Supplementary Material for

"Multiway Cluster Robust Double/Debiased Machine Learning"

February 8, 2021

Abstract

This supplementary material contains mathematical proofs and extended results. Appendix A presents mathematical proofs of the main results presented in the main text. Appendix B presents a couple of useful auxiliary lemmas. Appendix C presents extended results for the case of general multiway clustering. Appendix D presents mathematical proofs of the extended results. Appendix E presents a generalization of Lemma 1.

Appendix

A Proofs of the Main Results

For any $(i, j) \in I_k \times J_\ell$, we use the shorthand notation $\mathrm{E}_P[f(W_{ij})|I_k^c \times J_\ell^c]$ to denote the conditional expectation $\mathrm{E}_P[f(W_{ij})|(W_{i'j'})_{(i',j')\in([N]\setminus I_k)\times([M]\setminus J_\ell)}]$ whenever one exists.

A.1 Proof of Theorem 1

Proof. In this proof we try to follow as closely as possible the five steps of the proof of Theorem 3.1 of CCDDHNR (2018) although all the asymptotic arguments are properly modified to account for multiway cluster sampling.

Denote $\mathcal{E}_n$ for the event $\widehat{\eta}_{k\ell} \in \mathcal{T}_n$ for all $k, \ell \in [K]^2$. Assumption 3 (i) implies $P(\mathcal{E}_n) \geq 1 - K^2\Delta_n$.

Step 1. This is the main step showing linear representation and asymptotic normality for the proposed estimator. Denote

$$\widehat{J} := \frac{1}{K^2}\sum_{(k,\ell)\in[K]^2} \mathbb{E}_{n,k\ell}[\psi^a(W; \widehat{\eta}_{k\ell})], \qquad R_{n,1} := \widehat{J} - J_0,$$

$$R_{n,2} := \frac{1}{K^2}\sum_{(k,\ell)\in[K]^2} \mathbb{E}_{n,k\ell}[\psi(W; \theta_0, \widehat{\eta}_{k\ell})] - \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\psi(W_{ij}; \theta_0, \eta_0).$$

We will later show in Steps 2, 3, 4 and 5, respectively, that

$$\|R_{n,1}\| = O_{P_n}(\underline{C}^{-1/2} + r_n), \tag{A.1}$$

$$\|R_{n,2}\| = O_{P_n}(\underline{C}^{-1/2}r'_n + \lambda_n + \lambda'_n), \tag{A.2}$$

$$\left\|\sqrt{\underline{C}}(NM)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{M}\psi(W_{ij}; \theta_0, \eta_0)\right\| = O_{P_n}(1), \tag{A.3}$$

$$\|\sigma^{-1}\| = O_{P_n}(1). \tag{A.4}$$

Then, under Assumptions 2 and 3, $\underline{C}^{-1/2} + r_N \leq \rho_n = o(1)$ and all singular values of J_0 are bounded away from zero. Therefore, with P_n-probability at least $1 - o(1)$, all singular values of $\widehat{J}$ are bounded away from zero. Thus with the same P_n probability, the multiway DML solution is uniquely written as

$$\widetilde{\theta} = -\widehat{J}^{-1} \frac{1}{K^2} \sum_{(k,\ell)\in[K]^2} \mathbb{E}_{n,k\ell}[\psi^b(W; \widehat{\eta}_{k\ell})],$$

and

$$\begin{aligned}
\sqrt{\underline{C}}(\widetilde{\theta} - \theta_0) =& - \sqrt{\underline{C}}\widehat{J}^{-1} \frac{1}{K^2} \sum_{(k,\ell)\in[K]^2} \Big(\mathbb{E}_{n,k\ell}[\psi^b(W; \widehat{\eta}_{k\ell})] + \widehat{J}\theta_0\Big) \\
=& - \sqrt{\underline{C}}\widehat{J}^{-1} \frac{1}{K^2} \sum_{(k,\ell)\in[K]^2} \mathbb{E}_{n,k\ell}[\psi(W; \theta_0, \widehat{\eta}_{k\ell})] \\
=& - \Big(J_0 + R_{n,1}\Big)^{-1} \times \Big(\frac{\sqrt{\underline{C}}}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \psi(W_{ij}; \theta_0, \eta_0) + \sqrt{\underline{C}}R_{n,2}\Big).
\end{aligned} \tag{A.5}$$

Using the fact that

$$\Big(J_0 + R_{n,1}\Big)^{-1} - J_0^{-1} = -(J_0 + R_{n,1})^{-1} R_{n,1} J_0^{-1},$$

we have

$$\begin{aligned}
\|(J_0 + R_{n,1})^{-1} - J_0^{-1}\| =& \|(J_0 + R_{n,1})^{-1} R_{n,1} J_0^{-1}\| \leq \|(J_0 + R_{n,1})^{-1}\| \, \|R_{n,1}\| \, \|J_0^{-1}\| \\
=& O_{P_n}(1) O_{P_n}(\underline{C}^{-1/2} + r_n) O_{P_n}(1) = O_{P_n}(\underline{C}^{-1/2} + r_n).
\end{aligned}$$

Furthermore, $r_n' + \sqrt{\underline{C}}(\lambda_n + \lambda_n') \leq \rho_n = o(1)$, it holds that

$$\begin{aligned}
\Big\|\frac{\sqrt{\underline{C}}}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \psi(W_{ij}; \theta_0, \eta_0) + \sqrt{\underline{C}}R_{n,2}\Big\| \leq& \Big\|\frac{\sqrt{\underline{C}}}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \psi(W_{ij}; \theta_0, \eta_0)\Big\| + \Big\|\sqrt{\underline{C}}R_{n,2}\Big\| \\
=& O_{P_n}(1) + o_{P_n}(1) = O_{P_n}(1),
\end{aligned}$$

where the first equality is due to (A.3) and (A.4). Combining above two bounds gives

$$\begin{aligned}
\Big\|\Big(J_0 + R_{n,1}\Big)^{-1} - J_0^{-1}\Big\| \times \Big\|\frac{\sqrt{\underline{C}}}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \psi(W_{ij}; \theta_0, \eta_0) + \sqrt{\underline{C}}R_{n,2}\Big\| =& O_{P_n}(\underline{C}^{-1/2} + r_n) O_{P_n}(1) \\
=& O_{P_n}(\underline{C}^{-1/2} + r_n).
\end{aligned} \tag{A.6}$$

Therefore, from (A.4), (A.5) and (A.6), we have

$$\sqrt{\underline{C}}\sigma^{-1}(\widetilde{\theta}-\theta_0) =\frac{\sqrt{\underline{C}}}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\bar{\psi}(W_{ij})+O_{P_n}(\rho_n).$$

The first term on the RHS above can be written as $\mathbb{G}_n\bar{\psi}$. Applying Lemma 1, we obtain the independent linear representation

$$H_n\bar{\psi}:=\sum_{i=1}^{N}\frac{\sqrt{\underline{C}}}{N}\mathrm{E}_{P_n}[\bar{\psi}(W_{ij})|U_{i0}]+\sum_{j=1}^{M}\frac{\sqrt{\underline{C}}}{M}\mathrm{E}_{P_n}[\bar{\psi}(W_{ij})|U_{0j}]$$

and it holds P_n-a.s. that

$$\begin{aligned} V(\mathbb{G}_n\bar{\psi}) =& V(H_n\bar{\psi})+O(\underline{C}^{-1}) = J_0^{-1}\Gamma(J_0^{-1})'+O(\underline{C}^{-1}) \qquad \text{and} \\ \mathbb{G}_n\bar{\psi} =& H_n\bar{\psi}+O_P(\underline{C}^{-1/2}) \end{aligned}$$

under Assumption 3 (iv). Recall that $q\geq 4$, the third moments of both summands of $H_n\bar{\psi}$ are bounded over n under Assumptions 2(v) and 3 (ii) (iv). We have verified all the conditions for Lyapunov's CLT. An application of Lyapunov's CLT and Cramer-Wold device gives

$$H_n\bar{\psi}\rightsquigarrow N(0,I_{d_\theta})$$

and an application of Theorem 2.7 of van der Vaart (1998) concludes the proof.

Step 2. Since K is fixed, it suffices to show for any $(k,\ell)\in[K]^2$,

$$\Big\|\mathbb{E}_{n,k\ell}[\psi^a(W;\widehat{\eta}_{k\ell})]-\mathrm{E}_P[\psi^a(W_{11};\eta_0)]\Big\|=O_{P_n}(\underline{C}^{-1/2}+r_n).$$

Fix $(k,\ell)\in[K]^2$,

$$\Big\|\mathbb{E}_{n,k\ell}[\psi^a(W;\widehat{\eta}_{k\ell})]-\mathrm{E}_{P_n}[\psi^a(W_{ij};\eta_0)]\Big\|\leq\mathcal{I}_{1,k\ell}+\mathcal{I}_{2,k\ell}.$$

where

$$\begin{aligned} \mathcal{I}_{1,k\ell} &:= \Big\|\mathbb{E}_{n,k\ell}[\psi^a(W;\widehat{\eta}_{k\ell})]-\mathrm{E}_{P_n}[\psi^a(W_{ij};\widehat{\eta}_{k\ell})|I_k^c\times J_\ell^c]\Big\| \\ \mathcal{I}_{2,k\ell} &:= \Big\|\mathrm{E}_{P_n}[\psi^a(W_{ij};\widehat{\eta}_{k\ell})|I_k^c\times J_\ell^c]-\mathrm{E}_{P_n}[\psi^a(W_{11};\eta_0)]\Big\|. \end{aligned}$$

Notice that $\mathcal{I}_{2,k\ell} \le r_n$ with P_n-probability $1-o(1)$ follows directly from Assumptions 1 (ii) and 3 (iii). Now denote $\widetilde{\psi}^a_{ij,m} = \psi^a_m(W_{ij};\widehat{\eta}_{k\ell}) - \mathrm{E}_{P_n}[\psi^a_m(W_{ij};\widehat{\eta}_{k\ell})|I^c_k \times J^c_\ell]$ and $\widetilde{\psi}^a_{ij} = (\widetilde{\psi}^a_{ij,m})_{m\in[d_\theta]}$. To bound $\mathcal{I}_{1,k\ell}$, note that conditional on $I^c_k \times J^c_\ell$, it holds that

