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 $\xi_k = \sup_{z\in\mathcal{Z}} \|P_k(z)\|$ and $Q_k = E\left[P_k(Z)P_k'(Z)\right]$, where $\mathcal{Z}$ 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) $Var\left[Y|\, Z\right] < C$; (iii) $C^{-1} \leq \lambda_{\min}(Q_k) \leq \lambda_{\max}(Q_k) \leq C$ for all k; (iv) there exist $\beta_{h,k} \in R^k$ and $r_h > 0$ such that*

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

(v) $\max_{j=1,2} \xi_{k_j}^2 \log(k_j) n_j^{-1} = o(1)$ 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

$$L_n(\theta) = n^{-1}\sum_{i\in I} w_n(Z_i)\left|h_0(Z_i) - \phi(Z_i,\theta)\right|^2 \text{ and } L_n^*(\theta) = E\left[w_n(Z)\left|h_0(Z) - \phi(Z,\theta)\right|^2\right]$$

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

Assumption 2. *(i) $\sup_{\theta\in\Theta} E\left[\phi^2(Z,\theta)\right] < C$; (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*

$$E\left[\left|h_0(Z) - \phi(Z,\theta)\right|^2\right] > \eta_\varepsilon \text{ for any } \theta \in \Theta \text{ with } ||\theta - \theta_0|| \geq \varepsilon;$$

(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 $\widehat{\phi}_{n_2}(Z_i,\theta)$ 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 $w(Z_i)\left|h_0(Z_i) - \phi(Z_i,\theta)\right|^2$ indexed by θ. Assumption 2(v) requires that $\widehat{w}_n(\cdot)$ 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 $\widehat{w}_n(\cdot)$ is the identity function.

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

For ease of notations, we define

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

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 $\mathcal{N}_{\delta_n} = \{\theta \in \Theta : ||\theta - \theta_0|| \leq \delta_n\}$. Define $H_{0,n} = E\left[w_n(Z)\phi_\theta(Z,\theta_0)\phi_\theta'(Z,\theta_0)\right]$. Let $\phi_{\theta_j}(z,\theta)$

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

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

Assumption 3. *The following conditions hold:*

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

(iv) $E\left[\|\phi_\theta(Z,\theta_0)\|^4\right] < \infty$;

(v) $E\left[u^2\middle| Z\right] > C^{-1}$, $E\left[\varepsilon^2\middle| Z\right] > C^{-1}$ *and* $E\left[u^4+\varepsilon^4\middle| Z\right] < C$;

(vi) $\sup_{z\in\mathcal{Z}}|w_n(z)\phi_{\theta_j}(z,\theta_0) - P_k'(z)\beta_{w\phi_j,n,k}| = o(1)$ *where* $\beta_{w\phi_j,n,k}\in R^k$ $(j=1,\ldots,d_\theta)$;

(vii) $\max_{j=1,2}(k_j n_j^{-1/2} + k_j^{-r_h} n_j^{1/2}) = o(1)$.

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 $\widehat{\phi}_{\theta,n_2}(Z_i,\theta_0)$ 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 $\sigma_u^2(Z) = E\left[u^2\middle| Z\right]$, $\sigma_\varepsilon^2(Z) = E\left[\varepsilon^2\middle| Z\right]$ and $\phi_{w\theta,n} = (w_n(Z_i)\phi_\theta(Z_i,\theta))_{i\in I}$. Define

$$\Sigma_{n_1} \equiv \frac{\phi_{w\theta,n}P_{n,k_1}Q_{n_1,k_1}^{-1}Q_{n_1,u}Q_{n_1,k_1}^{-1}P_{n,k_1}'\phi_{w\theta,n}'}{n^2 n_1}$$

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

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

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

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

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*

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

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

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

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

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

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

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

where $\widehat{\theta}_{j,n}=\iota_j^{*\prime}\widehat{\theta}_n$ and $\theta_{j,0}=\iota_j^{*\prime}\theta_0$. 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.

4 Optimal Weighting

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 variance-covariance 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 $V_{n,\theta}=H_{0,n}^{-1}E\left[w_n^2(Z)\left(n_1^{-1}\sigma_u^2(Z)+n_2^{-1}\sigma_\varepsilon^2(Z)\right)\phi_\theta(Z,\theta_0)\phi_\theta'(Z,\theta_0)\right]H_{0,n}^{-1}$.

If the sequence of the weight function is set to be

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

then the finite sample variance of the MD estimator becomes

$$V_{n,\theta}^*=\left(E\left[\left(\frac{\sigma_u^2(Z)}{n_1}+\frac{\sigma_\varepsilon^2(Z)}{n_2}\right)^{-1}\phi_\theta(Z,\theta_0)\phi_\theta'(Z,\theta_0)\right]\right)^{-1}. \tag{15}$$

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 $V_{n,\theta}\geq V_{n,\theta}^*$ 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 $z \in \mathcal{Z}$ 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 $\widehat{u}_i = Y_i - \widehat{h}_{n_1}(Z_i)$ for any $i \in I_1$, and $\widehat{\varepsilon}_i = g(Z_i, \widehat{\theta}_n) - \widehat{\phi}_{n_2}(Z_i, \widehat{\theta}_n)$ for any $i \in I_2$. Define

$$\widehat{H}_n = n^{-1} \sum_{i\in I} \widehat{w}_n(Z_i) \widehat{\phi}_{\theta,n_2}(Z_i, \widehat{\theta}_n) \widehat{\phi}_{\theta,n_2}(Z_i, \widehat{\theta}_n)',$$
$$\widehat{\Sigma}_{n_1} = \frac{\widehat{\phi}_{w\theta,n} P_{n,k_1} Q_{n_1,k_1}^{-1} \widehat{Q}_{n_1,u} Q_{n_1,k_1}^{-1} P'_{n,k_1} \widehat{\phi}'_{w\theta,n}}{n^2 n_1},$$
$$\widehat{\Sigma}_{n_2} = \frac{\widehat{\phi}_{w\theta,n} P_{n,k_2} Q_{n_2,k_2}^{-1} \widehat{Q}_{n_2,\varepsilon} Q_{n_2,k_2}^{-1} P'_{n,k_2} \widehat{\phi}'_{w\theta,n}}{n^2 n_2},$$

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

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

The following conditions are needed to show the consistency of $\widehat{V}_n$ 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 $\beta_{u,k} \in R^k$ and $r_u > 0$ such that*

$$\sup_{z\in\mathcal{Z}} \left|\sigma_u^2(z) - P_k(z)'\beta_{u,k}\right| = O(k^{-r_u}); \tag{17}$$

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

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

(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*

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

and moreover,

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

for any non-random sequence $\gamma_n \in R^{d_\theta}$ 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

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

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

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

which together with (13) and the CMT implies that

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

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 $\widehat{\theta}_{1,n}$ defined as*

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

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

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

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

$$\widehat{w}_n^*(z)=(n_1^{-1}+n_2^{-1})(n_1^{-1}\widehat{\sigma}_{n,u}^2(z)+n_2^{-1}\widehat{\sigma}_{n,\varepsilon}^2(z))^{-1}, \tag{25}$$

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

$$\widehat{\sigma}^2_{n,u}(z) = n_1^{-1} P'_{k_1}(z) Q^{-1}_{n_1,k_1} P'_{n_1,k_1} \widehat{U}_{2,n_1} \text{ and } \widehat{\sigma}^2_{n,\varepsilon}(z) = n_2^{-1} P'_{k_2}(z) Q^{-1}_{n_2,k_2} P'_{n_2,k_2} \widehat{e}_{2,n_2}, \tag{26}$$

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

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

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

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

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

and

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

Remark 5. *Under Assumption 4(iv) and*

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

Lemma 3 implies that

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

which means that $\widehat{\sigma}^2_{n,u}(z)$ and $\widehat{\sigma}^2_{n,\varepsilon}(z)$ are consistent estimators of $\sigma^2_u(z)$ and $\sigma^2_\varepsilon(z)$ 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 $\delta_{w,n} = \max_{j=1,2}(\xi_{k_j} k_j^{1/2} n_j^{-1/2} + \xi^2_{k_j} k_j^{-2r_h}) + \xi_{k_1} k_1^{-r_u} + \xi_{k_2} k_2^{-r_\varepsilon} + \xi^2_{k_2} n_1^{-1}$.

