0}) - \phi(Z_{i},\theta_{0})) P_{k_{2}}(Z_{i}) \right\| \\
= O_{p}(\xi_{k_{2}}^{2} n_{2}^{-1/2} (k_{2}^{1/2} n_{2}^{-1/2} + k_{2}^{-r} \left\{ j \right\}).$$
(234)

By Assumptions 1(i) and 1(iii),

$$E\left[\left\|n_{2}^{-1}\sum_{i\in I_{2}}\varepsilon_{i}P_{k_{2}}(Z_{i})P_{k_{2}}(Z_{i})'\right\|^{2}\right] = n_{2}^{-1}E\left[\sigma_{\varepsilon}^{2}(Z_{i})|P_{k_{2}}(Z_{i})'P_{k_{2}}(Z_{i})|^{2}\right] = O(\xi_{k_{2}}^{2}k_{2}n_{2}^{-1}), \quad (235)$$

which together with the Markov inequality implies that

$$\left\| n_2^{-1} \sum_{i \in I_2} \left\{ i P_{k_2}(Z_i) P_{k_2}(Z_i)' \right\|_{\ell} = O_p(\xi_{k_2} k_2^{1/2} n_2^{-1/2}).$$
(236)

Recall that $g_{n_2}(\hat{\theta}_{1,n}) = (g(X_i, \hat{\theta}_{1,n}))'_{i \in I_2}$ and $g_{n_2}(\theta_0) = (g(X_i, \theta_0))'_{i \in I_2}$. By (42), (154), (236), the triangle inequality and the Cauchy-Schwarz inequality

$$\sup_{z \in \mathcal{Z}} \left| n_{2}^{-1} P_{k_{2}}'(z) Q_{n_{2},k_{2}}^{-1} \sum_{i \in I_{2}} \left\{ i(\widehat{\phi}_{n_{2}}(Z_{i},\widehat{\theta}_{1,n}) - \widehat{\phi}_{n_{2}}(Z_{i},\theta_{0})) P_{k_{2}}(Z_{i}) \right| \left(\left(\widehat{\phi}_{n_{2}}(Z_{i},\widehat{\theta}_{1,n}) - \widehat{\phi}_{n_{2}}(Z_{i},\theta_{0}) \right) P_{k_{2}}(Z_{i}) \right) \right| \\
= \left| n_{2}^{-1} P_{k_{2}}'(z) Q_{n_{2},k_{2}}^{-1} \sum_{i \in I_{2}} \left\{ iP_{k_{2}}(Z_{i}) P_{k_{2}}(Z_{i})' (P_{n_{2},k_{2}}' P_{n_{2},k_{2}})^{-1} P_{n_{2},k_{2}}'(g_{n_{2}}(\widehat{\theta}_{1,n}) - g_{n_{2}}(\theta_{0})) \right) \right| \\
\leq \sup_{z \in \mathcal{Z}} \left(P_{k_{2}}'(z) Q_{n_{2},k_{2}}^{-2} P_{k_{2}}(z) \right)^{1/2} \left\| n_{2}^{-1} \sum_{i \in I_{2}} \left\{ iP_{k_{2}}(Z_{i}) P_{k_{2}}(Z_{i})' \right\| \\
\times \sqrt{(g_{n_{2}}(\widehat{\theta}_{1,n}) - g_{n_{2}}(\theta_{0}))' P_{n_{2},k_{2}}(P_{n_{2},k_{2}}' P_{n_{2},k_{2}})^{-2} P_{n_{2},k_{2}}'(g_{n_{2}}(\widehat{\theta}_{1,n}) - g_{n_{2}}(\theta_{0}))} \right) \\
= O_{p}(\xi_{k_{2}}^{2} k_{2}^{1/2} n_{2}^{-1/2}) \sqrt{\left(\sum_{i \in I_{2}} \left(g(X_{i},\widehat{\theta}_{1,n}) - g(X_{i},\theta_{0}) \right)^{2} \right) \right)} \right| (237)$$

Combining the results in (234) and (237), we get

$$\sup_{z \in \mathcal{Z}} \left| n_2^{-1} P_{k_2}'(z) Q_{n_2,k_2}^{-1} \sum_{i \in I_2} \left\{ i(\widehat{\phi}_{n_2}(Z_i, \widehat{\theta}_{1,n}) - \phi(Z_i, \theta_0)) P_{k_2}(Z_i) \right|$$

$$= O_p(\xi_{k_2}^4 k_2^{1/2} n_2^{-1} + \xi_{k_2}^2 n_2^{-1/2} k_2^{-r_h} + \xi_{k_2}^2 k_2^{1/2} n_2^{-1/2} n_1^{-1/2}).$$
(238)