$$
\begin{aligned}
\mathrm{E}_{P_n}[\mathcal{I}^2_{1,k\ell}|I^c_k \times J^c_\ell] =& \mathrm{E}_{P_n}\Big[\Big\|\mathbb{E}_{n,k\ell}[\psi^a(W;\widehat{\eta}_{k\ell})] - \mathrm{E}_{P_n}[\psi^a(W_{ij};\widehat{\eta}_{k\ell})|I^c_k \times J^c_\ell]\Big\|^2\Big|I^c_k \times J^c_\ell\Big]\\
=& \frac{1}{(|I||J|)^2}\mathrm{E}_{P_n}\Big[\sum_{m=1}^{d_\theta}\Big(\sum_{(i,j)\in I_k\times J_\ell}\widetilde{\psi}^a_{ij,m}\Big)^2\Big|I^c_k \times J^c_\ell\Big]\\
=& \frac{1}{(|I||J|)^2}\sum_{(i,j)\in I_k\times J_\ell}\sum_{j'\in J_\ell, j'\neq j}\mathrm{E}_{P_n}\Big[\sum_{m=1}^{d_\theta}\widetilde{\psi}^a_{ij,m}\widetilde{\psi}^a_{ij',m}\Big|I^c_k \times J^c_\ell\Big]\\
&+ \frac{1}{(|I||J|)^2}\sum_{(i,j)\in I_k\times J_\ell}\sum_{i'\in I_k, i'\neq i}\mathrm{E}_{P_n}\Big[\sum_{m=1}^{d_\theta}\widetilde{\psi}^a_{ij,m}\widetilde{\psi}^a_{i'j,m}\Big|I^c_k \times J^c_\ell\Big]\\
&+ \frac{1}{(|I||J|)^2}\sum_{(i,j)\in I_k\times J_\ell}\mathrm{E}_{P_n}\Big[\sum_{m=1}^{d_\theta}(\widetilde{\psi}^a_{ij,m})^2\Big|I^c_k \times J^c_\ell\Big] + 0\\
=& \frac{1}{(|I||J|)^2}\sum_{(i,j)\in I_k\times J_\ell}\sum_{j'\in J_\ell, j'\neq j}\mathrm{E}_{P_n}[\langle\widetilde{\psi}^a_{ij},\widetilde{\psi}^a_{ij'}\rangle|I^c_k \times J^c_\ell]\\
&+ \frac{1}{(|I||J|)^2}\sum_{(i,j)\in I_k\times J_\ell}\sum_{i'\in I_k, i'\neq i}\mathrm{E}_{P_n}[\langle\widetilde{\psi}^a_{ij},\widetilde{\psi}^a_{i'j}\rangle|I^c_k \times J^c_\ell]\\
&+ \frac{1}{(|I||J|)^2}\sum_{(i,j)\in I_k\times J_\ell}\mathrm{E}_{P_n}[\|\widetilde{\psi}^a_{ij}\|^2|I^c_k \times J^c_\ell]\\
\lesssim& \frac{1}{|I|\wedge|J|}\mathrm{E}_{P_n}\Big[\Big\|\psi^a(W_{ij};\widehat{\eta}_{k\ell}) - \mathrm{E}_{P_n}[\psi^a(W_{ij};\widehat{\eta}_{k\ell})|I^c_k \times J^c_\ell]\Big\|^2\Big|I^c_k \times J^c_\ell\Big]\\
\le& \frac{1}{|I|\wedge|J|}\mathrm{E}_{P_n}[\|\psi^a(W_{ij};\widehat{\eta}_{k\ell})\|^2|I^c_k \times J^c_\ell]\\
\le& c_1^2/|I|\wedge|J|
\end{aligned}
$$

under an application of Cauchy-Schwarz's inequality and Assumptions 1 and 3 (ii). Note that $\underline{C} \lesssim |I|\wedge|J| \lesssim \underline{C}$. Hence an application of Lemma 2 (i) implies $\mathcal{I}_{1,k\ell} = O_{P_n}(\underline{C}^{-1/2})$. This completes a proof of (A.1).

Step 3. It again suffices to show that for any $(k,\ell)\in[K]^2$, one has

$$\Big\|\mathbb{E}_{n,k\ell}[\psi(W;\theta_0,\widehat{\eta}_{k\ell})]-\frac{1}{|I||J|}\sum_{(i,j)\in I_k\times J_\ell}\psi(W_{ij};\theta_0,\eta_0)\Big\|=O_{P_n}(\underline{C}^{-1/2}r_n'+\lambda_n+\lambda_n')$$

Denote

$$\mathbb{G}_{n,k\ell}[\phi(W)]=\frac{\sqrt{\underline{C}}}{|I||J|}\sum_{(i,j)\in I_k\times J_\ell}\Big(\phi(W_{ij})-\int\phi(w)dP_n\Big),$$

where ϕ is P_n an integrable function on $\mathrm{supp}(W)$. Then

$$\Big\|\mathbb{E}_{n,k\ell}[\psi(W;\theta_0,\widehat{\eta}_{k\ell})]-\frac{1}{|I||J|}\sum_{(i,j)\in I_k\times J_\ell}\psi(W_{ij};\theta_0,\eta_0)\Big\|\le\frac{\mathcal{I}_{3,k\ell}+\mathcal{I}_{4,k\ell}}{\sqrt{\underline{C}}}$$

where

$$\begin{aligned}
\mathcal{I}_{3,k\ell}:=&\Big\|\mathbb{G}_{n,k\ell}[\psi(W;\theta_0,\widehat{\eta}_{k\ell})]-\mathbb{G}_{n,k\ell}[\psi(W;\theta_0,\eta_0)]\Big\|,\\
\mathcal{I}_{4,k\ell}:=&\sqrt{\underline{C}}\Big\|\mathrm{E}_{P_n}[\psi(W_{ij};\theta_0,\widehat{\eta}_{k\ell})|I_k\times J_\ell]-\mathrm{E}_{P_n}[\psi(W_{11};\theta_0,\eta_0)]\Big\|.
\end{aligned}$$

Denote $\widetilde{\psi}_{ij,m}:=\psi_m(W_{ij};\theta_0,\widehat{\eta}_{k\ell})-\psi_m(W_{ij};\theta_0,\eta_0)$ and $\widetilde{\psi}_{ij}=(\widetilde{\psi}_{ij,m})_{m\in[d_\theta]}$. To bound $\mathcal{I}_{3,k\ell}$, notice that using a similar argument as for the bound of $\mathcal{I}_{1,k\ell}$, one has

$$\begin{aligned}
&\mathrm{E}_{P_n}[\|\mathcal{I}_{3,k\ell}\|^2|I_k^c\times J_\ell^c]=\mathrm{E}_{P_n}[\|\mathbb{G}_{n,k\ell}[\psi(W_{ij};\theta_0,\widehat{\eta}_{k\ell})-\psi(W_{ij};\theta_0,\eta_0)]\|^2|I_k^c\times J_\ell^c]\\
=&\mathrm{E}_{P_n}\Big[\frac{\underline{C}}{(|I||J|)^2}\sum_{m=1}^{d_\theta}\Big\{\sum_{(i,j)\in I_k\times J_\ell}\Big(\widetilde{\psi}_{ij,m}-\mathrm{E}_{P_n}\widetilde{\psi}_{ij,m}\Big)\Big\}^2\Big|I_k^c\times J_\ell^c\Big]\\
=&\frac{\underline{C}}{(|I||J|)^2}\sum_{(i,j)\in I_k\times J_\ell}\sum_{j'\in J_\ell,j'\neq j}\mathrm{E}_{P_n}\Big[\sum_{m=1}^{d_\theta}\Big(\widetilde{\psi}_{ij,m}-\mathrm{E}_{P_n}\widetilde{\psi}_{ij,m}\Big)\Big(\widetilde{\psi}_{ij',m}-\mathrm{E}_{P_n}\widetilde{\psi}_{ij',m}\Big)\Big|I_k^c\times J_\ell^c\Big]\\
&+\frac{\underline{C}}{(|I||J|)^2}\sum_{(i,j)\in I_k\times J_\ell}\sum_{i'\in I_k,i'\neq i}\mathrm{E}_{P_n}\Big[\sum_{m=1}^{d_\theta}\Big(\widetilde{\psi}_{ij,m}-\mathrm{E}_{P_n}\widetilde{\psi}_{ij,m}\Big)\Big(\widetilde{\psi}_{i'j,m}-\mathrm{E}_{P_n}\widetilde{\psi}_{i'j,m}\Big)\Big|I_k^c\times J_\ell^c\Big]\\
&+\frac{\underline{C}}{(|I||J|)^2}\sum_{(i,j)\in I_k\times J_\ell}\mathrm{E}_{P_n}\Big[\sum_{m=1}^{d_\theta}\Big(\widetilde{\psi}_{ij,m}-\mathrm{E}_{P_n}\widetilde{\psi}_{ij,m}\Big)^2\Big|I_k^c\times J_\ell^c\Big]+0\\
=&\frac{\underline{C}}{(|I||J|)^2}\sum_{(i,j)\in I_k\times J_\ell}\sum_{j'\in J_\ell,j'\neq j}\mathrm{E}_{P_n}\Big[\langle\widetilde{\psi}_{ij}-\mathrm{E}_{P_n}\widetilde{\psi}_{ij},\widetilde{\psi}_{ij'}-\mathrm{E}_{P_n}\widetilde{\psi}_{ij'}\rangle\Big|I_k^c\times J_\ell^c\Big]\\
&+\frac{\underline{C}}{(|I||J|)^2}\sum_{(i,j)\in I_k\times J_\ell}\sum_{i'\in I_k,i'\neq i}\mathrm{E}_{P_n}\Big[\langle\widetilde{\psi}_{ij}-\mathrm{E}_{P_n}\widetilde{\psi}_{ij},\widetilde{\psi}_{i'j}-\mathrm{E}_{P_n}\widetilde{\psi}_{i'j}\rangle\Big|I_k^c\times J_\ell^c\Big]
\end{aligned}$$

$$
\begin{aligned}
&+\frac{C}{(|I||J|)^2}\sum_{(i,j)\in I_k\times J_\ell}\mathrm{E}_{P_n}\Big[\Big\|\widetilde{\psi}_{ij}-\mathrm{E}_{P_n}\widetilde{\psi}_{ij}\Big\|^2\Big|I_k^c\times J_\ell^c\Big]\\
\lesssim&\mathrm{E}_{P_n}\Big[\Big\|\psi(W_{ij};\theta_0,\widehat{\eta})-\psi(W_{ij};\theta_0,\eta_0)-\mathrm{E}_{P_n}[\psi(W_{ij};\theta_0,\widehat{\eta})-\psi(W_{ij};\theta_0,\eta_0)]\Big\|^2\Big|I_k^c\times J_\ell^c\Big]\\
\le&\mathrm{E}_{P_n}[\|\psi(W_{ij};\theta_0,\widehat{\eta})-\psi(W_{ij};\theta_0,\eta_0)\|^2|I_k^c\times J_\ell^c]\\
\le&\sup_{\eta\in\mathcal{T}_n}\mathrm{E}_{P_n}[\|\psi(W_{00};\theta_0,\eta)-\psi(W_{00};\theta_0,\eta_0)\|^2|I_k^c\times J_\ell^c]\\
=&\sup_{\eta\in\mathcal{T}_n}\mathrm{E}_{P_n}[\|\psi(W_{00};\theta_0,\eta)-\psi(W_{00};\theta_0,\eta_0)\|^2]=(r_n')^2,
\end{aligned}
$$

where the first inequality follows from Cauchy-Schwarz's inequality, the second-to-last equality is due to Assumption 1, and the last equality is due to Assumption 3 (iii).

Hence, $\mathcal{I}_{3,k\ell}=O_{P_n}(r_n')$. To bound $\mathcal{I}_{4,k\ell}$, let

$$
f_{k\ell}(r):=\mathrm{E}_{P_n}[\psi(W_{ij};\theta_0,\eta_0+r(\widehat{\eta}_{k\ell}-\eta_0))|I_k^c\times J_\ell^c]-\mathrm{E}_{P_n}[\psi(W_{11};\theta_0,\eta_0)],\qquad r\in[0,1].
$$

An application of the mean value expansion coordinate-wise gives

$$
f_{k\ell}(1)=f_{k\ell}(0)+f_{k\ell}'(0)+f_{k\ell}''(\widetilde{r})/2,
$$

where $\widetilde{r}\in(0,1)$. Note that $f_{k\ell}(0)=0$ under Assumption 2 (i), and

$$
\|f_{k\ell}'(0)\|=\Big\|\partial_\eta\mathrm{E}_{P_n}\psi(W;\theta_0,\eta_0)[\widehat{\eta}_{k\ell}-\eta_0]\Big\|\le\lambda_n
$$

under Assumption 2 (iv). Moreover, under Assumption 3 (iii), on the event $\mathcal{E}_n$, we have

$$
\|f_{k\ell}''(\widetilde{r})\|\le\sup_{r\in(0,1)}\|f_{k\ell}''(r)\|\le\lambda_n'.
$$

This completes a proof of (A.2).