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

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

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

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

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* $\xi_{k_j} \leq C k_j^{1/2}$*. 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*

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

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

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

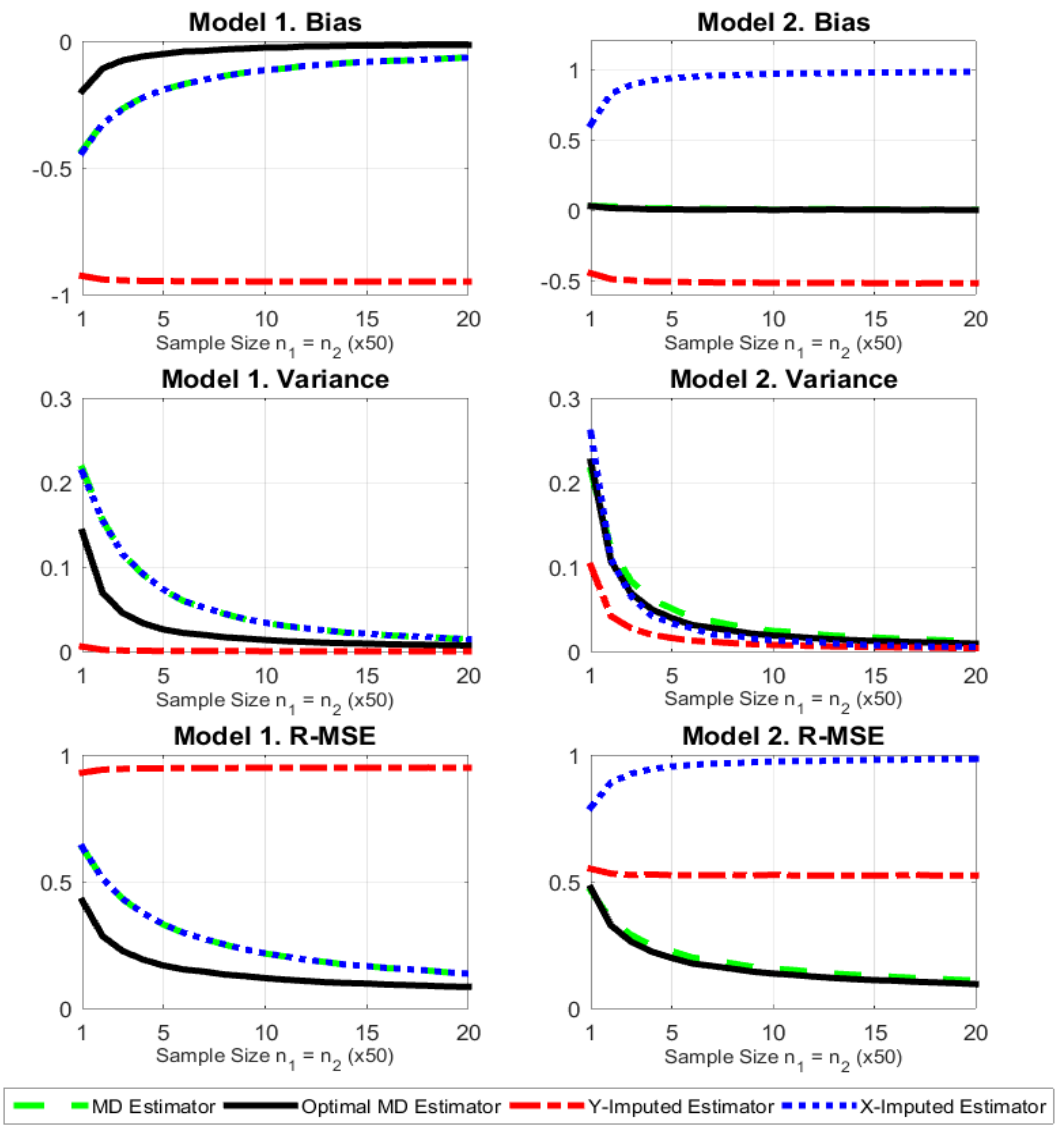
where $\theta_0 = 1$ is the unknown parameter.

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

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

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.

Figure 6.1. Properties of the MD and the Imputation Estimators ($n_1 = n_2$)



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

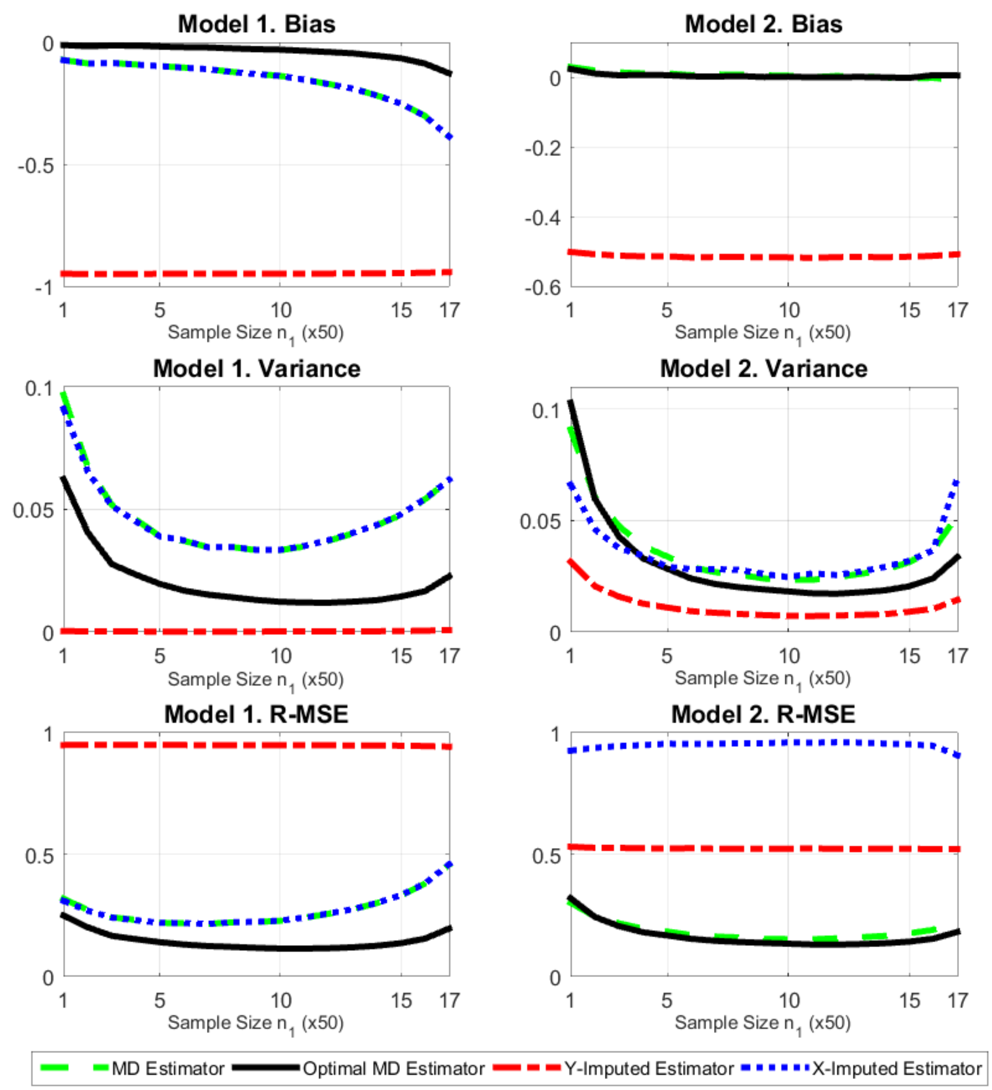
$$\widehat{\theta}_{X,n} = \arg\min_{\theta\in\Theta} n_1^{-1} \sum_{i\in I_1} (Y_i - g(\widehat{X}_i, \theta))^2 \tag{35}$$

where $\widehat{X}_i = n_2^{-1} P'_{k_2}(Z_i) Q^{-1}_{n_2,k_2} \sum_{i\in I_2} X_i P_{k_2}(Z_i)$ 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