Combining the results in (221), (222), (223), (224), (225), (232) and (238), and then applying Assumption 4(iv), we get

$$\sup_{z\in\mathcal{Z}} \left| \widehat{\sigma}_{n,\varepsilon}^2(z) - \sigma_{\varepsilon}^2(z) \right| = O_p(\xi_{k_2}^2 n_1^{-1} + \xi_{k_2}^2 k_2^{-2r_h} + \xi_{k_2} k_2^{1/2} n_2^{-1/2} + \xi_{k_2} k_2^{-r_{\varepsilon}}).$$

Proof. [Proof of Theorem 5] By the triangle inequality,

$$\begin{aligned} |\widehat{w}_{n}^{*}(z) - w^{*}(z)| &= (n_{1}^{-1} + n_{2}^{-1}) \left| (n_{1}^{-1}\widehat{\sigma}_{n,u}^{2}(z) + n_{2}^{-1}\widehat{\sigma}_{n,\varepsilon}^{2}(z))^{-1} - (n_{1}^{-1}\sigma_{u}^{2}(z) + n_{2}^{-1}\sigma_{\varepsilon}^{2}(z)) \right| \\ &\leq (n_{1}^{-1} + n_{2}^{-1}) \frac{n_{1}^{-1} \left| \widehat{\sigma}_{n,u}^{2}(z) - \sigma_{u}^{2}(z) \right| + n_{2}^{-1} \left| \widehat{\sigma}_{n,\varepsilon}^{2}(z) - \sigma_{\varepsilon}^{2}(z) \right|}{\left| (n_{1}^{-1}\widehat{\sigma}_{n,u}^{2}(z) + n_{2}^{-1}\widehat{\sigma}_{n,\varepsilon}^{2}(z))(n_{1}^{-1}\sigma_{u}^{2}(z) + n_{2}^{-1}\sigma_{\varepsilon}^{2}(z)) \right| \\ &\leq (n_{1}^{-1} + n_{2}^{-1})^{2} \frac{\left| \widehat{\sigma}_{n,u}^{2}(z) - \sigma_{u}^{2}(z) \right| + \left| \widehat{\sigma}_{n,\varepsilon}^{2}(z) - \sigma_{\varepsilon}^{2}(z) \right|}{\left| (n_{1}^{-1}\widehat{\sigma}_{n,u}^{2}(z) + n_{2}^{-1}\widehat{\sigma}_{n,\varepsilon}^{2}(z))(n_{1}^{-1}\sigma_{u}^{2}(z) + n_{2}^{-1}\sigma_{\varepsilon}^{2}(z)) \right|}. \end{aligned}$$

$$(239)$$

By Lemma 3 and the triangle inequality

$$\sup_{z \in \mathcal{Z}} \left| n_1^{-1} \widehat{\sigma}_{n,u}^2(z) + n_2^{-1} \widehat{\sigma}_{n,\varepsilon}^2(z) - (n_1^{-1} \sigma_u^2(z) + n_2^{-1} \sigma_{\varepsilon}^2(z)) \right| \\
\leq n_1^{-1} \sup_{z \in \mathcal{Z}} \left| \widehat{\sigma}_{n,u}^2(z) - \sigma_u^2(z) \right| + n_2^{-1} \sup_{z \in \mathcal{Z}} \left| \widehat{\sigma}_{n,\varepsilon}^2(z) - \sigma_{\varepsilon}^2(z) \right| \\
= O_p(\delta_{w,n}(n_1^{-1} + n_2^{-1})).$$
(240)

By Assumption 3(v),

$$n_1^{-1}\sigma_u^2(z) + n_2^{-1}\sigma_\varepsilon^2(z) \ge (n_1^{-1} + n_2^{-1})C^{-1}$$
(241)

for any $z \in \mathcal{Z}$, which together with (240) and $\delta_{w,n} = o(1)$ implies that

$$n_1^{-1}\widehat{\sigma}_{n,u}^2(z) + n_2^{-1}\widehat{\sigma}_{n,\varepsilon}^2(z) = (n_1^{-1}\sigma_u^2(z) + n_2^{-1}\sigma_\varepsilon^2(z))(1 + o_p(1))$$
(242)

uniformly over $z \in \mathbb{Z}$. Combining Lemma 3, the results in (239), (240) and (242), we have