Step 4. Note that

$$
\begin{aligned}
&\mathrm{E}_{P_n}\Big[\Big\|\frac{\sqrt{C}}{NM}\sum_{i=1}^N\sum_{j=1}^M\psi(W_{ij};\theta_0,\eta_0)\Big\|^2\Big]\\
=&\frac{C}{(NM)^2}\mathrm{E}_{P_n}\Big[\sum_{m=1}^{d_\theta}\Big(\sum_{i=1}^N\sum_{j=1}^M\psi_m(W_{ij};\theta_0,\eta_0)\Big)^2\Big]
\end{aligned}
$$

$$
\begin{aligned}
=&\frac{\underline{C}}{(NM)^2}\sum_{i=1}^{N}\sum_{1\le j<j'\le M}\mathrm{E}_{P_n}\Big[\sum_{m=1}^{d_\theta}\psi_m(W_{ij};\theta_0,\eta_0)\psi_m(W_{ij'};\theta_0,\eta_0)\Big]\\
&+\frac{\underline{C}}{(NM)^2}\sum_{1\le i<i'\le N}\sum_{j=1}^{M}\mathrm{E}_{P_n}\Big[\sum_{m=1}^{d_\theta}\psi_m(W_{ij};\theta_0,\eta_0)\psi_m(W_{i'j};\theta_0,\eta_0)\Big]\\
&+\frac{\underline{C}}{(NM)^2}\sum_{i=1}^{N}\sum_{j=1}^{M}\mathrm{E}_{P_n}\Big[\sum_{m=1}^{d_\theta}\psi_m^2(W_{ij};\theta_0,\eta_0)\Big]+0\\
\lesssim&\mathrm{E}_{P_n}[\|\psi(W_{ij};\theta_0,\eta_0)\|^2]\le c_1^2
\end{aligned}
$$

under Assumptions 1 and 3 (ii). Therefore, an application of Markov's inequality implies

$$
\Big\|\frac{\sqrt{\underline{C}}}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\psi(W_{ij};\theta_0,\eta_0)\Big\|=O_{P_n}(1).
$$

This completes a proof of (A.3).

Step 5. Note that all singular values of J_0 are bounded from above by c_1 under Assumption 2 (v) and all eigenvalues of Γ are bounded from below by c_0 under Assumption 3 (iv). Therefore, we have $\|\sigma^{-1}\|\le c_1/\sqrt{c_0}$ and thus $\|\sigma^{-1}\|=O_{P_n}(1)$. This completes a proof of (A.4). □

A.2 Proof of Theorem 2

Proof. Step 2 of the proof of Theorem 1 proves $\|\widehat{J}-J_0\|=O_p(\underline{C}^{-1/2}+r_n)$ and Assumption 2 (v) implies $\|J_0^{-1}\|\le c_0^{-1}$. Therefore, to prove the claim of the theorem, it suffices to show

$$
\begin{aligned}
\Big\|\frac{1}{K^2}\sum_{(k,\ell)\in[K]^2}&\Big\{\frac{|I|\wedge|J|}{(|I||J|)^2}\sum_{i\in I_k}\sum_{j,j'\in J_\ell}\psi(W_{ij};\widetilde{\theta},\widehat{\eta}_{k\ell})\psi(W_{ij'};\widetilde{\theta},\widehat{\eta}_{k\ell})'\\
&+\frac{|I|\wedge|J|}{(|I||J|)^2}\sum_{i,i'\in I_k}\sum_{j\in J_\ell}\psi(W_{ij};\widetilde{\theta},\widehat{\eta}_{k\ell})\psi(W_{i'j};\widetilde{\theta},\widehat{\eta}_{k\ell})'\Big\}\\
-\bar{\mu}_N\mathrm{E}_P[\psi(W_{11};\theta_0,\eta_0)&\psi(W_{12};\theta_0,\eta_0)']-\bar{\mu}_M\mathrm{E}_P[\psi(W_{11};\theta_0,\eta_0)\psi(W_{21};\theta_0,\eta_0)']\Big\|=O_P(\rho_n).
\end{aligned}
$$

Moreover, since K and d_θ are constants and $\mu_N\to\bar{\mu}_N\le 1$ and $\mu_M\to\bar{\mu}_M\le 1$, it suffices to show that for each $(k,\ell)\in[K]^2$ and $l,m\in[d_\theta]$, it holds that

$$
\Big|\frac{|I|\wedge|J|}{(|I||J|)^2}\sum_{i\in I_k}\sum_{j,j'\in J_\ell}\psi_l(W_{ij};\widetilde{\theta},\widehat{\eta}_{k\ell})\psi_m(W_{ij'};\widetilde{\theta},\widehat{\eta}_{k\ell})-\mu_N\mathrm{E}_P[\psi_l(W_{11};\theta_0,\eta_0)\psi_m(W_{12};\theta_0,\eta_0)]\Big|=O_P(\rho_n)
$$

and

$$\Big|\frac{|I|\wedge|J|}{(|I||J|)^2}\sum_{i,i'\in I_k}\sum_{j\in J_\ell}\psi_l(W_{ij};\widetilde{\theta},\widehat{\eta}_{k\ell})\psi_m(W_{i'j};\widetilde{\theta},\widehat{\eta}_{k\ell})-\mu_M \mathrm{E}_P[\psi_l(W_{11};\theta_0,\eta_0)\psi_m(W_{21};\theta_0,\eta_0)]\Big| = O_P(\rho_n).$$

We will show the second statement since the first one follows analogously. Denote the left-hand side of the equation as $\mathcal{I}_{k\ell,lm}$. First, note that $(|I|\wedge|J|)/|J| = \mu_M$, and apply the triangle inequality to get

$$\mathcal{I}_{k\ell,lm} \le \mathcal{I}_{k\ell,lm,1} + \mathcal{I}_{k\ell,lm,2},$$

where

$$\mathcal{I}_{k\ell,lm,1} := \Big|\frac{1}{|I|^2|J|}\sum_{i,i'\in I_k}\sum_{j\in J_\ell}\Big\{\psi_l(W_{ij};\widetilde{\theta},\widehat{\eta}_{k\ell})\psi_m(W_{i'j};\widetilde{\theta},\widehat{\eta}_{k\ell})-\psi_l(W_{ij};\theta_0,\eta_0)\psi_m(W_{i'j};\theta_0,\eta_0)\Big\}\Big|$$

$$\mathcal{I}_{k\ell,lm,2} := \Big|\frac{1}{|I|^2|J|}\sum_{i,i'\in I_k}\sum_{j\in J_\ell}\psi_l(W_{ij};\theta_0,\eta_0)\psi_m(W_{i'j};\theta_0,\eta_0)-\mathrm{E}_P[\psi_l(W_{11};\theta_0,\eta_0)\psi_m(W_{21};\theta_0,\eta_0)]\Big|.$$

We first find a bound for $\mathcal{I}_{k\ell,lm,2}$. Since $q>4$, it holds that

$$\begin{aligned}
\mathrm{E}_P[\mathcal{I}^2_{k\ell,lm,2}] =& \frac{1}{|I|^4|J|^2}\mathrm{E}_P\Big[\Big|\sum_{i,i'\in I_k}\sum_{j\in J_\ell}\psi_l(W_{ij};\theta_0,\eta_0)\psi_m(W_{i'j};\theta_0,\eta_0)-\mathrm{E}_P[\psi_l(W_{11};\theta_0,\eta_0)\psi_m(W_{21};\theta_0,\eta_0)]\Big|^2\Big]\\
\le& \frac{1}{|I|^4|J|^2}\mathrm{E}_P\Big[\sum_{i,i',i''\in I_k}\sum_{j,j'\in J_\ell}\psi_l(W_{ij};\theta_0,\eta_0)\psi_m(W_{i'j};\theta_0,\eta_0)\psi_l(W_{ij'};\theta_0,\eta_0)\psi_m(W_{i''j'};\theta_0,\eta_0)\Big]\\
&+\frac{1}{|I|^4|J|^2}\mathrm{E}_P\Big[\sum_{i,i',i'',i'''\in I_k}\sum_{j\in J_\ell}\psi_l(W_{ij};\theta_0,\eta_0)\psi_m(W_{i'j};\theta_0,\eta_0)\psi_l(W_{i''j};\theta_0,\eta_0)\psi_m(W_{i'''j};\theta_0,\eta_0)\Big]\\
&+o((|I|\wedge|J|)^{-1})+0\\
\lesssim& \frac{1}{|I|\wedge|J|}\mathrm{E}_P[\|\psi(W;\theta_0,\eta_0)\|^4]\lesssim c_1^4/\underline{C}=O(\underline{C}^{-1}).
\end{aligned}$$

Now, to bound $\mathcal{I}_{k\ell,lm,1}$, we make use of the following identity coming from the proof of Theorem 3.2 in CCDDHNR (2018): for any numbers a, b, δa, δb such that $|a|\vee|b|\le c$ and $|\delta a|\vee|\delta b|\le r$, it holds that $|(a+\delta a)(b+\delta b)-ab|\le 2r(c+r)$. Denote $\psi_{ij,h} := \psi_l(W_{ij};\theta_0,\eta_0)$ and $\widehat{\psi}_{ij,h} := \psi_l(W_{ij};\widetilde{\theta},\widehat{\eta}_{k\ell})$ for $h\in\{l,m\}$ and apply the above identity with $a=\psi_{ij,l}$, $b=\psi_{i'j,m}$,

$a+\delta a = \widehat{\psi}_{ij,l}$, $b+\delta b = \widehat{\psi}_{i'j,m}$, $r = |\widehat{\psi}_{ij,l} - \psi_{ij,l}| \vee |\widehat{\psi}_{i'j,m} - \psi_{i'j,m}|$ and $c = |\psi_{ij,l}| \vee |\psi_{i'j,m}|$. Then

$$\begin{aligned}
\mathcal{I}_{k\ell,lm,1} =& \Big| \frac{1}{|I|^2|J|} \sum_{i,i'\in I_k} \sum_{j\in J_\ell} \Big\{ \widehat{\psi}_{ij,l}\widehat{\psi}_{i'j,m} - \psi_{ij,l}\psi_{i'j,m} \Big\} \Big| \\
\le& \frac{1}{|I|^2|J|} \sum_{i,i'\in I_k} \sum_{j\in J_\ell} |\widehat{\psi}_{ij,l}\widehat{\psi}_{i'j,m} - \psi_{ij,l}\psi_{i'j,m}| \\
\le& \frac{2}{|I|^2|J|} \sum_{i,i'\in I_k} \sum_{j\in J_\ell} (|\widehat{\psi}_{ij,l} - \psi_{ij,l}| \vee |\widehat{\psi}_{i'j,m} - \psi_{i'j,m}|) \\
& \quad \times \Big(|\psi_{ij,l}| \vee |\psi_{i'j,m}| + |\widehat{\psi}_{ij,l} - \psi_{ij,l}| \vee |\widehat{\psi}_{i'j,m} - \psi_{i'j,m}| \Big) \\
\le& \Big(\frac{2}{|I|^2|J|} \sum_{i,i'\in I_k} \sum_{j\in J_\ell} |\widehat{\psi}_{ij,l} - \psi_{ij,l}|^2 \vee |\widehat{\psi}_{i'j,m} - \psi_{i'j,m}|^2 \Big)^{1/2} \\
& \quad \times \Big(\frac{2}{|I|^2|J|} \sum_{i,i'\in I_k} \sum_{j\in J_\ell} \Big\{ |\psi_{ij,l}| \vee |\psi_{i'j,m}| + |\widehat{\psi}_{ij,l} - \psi_{ij,l}| \vee |\widehat{\psi}_{i'j,m} - \psi_{i'j,m}| \Big\}^2 \Big)^{1/2} \\
\le& \Big(\frac{2}{|I|^2|J|} \sum_{i,i'\in I_k} \sum_{j\in J_\ell} |\widehat{\psi}_{ij,l} - \psi_{ij,l}|^2 \vee |\widehat{\psi}_{i'j,m} - \psi_{i'j,m}|^2 \Big)^{1/2} \\
& \times \Big\{ \Big(\frac{2}{|I|^2|J|} \sum_{i,i'\in I_k} \sum_{j\in J_\ell} |\psi_{ij,l}|^2 \vee |\psi_{i'j,m}|^2 \Big)^{1/2} \\
& + \Big(\frac{2}{|I|^2|J|} \sum_{i,i'\in I_k} \sum_{j\in J_\ell} |\widehat{\psi}_{ij,l} - \psi_{ij,l}|^2 \vee |\widehat{\psi}_{i'j,m} - \psi_{i'j,m}|^2 \Big)^{1/2} \Big\},
\end{aligned}$$

where the second to the last inequality follows from Cauchy-Schwarz's inequality and Minkowski's inequality. Notice that

$$\begin{aligned}
&\sum_{i,i'\in I_k} \sum_{j\in J_\ell} |\psi_{ij,l}|^2 \vee |\psi_{i'j,m}|^2 \le |I| \sum_{i=1}^N \sum_{j=1}^M \|\psi(W_{ij};\theta_0,\eta_0)\|^2, \\
&\sum_{i,i'\in I_k} \sum_{j\in J_\ell} |\widehat{\psi}_{ij,l} - \psi_{ij,l}|^2 \vee |\widehat{\psi}_{i'j,m} - \psi_{i'j,m}|^2 \le |I| \sum_{i=1}^N \sum_{j=1}^M \|\psi(W_{ij};\widetilde{\theta},\widehat{\eta}_{k\ell}) - \psi(W_{ij};\theta_0,\eta_0)\|^2.
\end{aligned}$$