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

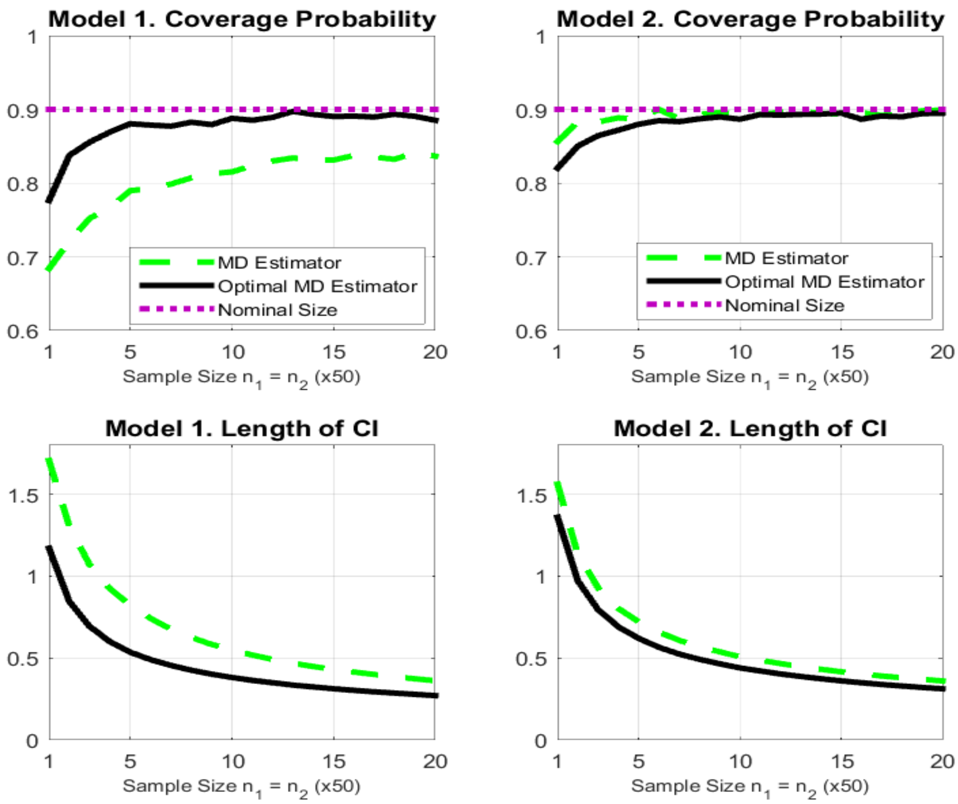
where $\widehat{Y}_i = n_1^{-1} P'_{k_1}(Z_i) Q^{-1}_{n_1,k_1} \sum_{i\in I_1} Y_i P_{k_1}(Z_i)$ for any $i \in I_2$ is the predicted value of Y_i in the second data set based on nonparametric regression. In the simulation studies, we set $k_1 = k_2 = 5$ and $P_{k_1}(Z) = P_{k_2}(Z) = (1, Z, Z^2, Z^3, Z^4)$. 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 ($n_1 + n_2 = 1000$)



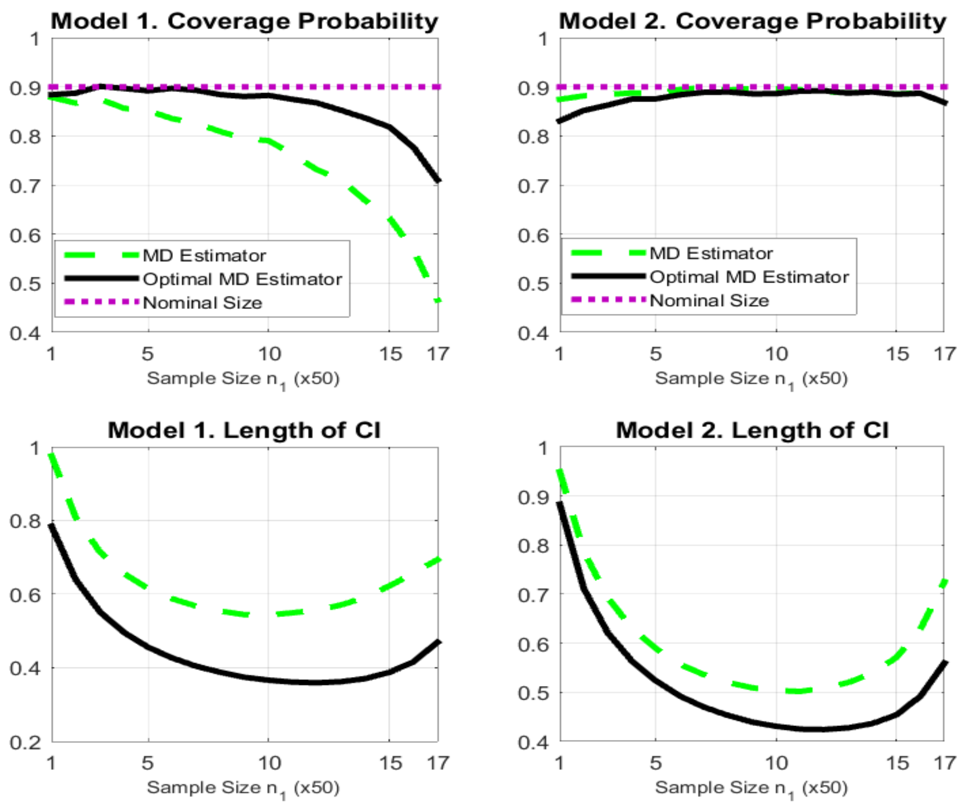
The finite sample properties of the identity weighted MD estimator (the green dashed line), the optimal weighted MD estimator (the black solid line), the X-imputed estimator (the blue dotted line) and the Y-imputed estimator (the red dash-dotted line) are presented in Figures 6.1 and 6.2. In Figure 6.1, 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 estimators under 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.

[9] 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 permanent-income 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

$$\widehat{L}_n(\theta) = n^{-1}\sum_{i\in I}\left[\widehat{w}_n(Z_i)\left|\widehat{h}_{n_1}(Z_i)-\widehat{\phi}_{n_2}(Z_i,\theta)\right|^2\right] \text{ for any } \theta\in\Theta. \tag{37}$$

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

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

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 $\widehat{\theta}_n$, it is sufficient to show that

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

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

$$\begin{aligned}\widehat{L}_n(\theta) &= n^{-1}\sum_{i\in I}\widehat{w}_n(Z_i)(|\widehat{h}_{n_1}(Z_i)-h_0(Z_i)|^2+|\widehat{\phi}_{n_2}(Z_i,\theta)-\phi(Z_i,\theta)|^2+|h_0(Z_i)-\phi(Z_i,\theta)|^2)\\ &-2n^{-1}\sum_{i\in I}\widehat{w}_n(Z_i)(\widehat{h}_{n_1}(Z_i)-h_0(Z_i))(\widehat{\phi}_{n_2}(Z_i,\theta)-\phi(Z_i,\theta))\\ &-2n^{-1}\sum_{i\in I}\widehat{w}_n(Z_i)(\widehat{\phi}_{n_2}(Z_i,\theta)-\phi(Z_i,\theta))(h_0(Z_i)-\phi(Z_i,\theta))\\ &+2n^{-1}\sum_{i\in I}\widehat{w}_n(Z_i)(\widehat{h}_{n_1}(Z_i)-h_0(Z_i))(h_0(Z_i)-\phi(Z_i,\theta)).\end{aligned} \tag{40}$$