$$\sup_{z \in \mathcal{Z}} |\widehat{w}_{n}^{*}(z) - w^{*}(z)| \leq (n_{1}^{-1} + n_{2}^{-1})^{2} \frac{\sup_{z \in \mathcal{Z}} |\widehat{\sigma}_{n,u}^{2}(z) - \sigma_{u}^{2}(z)| + \sup_{z \in \mathcal{Z}} |\widehat{\sigma}_{n,\varepsilon}^{2}(z) - \sigma_{\varepsilon}^{2}(z)|}{\inf_{z \in \mathcal{Z}} (n_{1}^{-1} \sigma_{u}^{2}(z) + n_{2}^{-1} \sigma_{\varepsilon}^{2}(z))^{2} (1 + o_{p}(1))} \\
\leq \sup_{z \in \mathcal{Z}} |\widehat{\sigma}_{n,u}^{2}(z) - \sigma_{u}^{2}(z)| + \sup_{z \in \mathcal{Z}} |\widehat{\sigma}_{n,\varepsilon}^{2}(z) - \sigma_{\varepsilon}^{2}(z)| = O_{p}(\delta_{w,n}) \tag{243}$$

which finishes the proof.

D Low-level Sufficient Conditions

In this section, we provide low-level sufficient conditions for Assumptions 2(i), 2(ii), 2(iv), 3(i), 3(iii)-(iv), 4(i) and 4(v).

Assumption 5. (i) For any θ , there exist $\beta_{\theta,k} \in \mathbb{R}^k$ and $r_{\varphi} > 0$, such that

$$\sup_{z \in \mathcal{Z}} |\varphi(z, \theta) - P_k(z)' \beta_{\theta, k}| = O(k^{-r_{\varphi}})$$

uniformly over $\theta \in \Theta$; (ii) $\sup_{x,\theta} [\|g(x,\theta)\| + \|g_{\theta}(x,\theta)\| + \|g_{\theta\theta}(x,\theta)\|] \le C$; (iii) Θ is a compact subspace of $R^{d_{\theta}}$; (iv) there exist $\beta_{\varphi_{\theta},j,k} \in R^{k}$ and $r_{\varphi_{\theta},j} > 0$, such that

$$\sup_{z \in \mathcal{Z}} \left| \varphi_{\theta_j}(z, \theta_0) - P_k(z)' \beta_{\varphi_{\theta}, j, k} \right| = O(k^{-r_{\varphi_{\theta}, j}})$$

for any $j = 1, ..., d_{\theta}$; (v) $\max_{j=1,...,d_{\theta}} n_2^{-1/4} k_2^{-r_{\varphi_{\theta},j}} = o(1).$

It is clear that Assumption 5(ii) implies that Assumptions 2(i), 3(i), 3(iv), 4(i) and 4(v) hold. In the rest of the section, we verify Assumptions 2(ii), 2(iv) and 3(iii).

Lemma 10. Under Assumptions 1(i), 3(v), 3(vii), 5(ii) and 5(iii),

$$\sup_{\gamma_{k_2} \in U_{k_2}} \left| n_2^{-1} \sum_{i \in I_2} \left(\sum_{i \in I_2} \left(\sum_{j \in I_2} \left(\sum_{i \in I_2} \left(\sum_{i \in I_2} \left(\sum_{j \in I_2} \left(\sum_{i \inI_2} \left(\sum_{i \in I_2} \left(\sum_{i \inI_2} \left(\sum_{i \in I_2} \left(\sum_{i \inI_2} \left(\sum_{i \in I_2} \left(\sum_{i \in I_2} \left(\sum_{i \inI_2} \left(\sum_{i \in$$

uniformly over $\theta \in \Theta$, where $U_{k_2} = \{\gamma_{k_2} \in \mathbb{R}^{k_2} : \gamma'_{k_2}\gamma_{k_2} = 1\}$

Proof. [Proof of Lemma 10] For any $\gamma_{1,k_2}, \gamma_{2,k_2} \in U_{k_2}$ and any $\theta_1, \theta_2 \in \Theta$, using the triangle inequality we get

$$\begin{aligned} \left| \gamma_{1,k_{2}}^{\prime} P_{k_{2}}(Z_{i}) \left(g(X_{i},\theta_{1}) - \varphi(Z_{i},\theta_{1}) \right) - \gamma_{2,k_{2}}^{\prime} P_{k_{2}}(Z_{i}) \left(g(X_{i},\theta_{2}) - \varphi(Z_{i},\theta_{2}) \right) \right| \\ \leq \left| \left(\gamma_{1,k_{2}} - \gamma_{2,k_{2}} \right)^{\prime} P_{k_{2}}(Z_{i}) \left(g(X_{i},\theta_{1}) - \varphi(Z_{i},\theta_{1}) \right) \right| \\ + \left| \gamma_{2,k_{2}}^{\prime} P_{k_{2}}(Z_{i}) (g(X_{i},\theta_{1}) - \varphi(Z_{i},\theta_{1}) - g(X_{i},\theta_{2}) + \varphi(Z_{i},\theta_{2})) \right|. \end{aligned}$$
(244)