Thus, the above bound for $\mathcal{I}_{k\ell,lm,1}$ implies that

$$\mathcal{I}^2_{k\ell,lm,1} \lesssim R_n \times \Big(\frac{1}{|I||J|} \sum_{(i,j)\in I_k\times J_\ell} \|\psi(W_{ij};\theta_0,\eta_0)\|^2 + R_n \Big),$$

where

$$R_n := \frac{1}{|I||J|} \sum_{(i,j)\in I_k\times J_\ell} \|\psi(W_{ij};\widetilde{\theta},\widehat{\eta}_{k\ell}) - \psi(W_{ij};\theta_0,\eta_0)\|^2.$$

Notice that

$$\frac{1}{|I||J|}\sum_{(i,j)\in I_k\times J_\ell}\|\psi(W_{ij};\theta_0,\eta_0)\|^2 = O_P(1),$$

which is implied by Markov's inequality and the calculations

$$\mathrm{E}_P\Big[\frac{1}{|I||J|}\sum_{(i,j)\in I_k\times J_\ell}\|\psi(W_{ij};\theta_0,\eta_0)\|^2\Big] = \mathrm{E}_P[\|\psi(W_{11};\theta_0,\eta_0)\|^2] \le c_1^2$$

under Assumptions 1 and 3 (ii). Finally, to bound R_n, using Assumption 2 (ii),

$$R_n \lesssim \frac{1}{|I||J|}\sum_{(i,j)\in I_k\times J_\ell}\|\psi^a(W_{ij};\widehat{\eta}_{k\ell})(\widetilde{\theta}-\theta_0)\|^2 + \frac{1}{|I||J|}\sum_{(i,j)\in I_k\times J_\ell}\|\psi(W_{ij};\theta_0,\widehat{\eta}_{k\ell}) - \psi(W_{ij};\theta_0,\eta_0)\|^2.$$

The first term on RHS is bounded by

$$\Big(\frac{1}{|I||J|}\sum_{(i,j)\in I_k\times J_\ell}\|\psi^a(W_{ij};\widehat{\eta}_{k\ell})\|^2\Big)\times\|\widetilde{\theta}-\theta_0\|^2 = O_P(1)\times O_P(\underline{C}^{-1}) = O_P(\underline{C}^{-1})$$

due to Assumption 3 (ii), Markov's inequality, and Theorem 1. Furthermore, given that $(W_{ij})_{(i,j)\in I_k^c\times J_\ell^c}$ satisfies $\widehat{\eta}_{k\ell}\in\mathcal{T}_n$,

$$\begin{aligned}
&\mathrm{E}_P\Big[\|\psi(W_{ij};\theta_0,\widehat{\eta}_{k\ell}) - \psi(W_{ij};\theta_0,\eta_0)\|^2\Big| I_k^c\times J_\ell^c\Big]\\
\le&\sup_{\eta\in\mathcal{T}_n}\mathrm{E}_P\Big[\|\psi(W_{ij};\theta_0,\eta) - \psi(W_{ij};\theta_0,\eta_0)\|^2\Big| I_k^c\times J_\ell^c\Big]\\
\le&(r_n')^2
\end{aligned}$$

due to Assumptions 1 and 3 (iii). Also, the event $\widehat{\eta}_{k\ell}\in\mathcal{T}_n$ happens with probability $1-o(1)$, we have $R_n = O_P(\underline{C}^{-1}+(r_n')^2)$. Thus we conclude that

$$\mathcal{I}_{k\ell,lm,1} = O_P(\underline{C}^{-1/2}+r_n').$$

This completes the proof. □

B Useful Lemmas

We collect some of the useful auxiliary results in this section.

First, for any $f : \text{supp}(W) \to \mathbb{R}^d$ for a fixed $d \in \mathbb{N}$, we use

$$\mathbb{G}_n f := \sqrt{\underline{C}}\Big\{\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} f(W_{ij}) - \mathrm{E}_P[f(W_{11})]\Big\}$$

to denote its multiway empirical process. The following is a multivariate version of Chiang and Sasaki (2019), Lemma 1; see also Lemma D.2 in Davezies, D'Haultfoeuille, and Guyonvarch (2018).

Lemma 1 (Independentization via Hájek Projections)**.** *If Assumption 1 holds and $f : \text{supp}(W) \to \mathbb{R}^d$ for some fixed $d \in \mathbb{N}$ and suppose $\mathrm{E}_P\|f(W_{11})\|^2 < K$ for a finite constant K that is independent of n, then there exist i.i.d. uniform random variables U_{i0} and U_{0j} such that the Hájek projection $H_n f$ of $\mathbb{G}_n f$ on*

$$\mathcal{G}_n = \Big\{\sum_{i=1}^{N} g_{i0}(U_{i0}) + \sum_{j=1}^{M} g_{0j}(U_{0j}) : g_{i0}, g_{0j} \in L^2(P_n)\Big\}$$

is equal to

$$H_n f = \frac{\sqrt{\underline{C}}}{N}\sum_{i=1}^{N} \mathrm{E}_P\Big[f(W_{i1}) - \mathrm{E}_P f(W_{11})\Big|U_{i0}\Big] + \frac{\sqrt{\underline{C}}}{M}\sum_{j=1}^{M} \mathrm{E}_P\Big[f(W_{1j}) - \mathrm{E}_P f(W_{11})\Big|U_{0j}\Big]$$

for each n. Furthermore,

$$V(\mathbb{G}_n f) = V(H_n f) + O(\underline{C}^{-1}) = \bar{\mu}_N Cov(f(W_{11}), f(W_{12})) + \bar{\mu}_M Cov(f(W_{11}), f(W_{21})) + O(\underline{C}^{-1})$$

holds a.s.

Proof. The proof is essentially the same as the proof for Lemma 1 of Chiang and Sasaki (2019) and is therefore omitted. □

The following re-states Lemma 6.1. of CCDDHNR (2018):

Lemma 2 (Conditional Convergence Implies Unconditional)**.** *Let* (X_n) *and* (Y_n) *be sequences of random vectors.*

(i) If for $\epsilon_n \to 0$*,* $P(\|X_n\| > \epsilon_n | Y_n) = o_P(1)$ *in probability, then* $P(\|X_n\| > \epsilon_n) = o(1)$*. In particular, this occurs if* $\mathrm{E}_P[\|X_n\|^q / \epsilon_n^q | Y_n] = o_P(1)$ *for some* $q \geq 1$*.*

(ii) Let (A_n) *be a sequence of positive constants. If* $\|X_n\| = O_P(A_n)$ *conditional on* Y_n*, then* $\|X_n\| = O_P(A_n)$ *unconditional, namely, for any* $l_n \to \infty$*,* $P(\|X_n\| > l_n A_n) = o(1)$*.*

C Extension to General Multiway Clustering

In this section, we extend the main results to general multiway cluster sampling framework. Notations in the current section are independent of those in the remaining parts of the paper – we introduce different notations in order to enhance the readability of the main results of the paper while economizing complicated notations in the current extension section. Consider the ℓ-way clustered data for a fixed dimension $\ell \in \mathbb{N}$. With $C_i \in \mathbb{N}$ denoting the number of clusters in the i-th cluster dimension for each $i \in \{1, ..., \ell\}$, each cell of the ℓ-way clustered sample is indexed by the ℓ-dimensional multiway cluster indices $\boldsymbol{j} = (j_1, ..., j_\ell) \in \times_{i=1}^{\ell} [C_i]$. The ℓ-dimensional size $(C_1, ..., C_\ell) \in \mathbb{N}^\ell$ of the ℓ-way clustered sample will be index by $n \in \mathbb{N}$ as $(C_1, ..., C_\ell) = (C_1(n), ..., C_\ell(n))$, where $C_i(n)$ is non-decreasing in n for each $i \in \{1, ..., \ell\}$ and $\prod_{i=1}^{\ell} C_i(n)$ is increasing in n. With this said, we will suppress the index notation and write $(C_1, ..., C_\ell)$ without n for simplicity. Also define the notations $\boldsymbol{C} = (C_1, ..., C_\ell)$, $\prod_C = \prod_{i=1}^{\ell} C_i$, $\underline{C} = \min_{1 \leq i \leq \ell} C_i$, $\overline{C} = \max_{1 \leq i \leq \ell} C_i$, and $\mu_i = \underline{C}/C_i$ for each $i \in \{1, ..., \ell\}$. Suppose that $\mu_i \to \bar{\mu}_i$ for some constant $\bar{\mu}_i$ for each $i \in \{1, ..., \ell\}$. The number of observations in the $\boldsymbol{j}$-th cell is denoted by $N_{\boldsymbol{j}}$, which is treated as an $\{0, 1, ..., \overline{N}\}$-valued random variable for some $\overline{N} \in \mathbb{N}$ not depending on n. When $[\cdot]$ takes the random variable $N_{\boldsymbol{j}}$ as an argument, we extend the definition of $[\cdot]$ to $[N_{\boldsymbol{j}}] := \{1, ..., N_{\boldsymbol{j}}\}$ if $N_{\boldsymbol{j}} \geq 1$ and $:= \emptyset$ if $N_{\boldsymbol{j}} = 0$. The observed vector for unit

$\imath \in [N_{\boldsymbol{j}}]$ in the $\boldsymbol{j}$-th cell is denoted by $W_{\imath,\boldsymbol{j}}$. Let $\{\mathcal{P}_n\}_n$ be a sequence of sets of probability laws of $(N_{\boldsymbol{j}}, (W_{\imath,\boldsymbol{j}})_{1\leq\imath\leq\overline{N}})_{\boldsymbol{j}\geq\mathbf{1}}$, where $\mathbf{1} := (1, ..., 1)$ for a short-hand notation and we write $\boldsymbol{j} \geq \boldsymbol{j}'$ for $j_i \geq j_i{}'$ for all $i \in \{1, ..., \ell\}$.

Example 1. The sampling setting in Section 3.2 fits in the current general framework with $\ell = 2$, $C_1 = N$, $C_2 = M$, and $(N_{\boldsymbol{j}}, W_{1,\boldsymbol{j}}) = (1, W_{j_1 j_2})$ for all $\boldsymbol{j} \in [N] \times [M]$ with probability one. $\square$

The econometric model has the true parameters $(\theta_0, \eta_0) \in \Theta \times T$ satisfying the score moment restriction

$$\mathrm{E}_P\Big[\sum_{\imath=1}^{N_{\mathbf{1}}} \psi(W_{\imath,\mathbf{1}}; \theta_0, \eta_0)\Big] = 0, \tag{C.1}$$

where we focus on the linear Neyman orthogonal score of the form

$$\psi(w; \theta, \eta) = \psi^a(w; \eta)\theta + \psi^b(w; \eta), \text{ for all } w \in \mathrm{supp}(W),\ \theta \in \Theta,\ \eta \in T \tag{C.2}$$

for supp(W):$= \cup_{\imath=1}^{\overline{N}} \mathrm{supp}(W_{\imath,\mathbf{1}})$, $\Theta \subset \mathbb{R}^{d_\theta}$ and a convex set T.