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

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

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),

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

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

$$\left\|\widehat{\beta}_{k_1,n_1}-\beta_{h,k_1}\right\|^2 = O_p(k_1n_1^{-1}+k_1^{-2r_h}). \tag{43}$$

By the triangle inequality,

$$\begin{aligned}
& n^{-1}\sum_{i\in I}\left|\widehat{h}_{n_1}(Z_i)-h_0(Z_i)\right|^2 \\
&\leq 2n^{-1}\sum_{i\in I}\left|\widehat{h}_{n_1}(Z_i)-h_{0,k_1}(Z_i)\right|^2+2n^{-1}\sum_{i\in I}\left|h_{0,k_1}(Z_i)-h_0(Z_i)\right|^2 \\
&\leq 2(\widehat{\beta}_{k_1,n_1}-\beta_{h,k_1})'Q_{k_1,n}(\widehat{\beta}_{k_1,n_1}-\beta_{h,k_1})+2\sup_{z\in\mathcal{Z}}\left|h_{0,k_1}(z)-h_0(z)\right|^2 \\
&\leq 2\lambda_{\max}(Q_{k_1,n})\left\|\widehat{\beta}_{k_1,n_1}-\beta_{h,k_1}\right\|^2+2\sup_{z\in\mathcal{Z}}\left|h_{0,k_1}(z)-h_0(z)\right|^2 \\
&= O_p(k_1n_1^{-1}+k_1^{-2r_h})=o_p(1)
\end{aligned} \tag{44}$$

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),

$$\sup_{z\in\mathcal{Z}}\left|\widehat{w}_n(z)\right|\leq\sup_{z\in\mathcal{Z}}\left|\widehat{w}_n(z)-w_n(z)\right|+\sup_{z\in\mathcal{Z}}\left|w_n(z)\right|<2C \tag{45}$$

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

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

By (45) and Assumption 2(ii),

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

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

$$\begin{aligned}
& \sup_{\theta\in\Theta}\left|n^{-1}\sum_{i\in I}\widehat{w}_n(Z_i)(\widehat{h}_{n_1}(Z_i)-h_0(Z_i))(\widehat{\phi}_{n_2}(Z_i,\theta)-\phi(Z_i,\theta))\right| \\
&\leq\sup_{z\in\mathcal{Z}}\left|\widehat{w}_n(z)\right|\sqrt{n^{-1}\sum_{i\in I}\left|\widehat{h}_{n_1}(Z_i)-h_0(Z_i)\right|^2}\sqrt{\sup_{\theta\in\Theta}n^{-1}\sum_{i\in I}\left|\widehat{\phi}_{n_2}(Z_i,\theta)-\phi(Z_i,\theta)\right|^2}=o_p(1).
\end{aligned} \tag{48}$$

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

$$\sup_{\theta\in\Theta}E\left[\left|h_0(Z)-\phi(Z,\theta)\right|^2\right]\leq 2E\left[h_0^2(Z)\right]+2\sup_{\theta\in\Theta}E\left[\phi^2(Z,\theta)\right]<C. \tag{49}$$

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

$$
\begin{aligned}
&\sup_{\theta\in\Theta} n^{-1}\sum_{i\in I}|h_0(Z_i)-\phi(Z_i,\theta)|^2\\
&\le C\sup_{\theta\in\Theta} n^{-1}\sum_{i\in I} w_n(Z_i)\,|h_0(Z_i)-\phi(Z_i,\theta)|^2\\
&\le (C+o_p(1))\sup_{\theta\in\Theta} E\left[w_n(Z_i)\,|h_0(Z_i)-\phi(Z_i,\theta)|^2\right]\\
&\le (C+o_p(1))\sup_{\theta\in\Theta} E\left[|h_0(Z_i)-\phi(Z_i,\theta)|^2\right]=O_p(1). \qquad (50)
\end{aligned}
$$

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

$$
\begin{aligned}
&\sup_{\theta\in\Theta}\left|n^{-1}\sum_{i\in I}\widehat{w}_n(Z_i)(\widehat{\phi}_{n_2}(Z_i,\theta)-\phi(Z_i,\theta))(h_0(Z_i)-\phi(Z_i,\theta))\right|\\
&\le \sup_{z\in\mathcal{Z}}|\widehat{w}_n(z)|\sup_{\theta\in\Theta}\sqrt{n^{-1}\sum_{i\in I}|h_0(Z_i)-\phi(Z_i,\theta)|^2}\sqrt{n^{-1}\sum_{i\in I}\left|\widehat{\phi}_{n_2}(Z_i,\theta)-\phi(Z_i,\theta)\right|^2}=o_p(1). \qquad (51)
\end{aligned}
$$

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

$$
\sup_{\theta\in\Theta}\left|n^{-1}\sum_{i\in I}\widehat{w}_n(Z_i)(\widehat{h}_{n_1}(Z_i)-h_0(Z_i))(h_0(Z_i)-\phi(Z_i,\theta))\right|=o_p(1). \qquad (52)
$$

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

$$
\sup_{\theta\in\Theta}L_n(\theta)=\sup_{\theta\in\Theta}n^{-1}\sum_{i\in I}\widehat{w}_n(Z_i)\,|h_0(Z_i)-\phi(Z_i,\theta)|^2+o_p(1). \qquad (53)
$$

By (50) and Assumption 2(v),

$$
\begin{aligned}
&\sup_{\theta\in\Theta}\left|n^{-1}\sum_{i\in I}(\widehat{w}_n(Z_i)-w_n(Z_i))\,|h_0(Z_i)-\phi(Z_i,\theta)|^2\right|\\
\le\ &\sup_{z\in\mathcal{Z}}|\widehat{w}_n(z)-w_n(z)|\sup_{\theta\in\Theta}n^{-1}\sum_{i\in I}|h_0(Z_i)-\phi(Z_i,\theta)|^2=o_p(1) \qquad (54)
\end{aligned}
$$

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

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

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*

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

Proof. [Proof of Lemma 4] By definition,

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

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,

$$\begin{aligned} n^{-1}\sum_{i\in I} \widehat{\phi}^2_{\theta_{j_1}\theta_{j_2},n_2}(Z_i,\theta) &= g_{\theta_{j_1}\theta_{j_2},n_2}(\theta)' P_{n_2,k_2} Q_{n_2,k_2}^{-1} Q_{n,k_2} Q_{n_2,k_2}^{-1} P'_{n_2,k_2} g_{\theta_{j_1}\theta_{j_2},n_2}(\theta) \\ &\le \frac{\lambda_{\max}(Q_{n,k_2})}{\lambda_{\min}(Q_{n_1,k_2})} \frac{g_{\theta_{j_1}\theta_{j_2},n_2}(\theta)' P_{n_2,k_2}(P'_{n_2,k_2}P_{n_2,k_2})^{-1} P'_{n_2,k_2} g_{\theta_{j_1}\theta_{j_2},n_2}(\theta)}{n_2} \\ &\le \frac{\lambda_{\max}(Q_{n,k_2})}{\lambda_{\min}(Q_{n_1,k_2})} n_2^{-1} \sum_{i\in I_2} (g_{\theta_{j_1}\theta_{j_2}}(X_i,\theta))^2 \end{aligned}$$

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

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

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)),

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

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

$$\begin{aligned} \left\| \widehat{\beta}_{\phi,n_2} - \beta_{h,k_2} \right\|^2 &= (g_{n_2}(\theta_0) - H_{n_2,k_2})' P_{n_2,k_2} (P'_{n_2,k_2}P_{n_2,k_2})^{-2} P'_{n_2,k_2} (g_{n_2}(\theta_0) - H_{n_2,k_2}) \\ &\le \frac{(g_{n_2}(\theta_0) - H_{n_2,k_2})' P_{n_2,k_2} (P'_{n_2,k_2}P_{n_2,k_2})^{-1} P'_{n_2,k_2} (g_{n_2}(\theta_0) - H_{n_2,k_2})}{n_2 \lambda_{\min}(Q_{n_2,k_2})}, \end{aligned} \tag{57}$$