By the Cauchy-Schwarz inequality,

$$\begin{aligned} &|(\gamma_{1,k_{2}} - \gamma_{2,k_{2}})' P_{k_{2}}(Z_{i}) \left(g(X_{i},\theta_{1}) - \varphi(Z_{i},\theta_{1})\right)| \\ &\leq \|\gamma_{1,k_{2}} - \gamma_{2,k_{2}}\| \sqrt{\left(g(X_{i},\theta_{1}) - \varphi(Z_{i},\theta_{1})\right)^{2} P_{k_{2}}'(Z_{i}) P_{k_{2}}(Z_{i})} \\ &\leq C\xi_{k_{1}} \|\gamma_{1,k_{1}} - \gamma_{2,k_{1}}\| \end{aligned}$$
(245)

where the second inequality is by Assumptions 5(ii) and the definition of $\varphi(z,\theta)$. By the mean value theorem,

$$g(X_i,\theta_1) - g(X_i,\theta_2) = g_\theta(X_i,\theta_i)'(\theta_1 - \theta_2), \qquad (246)$$

which together with the Cauchy-Schwarz inequality and Assumptions 5(ii) implies that for any x,

$$|g(x,\theta_1) - g(x,\theta_2)| \le C \, \|\theta_1 - \theta_2\| \,. \tag{247}$$

Similarly, we can show that for any z,

$$|\varphi(z,\theta_1) - \varphi(z,\theta_2)| \le C \left\|\theta_1 - \theta_2\right\|.$$
(248)

By the Cauchy-Schwarz inequality, the triangle inequality, (247) and (248),

$$\left|\gamma_{2,k_{2}}'P_{k_{2}}(Z_{i})(g(X_{i},\theta_{1})-\varphi(Z_{i},\theta_{1})-g(X_{i},\theta_{2})+\varphi(Z_{i},\theta_{2}))\right| \leq C \left\|\theta_{1}-\theta_{2}\right\|.$$
(249)

Combining the results in (244), (245) and (249), we get

$$\left|\gamma_{1,k_{2}}^{\prime}P_{k_{2}}(Z_{i})\left(g(X_{i},\theta_{1})-\varphi(Z_{i},\theta_{1})\right)-\gamma_{2,k_{2}}^{\prime}P_{k_{2}}(Z_{i})\left(g(X_{i},\theta_{2})-\varphi(Z_{i},\theta_{2})\right)\right|$$

$$\leq C\xi_{k_{1}}\left[\|\gamma_{1,k_{1}}-\gamma_{2,k_{1}}\|+\|\theta_{1}-\theta_{2}\|\right].$$
(250)

Let γ_{m_0,k_2} $(m_0 = 1, \dots, M_{\gamma,n})$ be a set of points such that $\min_{m_0 \le M_{\gamma,n}} \|\gamma_{k_2} - \gamma_{m_0,k_2}\| \le C^{-1} \log^{1/2}(n_2) k_2^{1/2} n_2^{-1/2} \xi_{k_2}^{-1}$

for any $\gamma_{k_2} \in U_{k_2}$. Similarly, let θ_{m_1} $(m_1 = 1, \dots, M_{\theta,n})$ be a set of points in \times such that $\min_{m_1 \leq M_{\theta,n}} \|\theta - \theta_{m_1}\| \leq C^{-1} \log^{1/2}(n_2)k_2^{1/2}n_2^{-1/2}\xi_{k_2}^{-1}$ for any $\theta \in \times$. As U_{k_2} is compact in R^{k_2} , we know that $M_{\gamma,n} \leq C(n_2^{1/2}\xi_{k_2}\log^{-1/2}(n_2)k_2^{-1/2})^{k_2}$. Similarly, $M_{\theta,n} \leq C(n_2^{1/2}\xi_{k_2}\log^{-1/2}(n_2)k_2^{-1/2})^{d_{\theta}}$, which implies that

$$M_{\gamma,n}M_{\theta,n} \le C(n_2^{1/2}\xi_{k_2}\log^{-1/2}(n_2)k_2^{-1/2})^{k_2+d_{\theta}}.$$
(251)