For a fixed integer $K > 1$, we randomly split the data into K folds in each of the ℓ cluster dimensions, resulting in K^ℓ folds in total. Specifically, randomly partition $[C_i]$ into K parts $\{I_i^1, ..., I_i^K\}$ for each $i \in \{1, ..., \ell\}$. We use the ℓ-dimensional indices $\boldsymbol{k} := (k_1, ..., k_\ell)$ to index the ℓ-way fold $I_{\boldsymbol{k}} := I_{k_1} \times \cdots \times I_{k_\ell}$ and its complementary product $I^c_{\boldsymbol{k}} := I^c_{k_1} \times \cdots \times I^c_{k_\ell}$ for each $\boldsymbol{k} \in [K]^\ell$. Let

$$\widehat{\eta}_{\boldsymbol{k}} = \widehat{\eta}(((W_{\imath,\boldsymbol{j}})_{\imath\in[N_{\boldsymbol{j}}]})_{\boldsymbol{j}\in I^c_{\boldsymbol{k}}})$$

be a machine learning estimate of η using the subsample $((W_{\imath,\boldsymbol{j}})_{\imath\in[N_{\boldsymbol{j}}]})_{\boldsymbol{j}\in I^c_{\boldsymbol{k}}}$ for each $\boldsymbol{k} \in [K]^\ell$. Let

$$\widehat{J} := \frac{1}{K^\ell} \sum_{\boldsymbol{k}\in[K]^\ell} \mathbb{E}_{n,\boldsymbol{k}}\Big[\sum_{\imath\in[N_{\boldsymbol{j}}]} \psi^a(W_{\imath,\boldsymbol{j}}; \widehat{\eta}_{\boldsymbol{k}})\Big] \qquad \text{where}$$

$$\mathbb{E}_{n,\boldsymbol{k}}\Big[\sum_{\imath\in[N_{\boldsymbol{j}}]} f(W_{\imath,\boldsymbol{j}})\Big] := \frac{1}{|I_{\boldsymbol{k}}|} \sum_{\boldsymbol{j}\in I_{\boldsymbol{k}}} \sum_{\imath\in[N_{\boldsymbol{j}}]} f(W_{\imath,\boldsymbol{j}}) \text{ for each } \boldsymbol{k} \in [K]^\ell$$

for any Borel measurable function f, the sum $\sum\limits_{\imath\in[N_{\boldsymbol{j}}]}$ is treated as zero when $N_{\boldsymbol{j}} = 0$, and $|I_{\boldsymbol{k}}| := \lfloor \frac{\prod_{i=1}^{\ell} C_i}{K^\ell} \rfloor$. With these setup and notations, the multiway DML estimator is defined by

$$\widetilde{\theta} = -\widehat{J}^{-1}\frac{1}{K^\ell}\sum_{\boldsymbol{k}\in[K]^\ell}\mathbb{E}_{n,\boldsymbol{k}}\Big[\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi^b(W_{\imath,\boldsymbol{j}};\widehat{\eta}_{\boldsymbol{k}})\Big]. \tag{C.3}$$

Let $|\underline{I_{\boldsymbol{k}}}| = \min\{|I_{k_1}|, ..., |I_{k_\ell}|\}$ for a short-hand notation. Also let $I(\boldsymbol{j})$ denote the multiway fold containing the $\boldsymbol{j}$-th multiway cluster, i.e., $I(\boldsymbol{j}) \subset \times_{i=1}^{\ell}[C_i]$ satisfies $I_{\boldsymbol{k}} = I(\boldsymbol{j})$ for some $\boldsymbol{k} \in [K]^\ell$ and $\boldsymbol{j} \in I(\boldsymbol{j})$. With these additional notations, we propose to estimate the asymptotic variance of $\sqrt{\underline{C}}(\widetilde{\theta} - \theta_0)$ by

$$\widehat{\sigma}^2 = \widehat{J}^{-1}\Big[\frac{1}{K^\ell}\sum_{\boldsymbol{k}\in[K]^\ell}\frac{|\underline{I_{\boldsymbol{k}}}|}{|I_{\boldsymbol{k}}|^2}\sum_{i=1}^{\ell}\sum_{\substack{\boldsymbol{j},\boldsymbol{j}'\in I_{\boldsymbol{k}}\\ I_i(\boldsymbol{j})=I_i(\boldsymbol{j}')}}\sum_{\imath\in[N_{\boldsymbol{j}}]}\sum_{\imath'\in[N_{\boldsymbol{j}'}]}\psi(W_{\imath,\boldsymbol{j}};\widetilde{\theta},\widehat{\eta}_{\boldsymbol{k}})\psi(W_{\imath,\boldsymbol{j}'};\widetilde{\theta},\widehat{\eta}_{\boldsymbol{k}})'\Big](\widehat{J}^{-1})'. \tag{C.4}$$

Example 1, Continued. The two-way DML in Section 3 is a special case of the current general methodological framework with $\{I_1^1, ..., I_1^K\} = \{I_1, ..., I_K\}$, $\{I_2^1, ..., I_2^K\} = \{J_1, ..., J_K\}$, $\widehat{\eta}_{(k_1,k_2)} = \widehat{\eta}((W_{j_1j_2})_{(j_1,j_2)\in([N]\setminus I_{k_1})\times([M]\setminus J_{k_2})})$, $\widehat{J} = \frac{1}{K^2}\sum_{(k_1,k_2)\in[K]^2}\mathbb{E}_{n,(k_1,k_2)}[\psi^a(W_{j_1j_2};\widehat{\eta}_{(k_1,k_2)})]$ where $\mathbb{E}_{n,(k_1,k_2)}[f(W_{j_1j_2})] = \frac{1}{|I_{k_1}||J_{k_2}|}\sum_{(j_1,j_2)\in I_{k_1}\times J_{k_2}} f(W_{j_1j_2})$, $\widetilde{\theta} = -\widehat{J}^{-1}\frac{1}{K^2}\sum_{(k_1,k_2)\in[K]^2}\mathbb{E}_{n,(k_1,k_2)}\Big[\psi^b(W_{j_1j_2};\widehat{\eta}_{(k_1,k_2)})\Big]$, and $\widehat{\sigma}^2 = \widehat{J}^{-1}\widehat{\Gamma}(\widehat{J}^{-1})'$ where

$$\begin{aligned}\widehat{\Gamma} = \frac{1}{K^2}\sum_{(k_1,k_2)\in[K]^2}\Big\{&\frac{|I_{k_1}|\wedge|J_{k_2}|}{(|I_{k_1}||J_{k_2}|)^2}\sum_{j_1\in I_{k_1}}\sum_{j_2,j_2'\in J_{k_2}}\psi(W_{j_1j_2};\widetilde{\theta},\widehat{\eta}_{(k_1,k_2)})\psi(W_{j_1j_2'};\widetilde{\theta},\widehat{\eta}_{(k_1,k_2)})'\\ &+\frac{|I_{k_1}|\wedge|J_{k_2}|}{(|I_{k_1}||J_{k_2}|)^2}\sum_{j_1,j_1'\in I_{k_1}}\sum_{j_2\in J_{k_2}}\psi(W_{j_1j_2};\widetilde{\theta},\widehat{\eta}_{(k_1,k_2)})\psi(W_{j_1'j_2};\widetilde{\theta},\widehat{\eta}_{(k_1,k_2)})'\Big\}.\end{aligned}$$

□

We now state assumptions under which (C.4) is an asymptotically valid variance estimator for $\sqrt{\underline{C}}(\widetilde{\theta} - \theta_0)$ with the multiway DML estimator (C.3). We write $a \lesssim b$ to mean $a \le cb$ for some $c > 0$ that does not depend on n. We also write $a \lesssim_P b$ to mean $a = O_P(b)$. For any finite dimensional vector v, $\|v\|$ denotes the ℓ_2 or Euclidean norm of v. For any matrix A, $\|A\|$ denotes the induced ℓ_2-norm of the matrix. The following assumption concerns the multiway clustered sampling.

Assumption 1 (Sampling)**.** The following conditions hold for each n.

(i) The array $(N_{\boldsymbol{j}}, (W_{\imath,\boldsymbol{j}})_{1\leq\imath\leq\overline{N}})_{\boldsymbol{j}\geq\mathbf{1}}$ is an infinite sequence of separately exchangeable random vector. That is, for any ℓ-tuple of permutations $(\pi_1, ..., \pi_\ell)$ of $\mathbb{N}$, we have

$$(N_{\boldsymbol{j}}, (W_{\imath,\boldsymbol{j}})_{1\leq\imath\leq\overline{N}})_{\boldsymbol{j}\geq\mathbf{1}} \stackrel{d}{=} (N_{\pi_1(j_1),...,\pi_\ell(j_\ell)}, (W_{\imath,\pi_1(j_1),...,\pi_\ell(j_\ell)})_{1\leq\imath\leq\overline{N}})_{\boldsymbol{j}\geq\mathbf{1}}.$$

(ii) $(N_{\boldsymbol{j}}, (W_{\imath,\boldsymbol{j}})_{1\leq\imath\leq\overline{N}})_{\boldsymbol{j}\geq\mathbf{1}}$ is dissociated. That is, for any $\mathbf{c} \geq \mathbf{1}$, $(N_{\boldsymbol{j}}, (W_{\imath,\boldsymbol{j}})_{1\leq\imath\leq\overline{N}})_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{c}}$ is independent of $(N_{\boldsymbol{j}'}, (W_{\imath',\boldsymbol{j}'})_{1\leq\imath'\leq\overline{N}})_{\boldsymbol{j}'\geq\boldsymbol{c}+\mathbf{1}}$

(iii) $E(N_{\mathbf{1}}) > 0$ and $N_{\boldsymbol{j}} \leq \overline{N}$ for each $\mathbf{1} \leq \boldsymbol{j} \leq \boldsymbol{C}$, where $\overline{N} \in \mathbb{N}$ does not depend on n.

(iv) The econometrician observes $(N_{\boldsymbol{j}}, (W_{\imath,\boldsymbol{j}})_{1\leq\imath\leq N_{\boldsymbol{j}}})_{\mathbf{1}\leq\boldsymbol{j}\leq\boldsymbol{C}}$.

Remark 1. The dependence among $(W_{\imath,\boldsymbol{j}})_{\imath\geq1}$ in each cell $\boldsymbol{j}$ is left unrestricted in this assumption. Assumption 1 is similar to Assumption 1 of Davezies et al. (2018), except for $\overline{N}$. We introduce $\overline{N}$ to simplify some concentration arguments.

Let $c_0 > 0$, $c_1 > 0$, $s > 0$, $q \geq 4$ be some finite constants with $c_0 \leq c_1$. Let $\{\delta_n\}_{n\geq1}$ (estimation errors) and $\{\Delta_n\}_{n\geq1}$ (probability bounds) be sequences of positive constants that converge to zero such that $\delta_n \geq \underline{C}^{-1/2}$. Let $K \geq 2$ be a fixed integer. Let $(N_{\mathbf{0}}, (W_{\imath,\mathbf{0}})_{0\leq\imath\leq\overline{N}})$ denote an independent copy of $(N_{\mathbf{1}}, (W_{\imath,\mathbf{1}})_{1\leq\imath\leq\overline{N}})$ and therefore is independent from the data and the random set $\mathcal{T}_n$ of nuisance realization. With these notations, we state the following assumptions for the model.

Assumption 2 (Linear Neyman Orthogonal Score)**.** For all $\underline{C} \geq 3$ and $P \in \mathcal{P}_n$, the following conditions hold.

(i) The true parameter value θ_0 satisfies (C.1).

(ii) ψ is linear in the sense that it satisfies (C.2).