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

$$
\begin{aligned}
& n_2^{-1}(H_{n_2} - H_{n_2,k_2})' P_{n_2,k_2}(P'_{n_2,k_2}P_{n_2,k_2})^{-1}P'_{n_2,k_2}(H_{n_2} - H_{n_2,k_2}) \\
& \leq n_2^{-1}(H_{n_2} - H_{n_2,k_2})'(H_{n_2} - H_{n_2,k_2}) = O(k_2^{-2r_h}),
\end{aligned} \tag{58}
$$

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

$$
\begin{aligned}
& E\left[n_2^{-1}(g_{n_2}(\theta_0) - H_{n_2})' P_{n_2,k_2}(P'_{n_2,k_2}P_{n_2,k_2})^{-1}P'_{n_2,k_2}(g_{n_2}(\theta_0) - H_{n_2}) \middle| \{Z_i\}_{i\in I_2}\right] \\
& = n_2^{-1} tr\left((P'_{n_2,k_2}P_{n_2,k_2})^{-1}P'_{n_2,k_2}E\left[(g_{n_2}(\theta_0) - H_{n_2})(g_{n_2}(\theta_0) - H_{n_2})' \middle| \{Z_i\}_{i\in I_2}\right] P_{n_2,k_2}\right) \\
& \leq \sup_{z\in\mathcal{Z}} \sigma_\varepsilon^2(z) k_2 n_2^{-1} = O(k_2 n_2^{-1})
\end{aligned}
$$

which together with the Markov inequality implies that

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

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

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

By (42) and (60)

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

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*

$$
n^{-1}\phi_{w\theta,n}P_{n,k_1}(P'_{n,k_1}P_{n,k_1})^{-1}P'_{n,k_1}\phi'_{w\theta,n} = E\left[w_n^2(Z)\phi_\theta(Z,\theta_0)\phi'_\theta(Z,\theta_0)\right] + o_p(1).
$$

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,

$$
\begin{aligned}
\phi_{w\theta,n}M_{k_1,n}\phi'_{w\theta,n} = {} & \phi_{w\theta,k_1,n}M_{k_1,n}\phi'_{w\theta,k_1,n} + (\phi_{w\theta,n} - \phi_{w\theta,k_1,n})\, M_{k_1,n}\,(\phi_{w\theta,n} - \phi_{w\theta,k_1,n})' \\
& + (\phi_{w\theta,n} - \phi_{w\theta,k_1,n})\, M_{k_1,n}\phi'_{w\theta,k_1,n} + \phi_{w\theta,k_1,n}M_{k_1,n}\,(\phi_{w\theta,n} - \phi_{w\theta,k_1,n})'.
\end{aligned} \tag{62}
$$

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 $j_2 = 1, \ldots, d_\theta$,

$$\begin{aligned}
&\left|n^{-1}(\phi_{w\theta_{j_1},n} - \phi_{w\theta_j,k_1,n})M_{k_1,n}(\phi_{w\theta_{j_2},n} - \phi_{w\theta_{j_2},k_1,n})'\right|^2 \\
&\leq n^{-1}(\phi_{w\theta_{j_1},n} - \phi_{w\theta_j,k_1,n})M_{k_1,n}(\phi_{w\theta_{j_1},n} - \phi_{w\theta_j,k_1,n})' \\
&\times n^{-1}(\phi_{w\theta_{j_2},n} - \phi_{w\theta_{j_2},k_1,n})M_{k_1,n}(\phi_{w\theta_{j_2},n} - \phi_{w\theta_{j_2},k_1,n})' \\
&\leq n^{-1}\sum_{i\in I}\left|w_n(Z_i)\phi_{\theta_{j_1}}(Z_i,\theta_0) - \phi_{w\theta_{j_1},k_1}(Z_i,\theta_0)\right|^2 \\
&\times n^{-1}\sum_{i\in I}\left|w_n(Z_i)\phi_{\theta_{j_2}}(Z_i,\theta_0) - \phi_{w\theta_{j_2},k_1}(Z_i,\theta_0)\right|^2 = o(1)
\end{aligned} \tag{63}$$

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

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

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

$$\begin{aligned}
&n^{-1}\phi_{w\theta_{j_1},k_1,n}M_{k_1,n}\phi'_{w\theta_{j_2},k_1,n} \\
&= n^{-1}\beta'_{w\phi_{j_1},k}P'_{n,k_1}P_{n,k_1}(P'_{n,k_1}P_{n,k_1})^{-1}P'_{n,k_1}\phi'_{w\theta_{j_2},k_1,n} \\
&= n^{-1}\sum_{i\in I}\phi_{w\theta_{j_1},k_1,n}(Z_i,\theta_0)\phi_{w\theta_{j_2},k_1,n}(Z_i,\theta_0) \\
&= n^{-1}\sum_{i\in I}w_n^2(Z_i)\phi_{\theta_{j_1}}(Z_i,\theta_0)\phi_{\theta_{j_2}}(Z_i,\theta_0) \\
&+ n^{-1}\sum_{i\in I}(\phi_{w\theta_{j_1},k_1,n}(Z_i,\theta_0) - w_n(Z_i)\phi_{\theta_{j_1}}(Z_i,\theta_0))\phi_{w\theta_{j_2},k_1,n}(Z_i,\theta_0) \\
&+ n^{-1}\sum_{i\in I}w_n(Z_i)\phi_{\theta_{j_1}}(Z_i,\theta_0)(\phi_{w\theta_{j_2},k_1}(Z_i,\theta_0) - w_n(Z_i)\phi_{\theta_{j_2}}(Z_i,\theta_0)).
\end{aligned} \tag{65}$$

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),

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

which implies that

$$\begin{aligned} & n^{-1}\sum_{i\in I}(\phi_{w\theta_j,k_1,n}(Z_i,\theta_0)-w_n(Z_i)\phi_{\theta_{j_1}}(Z_i,\theta_0))\phi_{w\theta_{j_2},k_1,n}(Z_i,\theta_0) \\ & = n^{-1}\sum_{i\in I}(\phi_{w\theta_j,k_1,n}(Z_i,\theta_0)-w_n(Z_i)\phi_{\theta_{j_1}}(Z_i,\theta_0))w_n(Z_i)\phi_{\theta_{j_2}}(Z_i,\theta_0)+o_p(1). \end{aligned} \tag{69}$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned} & \left|n^{-1}\sum_{i\in I}(\phi_{w\theta_j,k_1,n}(Z_i,\theta_0)-w_n(Z_i)\phi_{\theta_{j_1}}(Z_i,\theta_0))w_n(Z_i)\phi_{\theta_{j_2}}(Z_i,\theta_0)\right|^2 \\ & \leq n^{-1}\sum_{i\in I}\left|\phi_{w\theta_j,k_1,n}(Z_i,\theta_0)-w_n(Z_i)\phi_{\theta_{j_1}}(Z_i,\theta_0)\right|^2 n^{-1}\sum_{i\in I}w_n^2(Z_i)\phi_{\theta_{j_2}}^2(Z_i,\theta_0)=o_p(1) \end{aligned} \tag{70}$$

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

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

Similarly, we can show that

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

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

$$n^{-1}\phi_{w\theta_{j_1},k_1,n}M_{k_1,n}\phi'_{w\theta_{j_2},k_1,n}=E\left[w_n^2(Z_i)\phi_{\theta_{j_1}}(Z_i,\theta_0)\phi_{\theta_{j_2}}(Z_i,\theta_0)\right]+o_p(1) \tag{73}$$

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

$$n^{-1}\phi_{w\theta,k_1,n}M_{k_1,n}\phi'_{w\theta,k_1,n}=E\left[w_n^2(Z_i)\phi_\theta(Z_i,\theta_0)\phi'_\theta(Z_i,\theta_0)\right]+o_p(1). \tag{74}$$