Hence, by the triangle inequality,

$$\sup_{\theta \in \times, \ \gamma_{k_{2}} \in U_{k_{2}}} \left| n_{2}^{-1} \sum_{i \in I_{2}} \left(\gamma_{k_{2}}' P_{k_{2}}(Z_{i}) \left(g(X_{i}, \theta) - \varphi(Z_{i}, \theta) \right) \right) \right| \\
\leq 2C \log^{1/2}(n_{2}) \left| q_{2}^{1/2} n_{2}^{-1/2} \right| \\
+ \max_{m_{0} \leq M_{\gamma,n}, m_{1} \leq M_{\theta,n}} \left| n_{2}^{-1} \gamma_{m_{0},k_{2}}' \sum_{i \in I_{2}} \left(P_{k_{2}}(Z_{i}) \left[g(X_{i}, \theta_{m_{1}}) - \varphi(Z_{i}, \theta_{m_{1}}) \right] \right| \right| . \tag{252}$$

For any m_0 and m_1 and for any *i*, by the Cauchy-Schwarz inequality, and Assumption $\mathfrak{Z}(\mathbf{i})$

$$\left| n_2^{-1} \gamma'_{m_0, k_2} P_{k_2}(Z_i) \left[g(X_i, \theta_{m_1}) - \varphi(Z_i, \theta_{m_1}) \right] \right| \le C \xi_{k_2} n_2^{-1}.$$
(253)

By Assumptions 1(i) and 3(v)

$$E\left[\left|n_{2}^{-1}\gamma_{m_{0},k_{2}}'P_{k_{2}}(Z_{i})\left[g(X_{i},\theta_{m_{1}})-\varphi(Z_{i},\theta_{m_{1}})\right]\right|^{2}\right] \le Cn_{2}^{-2}\gamma_{m_{0},k_{2}}'E[P_{k_{2}}(Z_{i})P_{k_{2}}'(Z_{i})]\gamma_{m_{0},k_{2}} \le C\lambda_{\max}(Q_{k_{2}})n_{2}^{-2} \le Cn_{2}^{-2}.$$
(254)

By (253) and (254), we can apply the Bernstein inequality to get

$$\Pr \left| n_{2}^{-1} \sum_{i \in I_{2}} \left(\binom{m_{0}, k_{2}}{m_{0}, k_{2}} P_{k_{2}}(Z_{i})(g(X_{i}, \theta_{m_{1}}) - \varphi(Z_{i}, \theta_{m_{1}})) \right| > B \log^{1/2}(n_{2}) k_{2}^{1/2} n_{2}^{-1/2} \right) \left($$

$$\leq 2 \exp \left[-\frac{B^{2} \log(n_{2}) k_{2} n_{2}^{-1}}{2C(n_{2}^{-1} + B \log^{1/2}(n_{2}) k_{2}^{1/2} n_{2}^{-3/2})} \right] \left(\left(\frac{B^{2} \log(n_{2}) k_{2}}{2C \left(1 + B \log^{1/2}(n_{2}) k_{2}^{1/2} n_{2}^{-1/2}\right)} \right) \right] \le 2 \exp \left[\left(\frac{B \log(n_{2}) k_{2}}{2C} \right),$$

$$(255)$$

where the last inequality is by Assumption 3(vii). (255) together with the Bonferroni inequality implies that

$$\Pr\left(\max_{m_{0} \leq M_{\gamma,n}, m_{1} \leq M_{\theta,n}} \left| n_{2}^{-1} \gamma_{m_{0},k_{2}}' P_{n_{2},k_{2}}'(g_{n_{2}}(\theta_{m_{1}}) - \varphi_{n_{2}}(\theta_{m_{1}})) \right| > B \log^{1/2}(n_{2}) k_{2}^{1/2} n_{2}^{-1/2} \right)$$

$$\leq 2M_{\gamma,n} M_{\theta,n} \exp\left[\left(\frac{B \log(n_{2}) k_{2}}{2C}\right] \left($$

$$\leq 2C(n_{2}^{1/2} \xi_{k_{2}} \log^{-1/2}(n_{2}) k_{2}^{-1/2})^{k_{2}+d_{\theta}} \exp\left[\left(\frac{B \log(n_{2}) k_{2}}{2C}\right]\right]$$

$$\leq 2C \exp\left[\left(\frac{B \log(n_{2}) k_{2}}{2C} + (k_{2} + d_{\theta}) \left(2^{-1} \log(n_{2}) + \log(\xi_{k_{2}}) - 2^{-1} \log(k_{2})\right)\right]\right]$$