(iii) The map $\eta \mapsto \mathrm{E}_P\Big[\sum_{\imath\in[N_{\mathbf{0}}]} \psi(W_{\imath,\mathbf{0}};\theta,\eta)\Big]$ is twice continuously Gateaux differentiable on T.

(iv) ψ satisfies the Neyman near orthogonality condition at (θ_0, η_0) as

$$\lambda_n := \sup_{\eta\in\mathcal{T}_n} \Big\|\partial_\eta \mathrm{E}_P\Big[\sum_{\imath\in[N_{\mathbf{0}}]} \psi(W_{\imath,\mathbf{0}};\theta_0,\eta_0)[\eta-\eta_0]\Big]\Big\| \le \delta_n \underline{C}^{-1/2}.$$

(v) The identification condition holds as the singular values of the matrix $J_0 := \mathrm{E}_P\Big[\sum_{\imath\in[N_{\mathbf{0}}]} \psi^a(W_{\imath,\mathbf{0}};\eta_0)\Big]$ are between c_0 and c_1.

Assumption 3 (Score Regularity and Nuisance Parameter Estimators)**.** For all $\underline{C} \ge 3$ and $P \in \mathcal{P}_n$, the following conditions hold.

(i) The realization set $\mathcal{T}_n$ contains η_0, and the nuisance parameter estimator $\widehat{\eta}_{\boldsymbol{k}} = \widehat{\eta}((W_{\imath,\boldsymbol{j}})_{\imath\in[N_{\boldsymbol{j}}]})_{\boldsymbol{j}\in I_{\boldsymbol{k}}^c}$ belongs to the realization set $\mathcal{T}_n$ for each $\boldsymbol{k} \in [K]^\ell$ with probability at least $1 - \Delta_n$.

(ii) The following moment conditions hold:

$$m_n := \sup_{\eta\in\mathcal{T}_n} (\mathrm{E}_P\Big[\Big\|\sum_{\imath\in[N_{\mathbf{0}}]} \psi(W_{\imath,\mathbf{0}};\theta_0,\eta)\Big\|^q\Big])^{1/q} \le c_1,$$

$$m_n' := \sup_{\eta\in\mathcal{T}_n} (\mathrm{E}_P\Big[\Big\|\sum_{\imath\in[N_{\mathbf{0}}]} \psi^a(W_{\imath,\mathbf{0}};\eta)\Big\|^q\Big])^{1/q} \le c_1.$$

(iii) The following conditions on the rates r_n, r_n' and λ_n' hold:

$$r_n := \sup_{\eta\in\mathcal{T}_n} \Big\|\mathrm{E}_P\Big[\sum_{\imath\in[N_{\mathbf{0}}]} \psi^a(W_{\imath,\mathbf{0}};\eta)\Big] - \mathrm{E}_P\Big[\sum_{\imath\in[N_{\mathbf{0}}]} \psi^a(W_{\imath,\mathbf{0}};\eta_0)\Big]\Big\| \le \delta_n,$$

$$r_n' := \sup_{\eta\in\mathcal{T}_n} \Big(\Big\|\mathrm{E}_P\Big[\sum_{\imath\in[N_{\mathbf{0}}]} \psi(W_{\imath,\mathbf{0}};\theta_0,\eta)\Big] - \mathrm{E}_P\Big[\sum_{\imath\in[N_{\mathbf{0}}]} \psi(W_{\imath,\mathbf{0}};\theta_0,\eta_0)\Big]\Big\|^2\Big)^{1/2} \le \delta_n,$$

$$\lambda_n' = \sup_{r\in(0,1),\eta\in\mathcal{T}_n} \Big\|\partial_r^2 \mathrm{E}_P\Big[\sum_{\imath\in[N_{\mathbf{0}}]} \psi(W_{\imath,\mathbf{0}};\theta_0,\eta_0 + r(\eta-\eta_0))\Big]\Big\| \le \delta_n/\sqrt{\underline{C}}.$$

(iv) All eigenvalues of the matrix

$$\Gamma := \sum_{i=1}^{\ell} \bar{\mu}_i \Gamma_i = \sum_{i=1}^{\ell} \bar{\mu}_i \mathrm{E}_P\left[\sum_{\imath=1}^{N_{\mathbf{1}}}\sum_{\imath'=1}^{N_{2_{\mathbf{i}}}} \psi(W_{\imath,\mathbf{1}};\theta_0,\eta_0)\psi(W_{\imath',2_i};\theta_0,\eta_0)'\right]$$

are bounded from below by c_0, where 2_i denotes the ℓ-tuple vector with 2 in each entry but for 1 in the i-th entry.

The following theorems generalize Theorems 1 and 2 to cover general ℓ-way cluster sampling. Their proofs are contained in Section D.

Theorem 1 (Main Result)**.** *Suppose that Assumptions 1, 2 and 3 are satisfied. If* $\delta_n \geq \underline{C}^{-1/2}$ *for all* $\underline{C} \geq 1$*, then*

$$\sqrt{\underline{C}}\sigma^{-1}(\widetilde{\theta} - \theta_0) = \frac{\sqrt{\underline{C}}}{\prod_C} \sum_{j_1=1}^{C_1} \cdots \sum_{j_\ell=1}^{C_\ell} \sum_{\imath \in [N_{\mathbf{j}}]} \bar{\psi}(W_{\imath,\mathbf{j}}) + O_P(\rho_n) \rightsquigarrow N(0, I_{d_\theta})$$

holds uniformly for all $P \in \mathcal{P}_n$*, where* $\prod_C = \prod_{i=1}^{\ell} C_i$*, the influence function takes the form* $\bar{\psi}(\cdot) := -\sigma^{-1} J_0^{-1} \psi(\cdot; \theta_0, \eta_0)$*, the size of the remainder terms follows*

$$\rho_n := \underline{C}^{-1/2} + r_n + r_n' + \underline{C}^{1/2}\lambda_n + \underline{C}^{1/2}\lambda_n' \lesssim \delta_n,$$

and the asymptotic variance is given by

$$\sigma^2 := J_0^{-1}\Gamma(J_0^{-1})'. \tag{C.5}$$

Theorem 2 (Variance Estimator)**.** *Under the assumptions required by Theorem 1, we have*

$$\widehat{\sigma}^2 = \sigma^2 + O_P(\rho_n).$$

Furthermore, the statement of Theorem 1 holds true with $\widehat{\sigma}^2$ *in place of* σ^2*.*

D Proofs of the Extended Results

D.1 Proof of Theorem 1

Proof. Let $\mathcal{E}_n$ denote the event $\widehat{\eta}_{(k_1,\ldots,k_\ell)} \in \mathcal{T}_n$ for all $(k_1, \ldots k_\ell) \in [K]^\ell$ and define $\boldsymbol{k} := (k_1, \ldots, k_\ell)$. Assumption 3 (i) implies $P_n(\mathcal{E}_n) \geq 1 - K^\ell \Delta_n$. Let $\mathbf{e} \in \{0,1\}^\ell$, and define $\mathcal{A}_{\mathbf{e}} := \{(\boldsymbol{j}, \boldsymbol{j}') : \mathbf{1} \leq$

$\boldsymbol{j}, \boldsymbol{j}' \leq \mathbf{C} : \forall i = 1, ..., \ell, e_i = 1 \Leftrightarrow j_i = j_i'\}$, and $\boldsymbol{\varepsilon}_m := \{\mathbf{e} \in \{0,1\}^\ell : \sum_{i'=1}^{\ell} e_{i'} = m\}$.

Step 1. This is the main step showing linear representation and asymptotic normality for the proposed estimator. Denote

$$\widehat{J} := \frac{1}{K^\ell} \sum_{\boldsymbol{k} \in [K]^\ell} \mathbb{E}_{n,\boldsymbol{k}} \Big[\sum_{\imath \in [N_{\boldsymbol{j}}]} \psi^a(W_{\imath,\boldsymbol{j}}; \widehat{\eta}_{\boldsymbol{k}}) \Big], \qquad R_{n,1} := \widehat{J} - J_0,$$

$$R_{n,2} := \frac{1}{K^\ell} \sum_{\boldsymbol{k} \in [K]^\ell} \mathbb{E}_{n,\boldsymbol{k}} \Big[\sum_{\imath \in [N_{\boldsymbol{j}}]} \psi(W_{\imath,\boldsymbol{j}}; \theta_0, \widehat{\eta}_{\boldsymbol{k}}) \Big] - \frac{1}{\prod_C} \sum_{j_1=1}^{C_1} ... \sum_{j_\ell=1}^{C_\ell} \sum_{\imath \in [N_{\boldsymbol{j}}]} \psi(W_{\imath,\boldsymbol{j}}; \theta_0, \eta_0).$$

We will later show in Steps 2, 3, 4 and 5, respectively, that

$$\|R_{n,1}\| = O_{P_n}(\underline{C}^{-1/2} + r_n), \tag{D.1}$$

$$\|R_{n,2}\| = O_{P_n}(\underline{C}^{-1/2} r_n' + \lambda_n + \lambda_n'), \tag{D.2}$$

$$\Big\| \sqrt{\underline{C}} \frac{1}{\prod_C} \sum_{j_1=1}^{C_1} ... \sum_{j_\ell=1}^{C_\ell} \sum_{\imath \in [N_{\boldsymbol{j}}]} \psi(W_{\imath,\boldsymbol{j}}; \theta_0, \eta_0) \Big\| = O_{P_n}(1), \tag{D.3}$$

$$\|\sigma^{-1}\| = O_{P_n}(1). \tag{D.4}$$

Then, under Assumptions 2 and 3, $\underline{C}^{-1/2} + r_N \leq \rho_n = o(1)$ and all singular values of J_0 are bounded away from zero. Therefore, with P_n-probability at least $1 - o(1)$, all singular values of $\widehat{J}$ are bounded away from zero. Thus with the same P_n probability, the multiway DML solution is uniquely written as

$$\widetilde{\theta} = -\widehat{J}^{-1} \frac{1}{K^\ell} \sum_{\boldsymbol{k} \in [K]^\ell} \mathbb{E}_{n,\boldsymbol{k}} \Big[\sum_{\imath \in [N_{\boldsymbol{j}}]} \psi^b(W_{\imath,\boldsymbol{j}}; \widehat{\eta}_{\boldsymbol{k}}) \Big],$$

and

$$\begin{aligned}
\sqrt{\underline{C}}(\widetilde{\theta} - \theta_0) &= - \sqrt{\underline{C}} \widehat{J}^{-1} \frac{1}{K^\ell} \sum_{\boldsymbol{k} \in [K]^\ell} \Big(\mathbb{E}_{n,\boldsymbol{k}} \Big[\sum_{\imath \in [N_{\boldsymbol{j}}]} \psi^b(W_{\imath,\boldsymbol{j}}; \widehat{\eta}_{\boldsymbol{k}}) \Big] + \widehat{J}\theta_0 \Big) \\
&= - \sqrt{\underline{C}} \widehat{J}^{-1} \frac{1}{K^\ell} \sum_{\boldsymbol{k} \in [K]^\ell} \mathbb{E}_{n,\boldsymbol{k}} \Big[\sum_{\imath \in [N_{\boldsymbol{j}}]} \psi(W_{\imath,\boldsymbol{j}}; \theta_0, \widehat{\eta}_{\boldsymbol{k}}) \Big] \\
&= - \Big(J_0 + R_{n,1} \Big)^{-1} \times \Big(\frac{\sqrt{\underline{C}}}{\prod_C} \sum_{j_1=1}^{C_1} ... \sum_{j_\ell=1}^{C_\ell} \sum_{\imath \in [N_{\boldsymbol{j}}]} \psi(W_{\imath,\boldsymbol{j}}; \theta_0, \eta_0) + \sqrt{\underline{C}} R_{n,2} \Big).
\end{aligned} \tag{D.5}$$