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

$$\begin{aligned} & \left|\frac{(\phi_{w\theta_{j_1},n}-\phi_{w\theta_{j_1},k_1,n})M_{k_1,n}\phi'_{w\theta_{j_2},k_1,n}}{n}\right|^2 \\ & \leq \frac{(\phi_{w\theta_{j_1},n}-\phi_{w\theta_{j_1},k_1,n})M_{k_1,n}(\phi_{w\theta_{j_1},n}-\phi_{w\theta_{j_1},k_1,n})'}{n}\frac{\phi_{w\theta_{j_2},k_1,n}M_{k_1,n}\phi'_{w\theta_{j_2},k_1,n}}{n}=o_p(1) \end{aligned} \tag{75}$$

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

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

and similarly

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

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

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

$$n^{-1}\sum_{i\in I}\widehat{w}_n(Z_i)(\widehat{h}_{n_1}(Z_i)-\widehat{\phi}_{n_2}(Z_i,\widehat{\theta}_n))\widehat{\phi}_{\theta,n_2}(Z_i,\widehat{\theta}_n)=0. \tag{78}$$

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

$$\begin{aligned}0&=n^{-1}\sum_{i\in I}\widehat{w}_n(Z_i)(\widehat{h}_{n_1}(Z_i)-\widehat{\phi}_{n_2}(Z_i,\theta_0))\widehat{\phi}_{\theta,n_2}(Z_i,\theta_0)\\&+n^{-1}\sum_{i\in I}\widehat{w}_n(Z_i)\widehat{\phi}_{\theta,n_2}(Z_i,\widetilde{\theta}_n)\widehat{\phi}_{\theta,n_2}(Z_i,\widetilde{\theta}_n)'(\widehat{\theta}_n-\theta_0)\\&+n^{-1}\sum_{i\in I}\widehat{w}_n(Z_i)(\widehat{h}_{n_1}(Z_i)-\widehat{\phi}_{n_2}(Z_i,\widetilde{\theta}_n))\widehat{\phi}_{\theta\theta,n_2}(Z_i,\widetilde{\theta}_n)(\widehat{\theta}_n-\theta_0),\end{aligned} \tag{79}$$

where $\widetilde{\theta}_n$ is between $\widehat{\theta}_n$ 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,

$$\left|\widehat{\phi}_{\theta_j,n_2}(Z_i,\widetilde{\theta}_{j,n})-\widehat{\phi}_{\theta_j,n_2}(Z_i,\theta_0)\right|\leq\sup_{\theta\in\mathcal{N}_{\delta_n}}\left\|\widehat{\phi}_{\theta_j\theta,n_2}(Z_i,\theta)\right\|\left\|\widetilde{\theta}_{j,n}-\theta_0\right\|$$

which together with the triangle inequality and Lemma 4 implies that

$$n^{-1}\sum_{i\in I}(\widehat{\phi}_{\theta_j,n_2}(Z_i,\widetilde{\theta}_{j,n})-\widehat{\phi}_{\theta_j,n_2}(Z_i,\theta_0))^2\leq\sup_{\theta\in\mathcal{N}_{\delta_n}}n^{-1}\sum_{i\in I}\left\|\widehat{\phi}_{\theta_j\theta,n_2}(Z_i,\theta)\right\|^2\left\|\widetilde{\theta}_{j,n}-\theta_0\right\|^2=o_p(1). \tag{80}$$

By Assumption 3(iii) and (80),

$$\begin{aligned}&n^{-1}\sum_{i\in I}(\widehat{\phi}_{\theta_j,n_2}(Z_i,\widetilde{\theta}_{j,n})-\phi_{\theta_j}(Z_i,\theta_0))^2\\&\leq 2n^{-1}\sum_{i\in I}(\widehat{\phi}_{\theta_j,n_2}(Z_i\widetilde{\theta}_{j,n})-\widehat{\phi}_{\theta_j,n_2}(Z_i,\theta_0))^2+2n^{-1}\sum_{i\in I}(\widehat{\phi}_{\theta_j,n_2}(Z_i,\theta_0)-\phi_{\theta_j}(Z_i,\theta_0))^2=o_p(1).\end{aligned} \tag{81}$$

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

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

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

$$\begin{aligned}
&\left| n^{-1}\sum_{i\in I}\widehat{w}_n(Z_i)\widehat{\phi}_{\theta_{j_1},n_2}(Z_i,\widetilde{\theta}_{j_1,n})\widehat{\phi}_{\theta_{j_2},n_2}(Z_i,\widetilde{\theta}_{j_2,n}) - n^{-1}\sum_{i\in I} w_n(Z_i)\phi_{\theta_{j_1}}(Z_i,\theta_0)\phi_{\theta_{j_2}}(Z_i,\theta_0)\right| \\
&\leq \left| n^{-1}\sum_{i\in I}(\widehat{w}_n(Z_i)-w_n(Z_i))\widehat{\phi}_{\theta_{j_1},n_2}(Z_i,\widetilde{\theta}_{j_1,n})\widehat{\phi}_{\theta_{j_2},n_2}(Z_i,\widetilde{\theta}_{j_2,n})\right| \\
&+\left| n^{-1}\sum_{i\in I} w_n(Z_i)(\widehat{\phi}_{\theta_{j_1},n_2}(Z_i,\widetilde{\theta}_{j_1,n})-\phi_{\theta_{j_1}}(Z_i,\theta_0))\widehat{\phi}_{\theta_{j_2},n_2}(Z_i,\theta_0)\right| \\
&+\left| n^{-1}\sum_{i\in I} w_n(Z_i)\phi_{\theta_{j_1}}(Z_i,\theta_0)(\widehat{\phi}_{\theta_{j_2},n_2}(Z_i,\widetilde{\theta}_{j_2,n})-\phi_{\theta_{j_2}}(Z_i,\theta_0))\right| = o_p(1). \qquad (84)
\end{aligned}$$

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

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

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

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

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

$$\begin{aligned}
& n^{-1}\sum_{i\in I}\left|\phi(Z_i,\widetilde{\theta}_{j,n})-\phi(Z_i,\theta_0)\right|^2 \\
\leq\ & 2n^{-1}\sum_{i\in I}\|\phi_\theta(Z_i,\theta_0)\|^2\,||\widetilde{\theta}_{j,n}-\theta_0||^2 \\
& +2^{-1}\sup_{\theta\in\mathcal{N}_{\delta_n}} n^{-1}\sum_{i\in I}\|\phi_{\theta\theta}(Z_i,\theta)\|^2\,||\widetilde{\theta}_{j,n}-\theta_0||^4 = o_p(1) \qquad (87)
\end{aligned}$$

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

$$\begin{aligned}
& n^{-1}\sum_{i\in I}\left|\widehat{\phi}_{n_2}(Z_i,\widetilde{\theta}_{j,n})-\phi(Z_i,\theta_0)\right|^2 \\
\leq\ & 2n^{-1}\sum_{i\in I}\left|\widehat{\phi}_{n_2}(Z_i,\widetilde{\theta}_{j,n})-\phi(Z_i,\widetilde{\theta}_n)\right|^2 + 2n^{-1}\sum_{i\in I}\left|\phi(Z_i,\widetilde{\theta}_{j,n})-\phi(Z_i,\theta_0)\right|^2 = o_p(1) \qquad (88)
\end{aligned}$$

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

$$
\begin{aligned}
&\left\| n^{-1} \sum_{i \in I} \widehat{w}_n(Z_i)(\widehat{h}_{n_1}(Z_i) - \widehat{\phi}_{n_2}(Z_i, \widetilde{\theta}_n))\widehat{\phi}_{\theta\theta,n_2}(Z_i, \widetilde{\theta}_n) \right\| \\
&\leq \sup_z |\widehat{w}_n(z)| \sqrt{\max_{j=1,\ldots,d_\theta} n^{-1} \sum_{i \in I} (\widehat{h}_{n_1}(Z_i) - \widehat{\phi}_{n_2}(Z_i, \widetilde{\theta}_{j,n}))^2} \sqrt{\max_{j=1,\ldots,d_\theta} n^{-1} \sum_{i \in I} \left\| \widehat{\phi}_{\theta\theta,n_2}(Z_i, \widetilde{\theta}_{j,n}) \right\|^2} \\
&\leq 2 \sup_z |\widehat{w}_n(z)| \sqrt{\max_{j=1,\ldots,d_\theta} n^{-1} \sum_{i \in I} \left\| \widehat{\phi}_{\theta\theta,n_2}(Z_i, \widetilde{\theta}_n) \right\|^2} \\
&\times \sqrt{n^{-1} \sum_{i \in I} \left| \widehat{h}_{n_1}(Z_i) - h_0(Z_i) \right|^2 + \max_{j=1,\ldots,d_\theta} n^{-1} \sum_{i \in I} \left| \widehat{\phi}_{n_2}(Z_i, \widetilde{\theta}_{j,n}) - \phi(Z_i, \theta_0) \right|^2} = o_p(1) \qquad (89)
\end{aligned}
$$