$$\leq 2C \exp\left[\left(\frac{B \log(n_{2}) k_{2}}{8C}\right], \qquad (256)$$

where the last inequality is by the assumption that $k_2 \to \infty$, as $n_2 \to \infty$. As C is a fixed constant, from (256), we can choose B sufficiently large such that for any (fixed but) small $\varepsilon > 0$, there is

$$\Pr\left(\max_{m_0 \le M_{\gamma,n}, m_1 \le M_{\theta,n}} \left| n_2^{-1} \gamma'_{m_0,k_2} P'_{n_2,k_2} (g_{n_2}(\theta_{m_1}) - \varphi_{n_2}(\theta_{m_1})) \right| > B \log^{1/2}(n_2) k_2^{1/2} n_2^{-1/2} \right) \leqslant \varepsilon$$

for all large n_2 , which implies that

$$\max_{m_0 \le M_{\gamma,n}, m_1 \le M_{\theta,n}} \left| n_2^{-1} \gamma'_{m_0,k_2} P'_{n_2,k_2} (g_{n_2}(\theta_{m_1}) - \varphi_{n_2}(\theta_{m_1})) \right| = O_p(\log^{1/2}(n_2)k_2^{1/2}n_2^{-1/2}).$$
(257)

Combining the results in (252) and (257), the claimed result immediately follows.

Lemma 11. Under Assumptions 1(i), 1(iii), 1(v), 3(v), 3(vii), 5(i)-5(iii),

$$\sup_{\theta \in \Theta} n_2^{-1} \sum_{i \in I} \left| \widehat{\phi}_{n_2}(Z_i, \theta) - \phi(Z_i, \theta) \right|^2 = o_p(1).$$

Proof. [Proof of Lemma 11] Let $\varphi_{k_2}(z, \theta) = P_{k_2}(z)' \beta_{\theta, k_2}$. By the triangle inequality and Assumption 5(i),

$$n^{-1} \sum_{i \in I} \left| \widehat{\phi}_{n_{2}}(Z_{i}, \theta) - \phi(Z_{i}, \theta) \right|^{2} \\ \leq 2n^{-1} \sum_{i \in I} \left| \widehat{\phi}_{n_{2}}(Z_{i}, \theta) - \varphi_{k_{2}}(Z_{i}, \theta) \right|^{2} + 2n_{2}^{-1} \sum_{i \in I} |\varphi_{k_{2}}(Z_{i}, \theta) - \varphi(Z_{i}, \theta)|^{2} \\ \leq (\widehat{\beta}_{\theta, k_{2}} - \beta_{\theta, k_{2}})' Q_{n, k_{2}}(\widehat{\beta}_{\theta, k_{2}} - \beta_{\theta, k_{2}}) + o(1) \\ \leq \lambda_{\max}(Q_{n, k_{2}}) \left\| \widehat{\beta}_{\theta, k_{2}} - \beta_{\theta, k_{2}} \right\|_{\ell}^{2} + o(1).$$
(258)

By definition, $\widehat{\beta}_{\theta,k_2} - \beta_{\theta,k_2} = n^{-1}Q_{n,k_2}^{-1}\sum_{i\in I} P_{k_2}(Z_i)(g(X_i,\theta) - \varphi_{k_2}(Z_i,\theta))$. Hence

$$n_{2}^{-1} \sum_{i \in I} \left| \widehat{\phi}_{n_{2}}(Z_{i}, \theta) - \phi(Z_{i}, \theta) \right|^{2} \\ \leq 2(g_{n}(\theta) - \varphi_{n}(\theta))' P_{n,k_{2}}(P_{n,k_{2}}'P_{n,k_{2}})^{-2} P_{n,k_{2}}'(g_{n}(\theta) - \varphi_{n}(\theta)) \\ + 2(\varphi_{k_{2},n}(\theta) - \varphi_{n}(\theta))' P_{n,k_{2}}(P_{n,k_{2}}'P_{n,k_{2}})^{-2} P_{n,k_{2}}'(\varphi_{k_{2},n}(\theta) - \varphi_{n}(\theta)) \\ \leq 2\lambda_{\min}^{-2}(Q_{n,k_{2}})n^{-2}(g_{n}(\theta) - \varphi_{n}(\theta))' P_{n,k_{2}}P_{n,k_{2}}'(g_{n}(\theta) - \varphi_{n}(\theta)) \\ + 2n^{-1}\sum_{i \in I} \left| \varphi_{k_{2}}(Z_{i}, \theta) - \varphi(Z_{i}, \theta) \right|^{2}.$$
(259)