Using the fact that

$$\Big(J_0+R_{n,1}\Big)^{-1}-J_0^{-1}=-(J_0+R_{n,1})^{-1}R_{n,1}J_0^{-1},$$

we have

$$\begin{aligned}\|(J_0+R_{n,1})^{-1}-J_0^{-1}\|=&\|(J_0+R_{n,1})^{-1}R_{n,1}J_0^{-1}\|\le\|(J_0+R_{n,1})^{-1}\|\,\|R_{n,1}\|\,\|J_0^{-1}\|\\=&O_{P_n}(1)O_{P_n}(\underline{C}^{-1/2}+r_n)O_{P_n}(1)=O_{P_n}(\underline{C}^{-1/2}+r_n).\end{aligned}$$

Furthermore, $r_n'+\sqrt{\underline{C}}(\lambda_n+\lambda_n')\le\rho_n=o(1)$, it holds that

$$\begin{aligned}&\Big\|\frac{\sqrt{\underline{C}}}{\prod_C}\sum_{j_1=1}^{C_1}...\sum_{j_\ell=1}^{C_\ell}\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi(W_{\imath,\boldsymbol{j}};\theta_0,\eta_0)+\sqrt{\underline{C}}R_{n,2}\Big\|\\\le&\Big\|\frac{\sqrt{\underline{C}}}{\prod_C}\sum_{j_1=1}^{C_1}...\sum_{j_\ell=1}^{C_\ell}\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi(W_{\imath,\boldsymbol{j}};\theta_0,\eta_0)\Big\|+\Big\|\sqrt{\underline{C}}R_{n,2}\Big\|\\=&O_{P_n}(1)+o_{P_n}(1)=O_{P_n}(1),\end{aligned}$$

where the first equality is due to (D.3) and (D.4). Combining above two bounds gives

$$\begin{aligned}&\Big\|\Big(J_0+R_{n,1}\Big)^{-1}-J_0^{-1}\Big\|\times\Big\|\frac{\sqrt{\underline{C}}}{\prod_C}\sum_{j_1=1}^{C_1}...\sum_{j_\ell=1}^{C_\ell}\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi(W_{\imath,\boldsymbol{j}};\theta_0,\eta_0)+\sqrt{\underline{C}}R_{n,2}\Big\|\\=&O_{P_n}(\underline{C}^{-1/2}+r_n)O_{P_n}(1)\\=&O_{P_n}(\underline{C}^{-1/2}+r_n).\end{aligned}\tag{D.6}$$

Therefore, from (D.4), (D.5) and (D.6), we have

$$\sqrt{\underline{C}}\sigma^{-1}(\widetilde{\theta}-\theta_0)=\frac{\sqrt{\underline{C}}}{\prod_C}\sum_{j_1=1}^{C_1}...\sum_{j_\ell=1}^{C_\ell}\sum_{\imath\in[N_{\boldsymbol{j}}]}\bar{\psi}(W_{\imath,\boldsymbol{j}})+O_{P_n}(\rho_n).$$

The first term on the RHS above can be written as $\mathbb{G}_n\bar{\psi}$. Applying Lemma 3, we obtain the independent linear representation

$$H_n\bar{\psi}:=\sum_{j_1=1}^{C_1}\frac{\sqrt{\underline{C}}}{C_1}\mathrm{E}_{P_n}\Big[\sum_{\imath\in[N_{\boldsymbol{j}}]}\bar{\psi}(W_{\imath,\boldsymbol{j}})\Big|U_{j_1,0...0}\Big]+...+\sum_{j_\ell=1}^{C_\ell}\frac{\sqrt{\underline{C}}}{C_\ell}\mathrm{E}_{P_n}\Big[\sum_{\imath\in[N_{\boldsymbol{j}}]}\bar{\psi}(W_{\imath,\boldsymbol{j}})\Big|U_{0...0,j_\ell}\Big]$$

and it holds P_n-a.s. that

$$\begin{aligned}V_n(\mathbb{G}_n\bar{\psi}) =& V_n(H_n\bar{\psi}) + O(\underline{C}^{-1}) = J_0^{-1}\Gamma(J_0^{-1})' + O(\underline{C}^{-1}) \qquad \text{and}\\ \mathbb{G}_n\bar{\psi} =& H_n\bar{\psi} + O_P(\underline{C}^{-1/2}),\end{aligned}$$

where $V_n(\cdot) = \mathrm{E}_{P_n}[(\cdot - \mathrm{E}_{P_n}[\cdot])^2]$. Under Assumption 3 (iv). Recall that $q \geq 4$, the third moments of both summands of $H_n\bar{\psi}$ are bounded over n under Assumptions 2(v) and 3 (ii) (iv). We have verified all the conditions for Lyapunov's CLT. An application of Lyapunov's CLT and Cramer-Wold device gives

$$H_n\bar{\psi} \rightsquigarrow N(0, I_{d_\theta})$$

and an application of Theorem 2.7 of van der Vaart (1998) concludes the proof.

Step 2. Since K is fixed, it suffices to show for any $\boldsymbol{k} \in [K]^\ell$,

$$\Big\|\mathbb{E}_{n,\boldsymbol{k}}\Big[\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi^a(W_{\imath,\boldsymbol{j}};\widehat{\eta}_{\boldsymbol{k}})\Big] - \mathrm{E}_P\Big[\sum_{\imath\in[N_{\mathbf{0}}]}\psi^a(W_{\imath,\mathbf{0}};\eta_0)\Big]\Big\| = O_{P_n}(\underline{C}^{-1/2} + r_n).$$

Fix $\boldsymbol{k} \in [K]^\ell$,

$$\Big\|\mathbb{E}_{n,\boldsymbol{k}}\Big[\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi^a(W_{\imath,\boldsymbol{j}};\widehat{\eta}_{\boldsymbol{k}})\Big] - \mathrm{E}_{P_n}\Big[\sum_{\imath\in[N_{\mathbf{0}}]}\psi^a(W_{\imath,\mathbf{0}};\eta_0)\Big]\Big\| \leq \mathcal{I}_{1,\boldsymbol{k}} + \mathcal{I}_{2,\boldsymbol{k}},$$

where

$$\begin{aligned}\mathcal{I}_{1,\boldsymbol{k}} :=& \Big\|\mathbb{E}_{n,\boldsymbol{k}}\Big[\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi^a(W_{\imath,\boldsymbol{j}};\widehat{\eta}_{\boldsymbol{k}})\Big] - \mathrm{E}_{P_n}\Big[\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi^a(W_{\imath,\boldsymbol{j}};\widehat{\eta}_{\boldsymbol{k}})\Big| I^c_{k_1}\times...\times I^c_{k_\ell}\Big]\Big\|,\\ \mathcal{I}_{2,\boldsymbol{k}} :=& \Big\|\mathrm{E}_{P_n}\Big[\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi^a(W_{\imath,\boldsymbol{j}};\widehat{\eta}_{\boldsymbol{k}})\Big| I^c_{k_1}\times...\times I^c_{k_\ell}\Big] - \mathrm{E}_{P_n}\Big[\sum_{\imath\in[N_{\mathbf{0}}]}\psi^a(W_{\imath,\mathbf{0}};\eta_0)\Big]\Big\|.\end{aligned}$$

Notice that $\mathcal{I}_{2,\boldsymbol{k}} \leq r_n$ with P_n-probability $1 - o(1)$ follows directly from Assumptions 1 (ii) and 3 (iii). Now denote $\widetilde{\psi}^a_{\boldsymbol{j},m} = \sum_{\imath\in[N_{\boldsymbol{j}}]}\psi^a_m(W_{\imath,\boldsymbol{j}};\widehat{\eta}_{\boldsymbol{k}}) - \mathrm{E}_{P_n}\Big[\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi^a_m(W_{\imath,\boldsymbol{j}};\widehat{\eta}_{\boldsymbol{k}})\Big| I^c_{k_1}\times...\times I^c_{k_\ell}\Big]$ and $\widetilde{\psi}^a_{\boldsymbol{j}} = (\widetilde{\psi}^a_{\boldsymbol{j},m})_{m\in[d_\theta]}$, and $|\underline{I_{\boldsymbol{k}}}| = \min\{|I_{k_1}|,...,|I_{k_\ell}|\}$. Let us denote $I_{\boldsymbol{k}} := (I_{k_1}\times...\times I_{k_\ell})$ and $I^c_{\boldsymbol{k}} := (I^c_{k_1}\times...\times I^c_{k_\ell})$. Let $\boldsymbol{j} \mapsto I(\boldsymbol{j}) \in \mathcal{I}$, and define $\mathcal{B}_{\mathbf{e}} := \{(\boldsymbol{j},\boldsymbol{j}') : \forall i = 1,...,\ell, e_i = 1 \Leftrightarrow I_i(\boldsymbol{j}) = I_i(\boldsymbol{j}') :$

$\boldsymbol{j},\boldsymbol{j}'\in\mathcal{I}\}$, where $\mathcal{I}:=\{I_1^1,...,I_1^K\}\times...\times\{I_\ell^1,...,I_\ell^K\}$, and $\boldsymbol{\epsilon}_m:=\{\mathbf{e}\in\{0,1\}^\ell:\sum_{i'=1}^{\ell}e_{i'}=m\}$.

To bound $\mathcal{I}_{1,\boldsymbol{k}}$, note that conditional on $I_{\boldsymbol{k}}^c$, it holds that

$$
\begin{aligned}
\mathrm{E}_{P_n}[\mathcal{I}_{1,\boldsymbol{k}}^2|I_{\boldsymbol{k}}^c]=&\mathrm{E}_{P_n}\Big[\Big\|\mathbb{E}_{n,\boldsymbol{k}}\Big[\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi^a(W_{\imath,\boldsymbol{j}};\widehat{\eta}_{\boldsymbol{k}})\Big]-\mathrm{E}_{P_n}\Big[\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi^a(W_{\imath,\boldsymbol{j}};\widehat{\eta}_{\boldsymbol{k}})\Big|I_{\boldsymbol{k}}^c\Big]\Big\|^2\Big|I_{\boldsymbol{k}}^c\Big]\\
=&\frac{1}{|I_{\boldsymbol{k}}|^2}\mathrm{E}_{P_n}\Big[\sum_{m=1}^{d_\theta}\Big(\sum_{\boldsymbol{j}\in I_{\boldsymbol{k}}}\widetilde{\psi}_{\boldsymbol{j},m}^a\Big)^2\Big|I_{\boldsymbol{k}}^c\Big]\\
=&\frac{1}{|I_{\boldsymbol{k}}|^2}\sum_{\mathbf{e}\in\boldsymbol{\epsilon}_1}\sum_{(\boldsymbol{j}',\boldsymbol{j})\in\mathcal{B}_{\mathbf{e}}}\mathrm{E}_{P_n}\Big[\sum_{m=1}^{d_\theta}\widetilde{\psi}_{\boldsymbol{j},m}^a\widetilde{\psi}_{\boldsymbol{j}',m}^a\Big|I_{\boldsymbol{k}}^c\Big]\\
&+\frac{1}{|I_{\boldsymbol{k}}|^2}\sum_{r=2}^{\ell}\sum_{\mathbf{e}\in\boldsymbol{\epsilon}_r}\sum_{(\boldsymbol{j}',\boldsymbol{j})\in\mathcal{B}_{\mathbf{e}}}\mathrm{E}_{P_n}\Big[\sum_{m=1}^{d_\theta}\widetilde{\psi}_{\boldsymbol{j},m}^a\widetilde{\psi}_{\boldsymbol{j}',m}^a\Big|I_{\boldsymbol{k}}^c\Big]\\
&+\frac{1}{|I_{\boldsymbol{k}}|^2}\sum_{\mathbf{e}\in\boldsymbol{\epsilon}_0}\sum_{(\boldsymbol{j}',\boldsymbol{j})\in\mathcal{B}_{\mathbf{e}}}\mathrm{E}_{P_n}\Big[\sum_{m=1}^{d_\theta}\widetilde{\psi}_{\boldsymbol{j},m}^a\widetilde{\psi}_{\boldsymbol{j}',m}^a\Big|I_{\boldsymbol{k}}^c\Big]\\
=&\frac{1}{|I_{\boldsymbol{k}}|^2}\sum_{\mathbf{e}\in\boldsymbol{\epsilon}_1}\sum_{(\boldsymbol{j}',\boldsymbol{j})\in\mathcal{B}_{\mathbf{e}}}\mathrm{E}_{P_n}[\langle\widetilde{\psi}_{\boldsymbol{j}}^a,\widetilde{\psi}_{\boldsymbol{j}'}^a\rangle|I_{\boldsymbol{k}}^c]+R+0\\
\lesssim&\frac{1}{|\underline{I_{\boldsymbol{k}}}|}\mathrm{E}_{P_n}\Big[\Big\|\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi^a(W_{\imath,\boldsymbol{j}};\widehat{\eta}_{\boldsymbol{k}})-\mathrm{E}_{P_n}\Big[\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi^a(W_{\imath,\boldsymbol{j}};\widehat{\eta}_{\boldsymbol{k}})\Big|I_{\boldsymbol{k}}^c\Big]\Big\|^2\Big|I_{\boldsymbol{k}}^c\Big]\\
\le&\frac{1}{|\underline{I_{\boldsymbol{k}}}|}\mathrm{E}_{P_n}\Big[\Big\|\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi^a(W_{\imath,\boldsymbol{j}};\widehat{\eta}_{\boldsymbol{k}})\Big\|^2\Big|I_{\boldsymbol{k}}^c\Big]\le\frac{c_1^2}{|\underline{I_{\boldsymbol{k}}}|}.
\end{aligned}
$$