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

By definition,

$$
\begin{aligned}
&n^{-1} \sum_{i \in I} \widehat{w}_n(Z_i)(\widehat{h}_{n_1}(Z_i) - \widehat{\phi}_{n_2}(Z_i, \theta_0))\widehat{\phi}_{\theta,n_2}(Z_i, \theta_0) \\
&= n^{-1} \sum_{i \in I} w_n(Z_i)(\widehat{h}_{n_1}(Z_i) - \widehat{\phi}_{n_2}(Z_i, \theta_0))\phi_\theta(Z_i, \theta_0) \\
&+ n^{-1} \sum_{i \in I} (\widehat{w}_n(Z_i) - w_n(Z_i))(\widehat{h}_{n_1}(Z_i) - \widehat{\phi}_{n_2}(Z_i, \theta_0))\widehat{\phi}_{\theta,n_2}(Z_i, \theta_0) \\
&+ n^{-1} \sum_{i \in I} w_n(Z_i)(\widehat{h}_{n_1}(Z_i) - \widehat{\phi}_{n_2}(Z_i, \theta_0))(\widehat{\phi}_{\theta,n_2}(Z_i, \theta_0) - \phi_\theta(Z_i, \theta_0)). \qquad (90)
\end{aligned}
$$

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

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

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

$$
\begin{aligned}
&\left\| n^{-1} \sum_{i \in I} (\widehat{w}_n(Z_i) - w_n(Z_i))(\widehat{h}_{n_1}(Z_i) - \widehat{\phi}_{n_2}(Z_i, \theta_0))\widehat{\phi}_{\theta,n_2}(Z_i, \theta_0) \right\| \\
&\leq \sup_{z \in \mathcal{Z}} |\widehat{w}_n(z) - w_n(z)| \sqrt{n^{-1} \sum_{i \in I} (\widehat{h}_{n_1}(Z_i) - \widehat{\phi}_{n_2}(Z_i, \theta_0))^2} \sqrt{n^{-1} \sum_{i \in I} \left\| \widehat{\phi}_{\theta,n_2}(Z_i, \theta_0) \right\|^2} \\
&= o_p(n_1^{-1/2} + n_2^{-1/2}) \sqrt{n^{-1} \sum_{i \in I} \left\| \widehat{\phi}_{\theta,n_2}(Z_i, \theta_0) \right\|^2} = o_p(n_1^{-1/2} + n_2^{-1/2}). \qquad (92)
\end{aligned}
$$

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

$$
\begin{aligned}
&\left\| n^{-1} \sum_{i \in I} w_n(Z_i)(\widehat{h}_{n_1}(Z_i) - \widehat{\phi}_{n_2}(Z_i, \theta_0))(\widehat{\phi}_{\theta,n_2}(Z_i, \theta_0) - \phi_\theta(Z_i, \theta_0)) \right\| \\
&\leq \sup_{z \in \mathcal{Z}} |w_n(z)| \sqrt{n^{-1} \sum_{i \in I} (\widehat{h}_{n_1}(Z_i) - \widehat{\phi}_{n_2}(Z_i, \theta_0))^2} \sqrt{n^{-1} \sum_{i \in I} \left\| \widehat{\phi}_{\theta,n_2}(Z_i, \theta_0) - \phi_\theta(Z_i, \theta_0) \right\|^2} \\
&= o_p(n_1^{-1/2} + n_2^{-1/2}). \qquad (93)
\end{aligned}
$$

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

$$
\begin{aligned}
& n^{-1}\sum_{i\in I} w_n(Z_i)(\widehat{h}_{n_1}(Z_i)-\widehat{\phi}_{n_2}(Z_i,\theta_0))\widehat{\phi}_{\theta,n_2}(Z_i,\theta_0) \\
&= n^{-1}\sum_{i\in I} w_n(Z_i)(\widehat{h}_{n_1}(Z_i)-\widehat{\phi}_{n_2}(Z_i,\theta_0))\phi_\theta(Z_i,\theta_0)+o_p(n_1^{-1/2}+n_2^{-1/2}) \\
&= n^{-1}\sum_{i\in I} w_n(Z_i)(\widehat{h}_{n_1}(Z_i)-h_0(Z_i))\phi_\theta(Z_i,\theta_0) \\
&- n^{-1}\sum_{i\in I} w_n(Z_i)(\widehat{\phi}_{n_2}(Z_i,\theta_0)-h_0(Z_i))\phi_\theta(Z_i,\theta_0)+o_p(n_1^{-1/2}+n_2^{-1/2}).
\end{aligned} \tag{94}
$$

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

$$
\begin{aligned}
& n^{-1}\sum_{i\in I} w_n(Z_i)(\widehat{h}_{n_1}(Z_i)-h_0(Z_i))\phi_\theta(Z_i,\theta_0) \\
&= \frac{\phi_{w\theta,n}P_{n,k_1}(P'_{n_1,k_1}P_{n_1,k_1})^{-1}P'_{n_1,k_1}U_{n_1}}{n} \\
&+ \frac{\phi_{w\theta,n}P_{n,k_1}(P'_{n_1,k_1}P_{n_1,k_1})^{-1}P'_{n_1,k_1}(H_{n_1}-H_{n_1,k_1})}{n}+\frac{\phi_{w\theta,n}(H_n-H_{n,k_1})}{n}.
\end{aligned} \tag{95}
$$

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,

$$
\begin{aligned}
& \frac{\left|\phi_{w\theta,n}P_{n,k_1}(P'_{n_1,k_1}P_{n_1,k_1})^{-1}P'_{n_1,k_1}(H_{n_1}-H_{n_1,k_1})\right|^2}{n^2} \\
&\leq \frac{\phi_{w\theta,n}P_{n,k_1}P'_{n,k_1}\phi'_{w\theta,n}}{n^2}(H_{n_1}-H_{n_1,k_1})'P_{n_1,k_1}(P'_{n_1,k_1}P_{n_1,k_1})^{-2}P'_{n_1,k_1}(H_{n_1}-H_{n_1,k_1}) \\
&\leq \frac{\lambda_{\max}(Q_{n,k_1})}{\lambda_{\min}(Q_{n_1,k_1})}\frac{\phi_{w\theta,n}P_{n,k_1}(P'_{n,k_1}P_{n,k_1})^{-1}P'_{n,k_1}\phi'_{w\theta,n}}{n} \\
&\times \frac{(H_{n_1}-H_{n_1,k_1})'P_{n_1,k_1}(P'_{n_1,k_1}P_{n_1,k_1})^{-1}P'_{n_1,k_1}(H_{n_1}-H_{n_1,k_1})}{n_1} \\
&\leq \frac{\sup_{z\in\mathcal{Z}}\left|w_n^2(z)\right|\lambda_{\max}(Q_{k_1,n})}{\lambda_{\min}(Q_{k_1,n_1})}n^{-1}\sum_{i\in I}\|\phi_\theta(Z_i,\theta_0)\|^2\times n_1^{-1}\sum_{i\in I_1}|h_{0,k_1}(Z_i)-h_0(Z_i)|^2=O_p(k_1^{-2r_h}),
\end{aligned} \tag{96}
$$