By Lemma 10, we have

$$n^{-2}(g_{n}(\theta) - \varphi_{n}(\theta))'P_{n,k_{2}}P_{n,k_{2}}'(g_{n}(\theta) - \varphi_{n}(\theta))$$

$$\leq \left\|n^{-1}P_{n,k_{2}}'(g_{n}(\theta) - \varphi_{n}(\theta))\right\| \sup_{\gamma_{k_{2}} \in U_{k_{2}}} \left|n_{2}^{-1}\sum_{i \in I_{2}}\gamma_{k_{2}}'P_{k_{2}}(Z_{i})\left(g(X_{i},\theta) - \varphi(Z_{i},\theta)\right)\right|$$

$$= \left\|n^{-1}P_{n,k_{2}}'(g_{n}(\theta) - \varphi_{n}(\theta))\right\| O_{p}(\log^{1/2}(n_{2})k_{2}^{1/2}n_{2}^{-1/2}) \qquad (260)$$

uniformly over $\theta \in \Theta$. (260) together with Assumption 3(v) then implies that

$$\sup_{\theta \in \Theta} \left\| n^{-1} P'_{n,k_2}(g_n(\theta) - \varphi_n(\theta)) \right\| = o_p(1).$$
(261)

Combining the results in (258), (259) and (261), and then applying Assumption 5(i), we immediately prove the claim of the lemma. \Box

Lemma 12. Under Assumptions 1(i), 1(iii), 1(v), 2(v), 3(v), 3(vii), 5(i)-5(iii),

$$\sup_{\theta \in \Theta} |L_n(\theta) - L_n^*(\theta)| = o_p(1).$$

Proof. [Proof of Lemma 12] For any $\theta_1, \theta_2 \in \Theta$, using the triangle inequality and Assumption 5(ii), we get

$$\begin{vmatrix}
w_{n}(Z_{i}) |h_{0}(Z_{i}) - \phi(Z_{i},\theta_{1})|^{2} - w_{n}(Z_{i}) |h_{0}(Z_{i}) - \phi(Z_{i},\theta_{2})|^{2} \\
\leq w_{n}(Z_{i}) |(2h_{0}(Z_{i}) - \phi(Z_{i},\theta_{1}))(\phi(Z_{i},\theta_{2}) - \phi(Z_{i},\theta_{1}))| \\
+ w_{n}(Z_{i}) |(\phi(Z_{i},\theta_{2}) - \phi(Z_{i},\theta_{1}))^{2}| \\
\leq C |\phi(Z_{i},\theta_{2}) - \phi(Z_{i},\theta_{1})| + C |\phi(Z_{i}(\theta_{2}) - \phi(Z_{i},\theta_{1})|^{2} \leq C ||\theta_{2} - \theta_{1}||.$$
(262)

let θ_{m_1} $(m_1 = 1, \dots, M_{\theta,n})$ be a set of points in \times such that $\min_{m_1 \leq M_{\theta,n}} \|\theta - \theta_{m_1}\| \leq n^{-1/2} \log(n)$ for any $\theta \in \times$. As \times is compact in $R^{d_{\theta}}$, we know that $M_{\theta,n} \leq C(n^{1/2} \log(n))^{d_{\theta}}$. Hence, by the triangle inequality,

$$\sup_{\theta \in \times} |L_n(\theta) - L_n^*(\theta)| \le 2Cn^{-1/2} \log(n) + \max_{m_1 \le M_{\theta,n}} |L_n(\theta_{m_1}) - L_n^*(\theta_{m_1})|.$$
(263)

For any m_1 and for any *i*, by Assumptions 2(v) and 5(ii)

$$n^{-1}w_n(Z_i) \left| h_0(Z_i) - \phi(Z_i, \theta_{m_1}) \right|^2 \le C n^{-1}.$$
(264)

By Assumptions 1(i), 2(v) and 5(ii),

$$\operatorname{Var}\left[\eta^{-1}w_{n}(Z_{i})\left|h_{0}(Z_{i})-\phi(Z_{i},\theta_{m_{1}})\right|^{2}\right] \leq n^{-2}E\left[w_{n}^{2}(Z_{i})\left|h_{0}(Z_{i})-\phi(Z_{i},\theta_{m_{1}})\right|^{4}\right] \leq Cn^{-2}.$$
(265)