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

$$
\begin{aligned}
& \left\|n^{-1}\sum_{i\in I} w_n(Z_i)(h_{0,k_1}(Z_i)-h_0(Z_i))\phi_\theta(Z_i,\theta_0)\right\| \\
&\leq \sup_{z\in\mathcal{Z}}|w_n(z)|\,n^{-1}\sum_{i\in I}\|(h_{0,k_1}(Z_i)-h_0(Z_i))\phi_\theta(Z_i,\theta_0)\| \\
&\leq C\frac{\sup_z|h_0(z)-h_{k_1}(z)|}{n}\sum_{i\in I}\|\phi_\theta(Z_i,\theta_0)\|=O_p(k_1^{-r_h}),
\end{aligned} \tag{97}
$$

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

$$n^{-1}\sum_{i\in I} w_n(Z_i)(\widehat{h}_{n_1}(Z_i)-h_0(Z_i))\phi_\theta(Z_i,\theta_0)$$
$$=\frac{\phi_{w\theta,n}P_{n,k_1}(P'_{n_1,k_1}P_{n_1,k_1})^{-1}}{n}\sum_{i\in I_1} u_i P_{k_1}(Z_i)+O_p(k_1^{-r_h}). \tag{98}$$

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

$$n^{-1}\sum_{i\in I} w_n(Z_i)(\widehat{\phi}_{n_2}(Z_i,\theta_0)-h_0(Z_i))\phi_\theta(Z_i,\theta_0)$$
$$=\frac{\phi_{w\theta,n}P_{n,k_2}(P'_{n_2,k_2}P_{n_2,k_2})^{-1}}{n}\sum_{i\in I_2}\varepsilon_i P_{k_2}(Z_i)$$
$$+\frac{\phi_{w\theta,n}P_{n,k_2}(P'_{n_2,k_2}P_{n_2,k_2})^{-1}(H_{n_2}-H_{n_2,k_2})}{n}+\frac{\phi_{w\theta,n}(H_n-H_{n,k_2})}{n} \tag{99}$$

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

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

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

$$C^{-1}<\lambda_{\min}(Q_{n_1,u})\le\lambda_{\max}(Q_{n_1,u})<C, \tag{101}$$

with probability approaching 1. Similarly, we can show that

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

with probability approaching 1.

Under the i.i.d. assumption,

$$
\begin{aligned}
&E\left[\left|\frac{\phi_{w\theta,n}P_{n,k_1}(P'_{n_1,k_1}P_{n_1,k_1})^{-1}}{n}\sum_{i\in I_1}u_iP_{k_1}(Z_i)\right|^2\middle|\{Z_i\}_{i\in I}\right]\\
&=\frac{\phi_{w\theta,n}P_{n,k_1}(P'_{n_1,k_1}P_{n_1,k_1})^{-1}Q_{n_1,u}(P'_{n_1,k_1}P_{n_1,k_1})^{-1}P'_{n,k_1}\phi'_{w\theta,n}}{n^2n_1^{-1}}\\
&\leq\frac{C\lambda_{\max}(Q_{n_1,u})}{\lambda^2_{\min}(Q_{k_1,n_1})}\frac{\phi_{w\theta,n}P_{n,k_1}P'_{n,k_1}\phi'_{w\theta,n}}{n^2n_1}\\
&\leq\frac{\lambda_{\max}(Q_{n_1,u})\lambda_{\max}(Q_{n,k_1})}{n_1\lambda^2_{\min}(Q_{n_1,k_1})}\frac{\phi_{w\theta,n}P_{n,k_1}(P'_{n,k_1}P_{n,k_1})^{-1}P'_{n,k_1}\phi'_{w\theta,n}}{n}\\
&\leq\sup_{z\in\mathcal{Z}}\left|w_n^2(z)\right|\frac{\lambda_{\max}(Q_{n_1,u})\lambda_{\max}(Q_{n,k_1})}{n_1\lambda^2_{\min}(Q_{n_1,k_1})}n^{-1}\sum_{i\in I}\|\phi_\theta(Z_i,\theta_0)\|^2=O_p(n_1^{-1}),
\end{aligned}
\tag{103}
$$

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

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

Similarly, we can show that

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

By (86) and (89),

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

which together with (79) implies that

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

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

$$
\begin{aligned}
&n^{-1}\sum_{i\in I}w_n(Z_i)(\widehat{h}_{n_1}(Z_i)-\widehat{\phi}_{n_2}(Z_i,\theta_0))\widehat{\phi}_{\theta,n_2}(Z_i,\theta_0)\\
&=\frac{\phi_{w\theta,n}P_{n,k_1}(P'_{n_1,k_1}P_{n_1,k_1})^{-1}}{n}\sum_{i\in I_1}u_iP_{k_1}(Z_i)\\
&-\frac{\phi_{w\theta,n}P_{n,k_2}(P'_{n_2,k_2}P_{n_2,k_2})^{-1}}{n}\sum_{i\in I_2}\varepsilon_iP_{k_2}(Z_i)+o_p(n_1^{-1/2}+n_2^{-1/2})
\end{aligned}
\tag{108}
$$

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

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

Then by (110), we can write

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

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), $E\left[\gamma_n'\omega_{i,n}\middle|\mathcal{F}_{i-1,n}\right]=0$ 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:

$$\sum_{i=1}^{n}E\left[(\gamma_n'\omega_{i,n})^2\middle|\mathcal{F}_{i,n}\right]\to_p 1;\ \text{and} \tag{117}$$

$$\sum_{i=1}^{n}E\left[(\gamma_n'\omega_{i,n})^2 I\left\{|\gamma_n'\omega_{i,n}|>\varepsilon\right\}\middle|\mathcal{F}_{i,n}\right]\to_p 0\forall\varepsilon>0. \tag{118}$$

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

$$\begin{aligned}\sum_{i=1}^{n}E\left[(\gamma_n'\omega_{i,n})^2\middle|\mathcal{F}_{i,n}\right]&=\sum_{i=1}^{n}\gamma_n'E\left[\omega_{i,n}\omega_{i,n}'\middle|\mathcal{F}_{i,n}\right]\gamma_n\\&=\gamma_n'\frac{D_n\phi_{w\theta,n}P_{n,k_1}Q_{n_1,k_1}^{-1}Q_{n_1,u}Q_{n_1,k_1}^{-1}P_{n,k_1}'\phi_{w\theta,n}'D_n'}{n^2n_1}\gamma_n\\&+\gamma_n'\frac{D_n\phi_{w\theta,n}P_{n,k_2}Q_{n_2,k_2}^{-1}Q_{n_2,\varepsilon}Q_{n_2,k_2}^{-1}P_{n,k_2}'\phi_{w\theta,n}'D_n'}{n^2n_2}\gamma_n\\&=\gamma_n'D_n(\Sigma_{n_1}+\Sigma_{n_2})D_n'\gamma_n=\gamma_n'\gamma_n=1\end{aligned} \tag{119}$$

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

$$\begin{aligned}&\sum_{i=1}^{n}E\left[(\gamma_n'\omega_{i,n})^2 I\left\{|\gamma_n'\omega_{i,n}|>\varepsilon\right\}\middle|\mathcal{F}_{i,n}\right]\\&\le\frac{1}{\varepsilon^2}\sum_{i=1}^{n}E\left[(\gamma_n'\omega_{i,n})^4\middle|\mathcal{F}_{i,n}\right]\\&=\frac{1}{\varepsilon^2}\sum_{i\in I_1}E\left[\frac{\left|\gamma_n'D_n\phi_{w\theta,n}P_{n,k_1}Q_{n_1,k_1}^{-1}P_{k_1}(Z_i)u_i\right|^4}{n^4n_1^4}\middle|\mathcal{F}_{i,n}\right]\\&+\frac{1}{\varepsilon^2}\sum_{i\in I_2}E\left[\frac{\left|\gamma_n'D_n\phi_{w\theta,n}P_{n,k_2}Q_{n_2,k_2}^{-1}P_{k_2}(Z_i)\varepsilon_i\right|^4}{n^4n_2^4}\middle|\mathcal{F}_{i,n}\right].\end{aligned} \tag{120}$$

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

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

By (66) in the proof of Lemma 6,

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