By (264) and (265), we can apply the Bernstein inequality to get

$$\Pr\left(\left| \mathcal{L}_{n}(\theta_{m_{1}}) - L_{n}^{*}(\theta_{m_{1}}) \right| > Bn^{-1/2} \log(n) \right)$$

$$\leq 2 \exp\left[\left(\frac{B^{2} \log^{2}(n)n^{-1}}{2C(n^{-1} + B \log(n)n^{-3/2})} \right] \left(\frac{B^{2} \log^{2}(n)}{2C\left(1 + B \log(n)n^{-1/2}\right)} \right] \left(2 \exp\left[\left(\frac{B \log(n)}{2C} \right) \right], \quad (266)$$

where the last inequality is by Assumption 3(vii). (266) together with the Bonferroni inequality implies that

$$\Pr\left(\prod_{n_{1} \leq M_{\theta,n}} |L_{n}(\theta_{m_{1}}) - L_{n}^{*}(\theta_{m_{1}})| > Bn^{-1/2} \log(n) \right)$$

$$\leq 2M_{\theta,n} \exp\left[-\frac{B \log(n)}{2C} \right] \left($$

$$\leq 2C(n^{1/2} \log(n))^{d_{\theta}} \exp\left[\left(\frac{B \log(n)}{2C} \right) \right]$$

$$\leq 2C \exp\left[\left(\frac{B \log(n)}{2C} + 2d_{\theta} \log(n) \right) \right]$$

$$\leq 2C \exp\left[\left(\frac{B \log(n)}{2C} \log(n) \right], \qquad (267)$$

As C is a fixed constant, from (267), we can choose B sufficiently large such that for any (fixed but) small $\varepsilon > 0$, there is

$$\Pr\left(\max_{n_1 \le M_{\theta,n}} |L_n(\theta_{m_1}) - L_n^*(\theta_{m_1})| > Bn^{-1/2}\log(n)\right) \le \varepsilon$$

for all large n_2 , which implies that

$$\max_{m_1 \le M_{\theta,n}} |L_n(\theta_{m_1}) - L_n^*(\theta_{m_1})| = O_p(n^{-1/2}\log(n)).$$
(268)

Combining the results in (263) and (268), the claimed result immediately follows.

Lemma 13. Under Assumptions 1(i), 1(iii), 1(v), 2(v), 3(v), 3(vii), 5(i)-5(iv),

$$n^{-1} \sum_{i \in I} \left| \left| \widehat{\phi}_{\theta, n_2}(Z_i, \theta_0) - \phi_{\theta}(Z_i, \theta_0) \right| \right|^2 = o_p(n_2^{-1/2}).$$

Proof. [Proof of Lemma 13] By Assumption 5(ii),

$$E\left[\left|\left(g_{\theta}(X,\theta_{0})-\phi_{\theta}(Z,\theta_{0})\right)\right|^{2}\right|Z\right] \leqslant E\left[\left|\left|g_{\theta}(X,\theta_{0})\right|\right|^{2}\right|Z\right] \leqslant C$$

$$(269)$$

for any Z. Using Assumptions 1(i), 1(iii), 1(v), 5(iv) and (269), we can use similar arguments in showing (61) to show that

$$n^{-1} \sum_{i \in I} \left| \widehat{\phi}_{\theta_j, n_2}(Z_i, \theta_0) - \phi_{\theta_j, k_2}(Z_i, \theta_0) \right|_{\ell}^2 = O_p(k_2 n_2^{-1} + k_2^{-2r_{\varphi_\theta, j}})$$
(270)

for any $j = 1, \ldots, d_{\theta}$, where $\phi_{\theta_j, k_2}(z, \theta_0) = P_k(z)' \beta_{\varphi_{\theta}, j, k_2}$. By Assumption 5(iv) and (270),

$$n^{-1} \sum_{i \in I} \left| \widehat{\phi}_{\theta_{j}, n_{2}}(Z_{i}, \theta_{0}) - \phi_{\theta_{j}}(Z_{i}, \theta_{0}) \right|^{2} \left(\sum_{i \in I} \left| \widehat{\phi}_{\theta_{j}, n_{2}}(Z_{i}, \theta_{0}) - \phi_{\theta_{j}, k_{2}}(Z_{i}, \theta_{0}) \right|^{2} \right) + 2n^{-1} \sum_{i \in I} \left| \phi_{\theta_{j}, k_{2}}(Z_{i}, \theta_{0}) - \phi_{\theta_{j}}(Z_{i}, \theta_{0}) \right|^{2} \left| \sum_{i \in I} O_{p}(k_{2}n_{2}^{-1} + k_{2}^{-2r_{\varphi_{\theta}, j}}) \right|^{2} \right|$$

$$(271)$$

for any $j = 1, ..., d_{\theta}$. Using the result in (271), Assumptions 3(vii) and 5(v), we prove the claim of the lemma.