In the third equility, the last term corresponds to the covariance between cells sharing no common cluster. By independence, the last term is zero. Let us denote the second term in the third equality by R. Under Cauchy-Schwarz inequality and Assumption 1 (ii),

$$|R|\le\frac{1}{|I_{\boldsymbol{k}}|^2}\sum_{\mathbf{e}\in\cup_{l=2}^{\ell}\boldsymbol{\epsilon}_l}|\mathcal{B}_{\mathbf{e}}|\mathrm{E}_{P_n}\Big[\sum_{m=1}^{d_\theta}(\widetilde{\psi}_{\boldsymbol{j},m}^a)^2\Big|I_{\boldsymbol{k}}^c\Big]. \tag{D.7}$$

For $r\ge1$ and $\mathbf{e}\in\boldsymbol{\epsilon}_r$, we have

$$|\mathcal{B}_{\mathbf{e}}|=|I_{\boldsymbol{k}}|\times\prod_{i:e_i=0}(|I_{k_i}|-1). \tag{D.8}$$

Therefore, $R=O(|\underline{I_{\boldsymbol{k}}}|^{-2})$. Note that $\underline{C}\lesssim|\underline{I_{\boldsymbol{k}}}|\lesssim\underline{C}$. Hence an application of Lemma 2 (i) implies $\mathcal{I}_{1,\boldsymbol{k}}=O_{P_n}(\underline{C}^{-1/2})$. This completes a proof of (D.1).

Step 3. It again suffices to show that for any $\boldsymbol{k} \in [K]^\ell$, one has

$$\Big\|\mathbb{E}_{n,\boldsymbol{k}}\Big[\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi(W_{\imath,\boldsymbol{j}};\theta_0,\widehat{\eta}_{\boldsymbol{k}})\Big] - \frac{1}{|I_{\boldsymbol{k}}|}\sum_{\boldsymbol{j}\in I_{\boldsymbol{k}}}\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi(W_{\imath,\boldsymbol{j}};\theta_0,\eta_0)\Big\| = O_{P_n}(\underline{C}^{-1/2}r_n' + \lambda_n + \lambda_n').$$

Denote

$$\mathbb{G}_{n,\boldsymbol{k}}\Big[\sum_{\imath\in[N_{\boldsymbol{j}}]}\phi(W_{\imath,\boldsymbol{j}})\Big] = \frac{\sqrt{\underline{C}}}{|I_{\boldsymbol{k}}|}\sum_{\boldsymbol{j}\in I_{\boldsymbol{k}}}\sum_{\imath\in[N_{\boldsymbol{j}}]}\Big(\phi(W_{\imath,\boldsymbol{j}}) - \int\phi(w)dP_n\Big).$$

Then

$$\Big\|\mathbb{E}_{n,\boldsymbol{k}}\Big[\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi(W_{\imath,\boldsymbol{j}};\theta_0,\widehat{\eta}_{\boldsymbol{k}})\Big] - \frac{1}{|I_{\boldsymbol{k}}|}\sum_{\boldsymbol{j}\in I_{\boldsymbol{k}}}\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi(W_{\imath,\boldsymbol{j}};\theta_0,\eta_0)\Big\| \le \frac{\mathcal{I}_{3,\boldsymbol{k}}+\mathcal{I}_{4,\boldsymbol{k}}}{\sqrt{\underline{C}}},$$

where

$$\begin{aligned}
\mathcal{I}_{3,\boldsymbol{k}} :=& \Big\|\mathbb{G}_{n,\boldsymbol{k}}\Big[\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi(W_{\imath,\boldsymbol{j}};\theta_0,\widehat{\eta}_{\boldsymbol{k}})\Big] - \mathbb{G}_{n,\boldsymbol{k}}\Big[\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi(W_{\imath,\boldsymbol{j}};\theta_0,\eta_0)\Big]\Big\|,\\
\mathcal{I}_{4,\boldsymbol{k}} :=& \sqrt{\underline{C}}\Big\|\mathrm{E}_{P_n}\Big[\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi(W_{\imath,\boldsymbol{j}};\theta_0,\widehat{\eta}_{\boldsymbol{k}})\Big|I_{\boldsymbol{k}}\Big] - \mathrm{E}_{P_n}\Big[\sum_{\imath\in[N_{\boldsymbol{0}}]}\psi(W_{\imath,\boldsymbol{0}};\theta_0,\eta_0)\Big]\Big\|.
\end{aligned}$$

Denote $\widetilde{\psi}_{\boldsymbol{j},m} := \sum_{\imath\in[N_{\boldsymbol{j}}]}\psi_m(W_{\imath,\boldsymbol{j}};\theta_0,\widehat{\eta}_{\boldsymbol{k}}) - \sum_{\imath\in[N_{\boldsymbol{j}}]}\psi_m(W_{\imath,\boldsymbol{j}};\theta_0,\eta_0)$ and $\widetilde{\psi}_{\boldsymbol{j}} = (\widetilde{\psi}_{\boldsymbol{j},m})_{m\in[d_\theta]}$. To bound $\mathcal{I}_{3,\boldsymbol{k}}$, notice that using a similar argument as for the bound of $\mathcal{I}_{1,\boldsymbol{k}}$, one has

$$\begin{aligned}
&\mathrm{E}_{P_n}[\|\mathcal{I}_{3,\boldsymbol{k}}\|^2|I_{\boldsymbol{k}}^c]\\
=&\mathrm{E}_{P_n}\Big[\Big\|\mathbb{G}_{n,\boldsymbol{k}}\Big[\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi(W_{\imath,\boldsymbol{j}};\theta_0,\widehat{\eta}_{\boldsymbol{k}}) - \sum_{\imath\in[N_{\boldsymbol{j}}]}\psi(W_{\imath,\boldsymbol{j}};\theta_0,\eta_0)\Big]\Big\|^2\Big|I_{\boldsymbol{k}}^c\Big]\\
=&\mathrm{E}_{P_n}\Big[\frac{\underline{C}}{|I_{\boldsymbol{k}}|^2}\sum_{m=1}^{d_\theta}\Big\{\sum_{\boldsymbol{j}\in I_{\boldsymbol{k}}}\Big(\widetilde{\psi}_{\boldsymbol{j},m} - \mathrm{E}_{P_n}\widetilde{\psi}_{\boldsymbol{j},m}\Big)\Big\}^2\Big|I_{\boldsymbol{k}}^c\Big]\\
=&\frac{\underline{C}}{|I_{\boldsymbol{k}}|^2}\sum_{\mathbf{e}\in\boldsymbol{\epsilon}_1}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{B}_{\mathbf{e}}}\mathrm{E}_{P_n}\Big[\sum_{m=1}^{d_\theta}\Big(\widetilde{\psi}_{\boldsymbol{j},m} - \mathrm{E}_{P_n}\widetilde{\psi}_{\boldsymbol{j},m}\Big)\Big(\widetilde{\psi}_{\boldsymbol{j}',m} - \mathrm{E}_{P_n}\widetilde{\psi}_{\boldsymbol{j}',m}\Big)\Big|I_{\boldsymbol{k}}^c\Big]\\
&+\frac{\underline{C}}{|I_{\boldsymbol{k}}|^2}\sum_{r=2}^{\ell}\sum_{\mathbf{e}\in\boldsymbol{\epsilon}_r}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{B}_{\mathbf{e}}}\mathrm{E}_{P_n}\Big[\sum_{m=1}^{d_\theta}\Big(\widetilde{\psi}_{\boldsymbol{j},m} - \mathrm{E}_{P_n}\widetilde{\psi}_{\boldsymbol{j},m}\Big)\Big(\widetilde{\psi}_{\boldsymbol{j}',m} - \mathrm{E}_{P_n}\widetilde{\psi}_{\boldsymbol{j}',m}\Big)\Big|I_{\boldsymbol{k}}^c\Big]\\
&+\frac{\underline{C}}{|I_{\boldsymbol{k}}|^2}\sum_{\mathbf{e}\in\boldsymbol{\epsilon}_0}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{B}_{\mathbf{e}}}\mathrm{E}_{P_n}\Big[\sum_{m=1}^{d_\theta}\Big(\widetilde{\psi}_{\boldsymbol{j},m} - \mathrm{E}_{P_n}\widetilde{\psi}_{\boldsymbol{j},m}\Big)\Big(\widetilde{\psi}_{\boldsymbol{j}',m} - \mathrm{E}_{P_n}\widetilde{\psi}_{\boldsymbol{j}',m}\Big)\Big|I_{\boldsymbol{k}}^c\Big]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\underline{C}}{|I_{\boldsymbol{k}}|^2} \sum_{\mathbf{e}\in\boldsymbol{\epsilon}_1} \sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{B}_{\mathbf{e}}} \mathrm{E}_{P_n}\Big[\langle\widetilde{\psi}_{\boldsymbol{j}} - \mathrm{E}_{P_n}\widetilde{\psi}_{\boldsymbol{j}}, \widetilde{\psi}_{\boldsymbol{j}'} - \mathrm{E}_{P_n}\widetilde{\psi}_{\boldsymbol{j}'}\rangle\Big| I_{\boldsymbol{k}}^c\Big] + R' + 0 \\
&\lesssim \mathrm{E}_{P_n}\Big[\Big\| \sum_{\imath\in[N_{\boldsymbol{j}}]} \psi(W_{\imath,\boldsymbol{j}};\theta_0,\widehat{\eta}) - \sum_{\imath\in[N_{\boldsymbol{j}}]} \psi(W_{\imath,\boldsymbol{j}};\theta_0,\eta_0) - \mathrm{E}_{P_n}\Big[\sum_{\imath\in[N_{\boldsymbol{j}}]} \psi(W_{\imath,\boldsymbol{j}};\theta_0,\widehat{\eta}) - \sum_{\imath\in[N_{\boldsymbol{j}}]} \psi(W_{\imath,\boldsymbol{j}};\theta_0,\eta_0)\Big]\Big\|^2 \Big| I_{\boldsymbol{k}}^c\Big] \\
&\leq \mathrm{E}_{P_n}\Big[\Big\| \sum_{\imath\in[N_{\boldsymbol{j}}]} \psi(W_{\imath,\boldsymbol{j}};\theta_0,\widehat{\eta}) - \sum_{\imath\in[N_{\boldsymbol{j}}]} \psi(W_{\imath,\boldsymbol{j}};\theta_0,\eta_0)\Big\|^2 \Big| I_{\boldsymbol{k}}^c\Big] \\
&\leq \sup_{\eta\in\mathcal{T}_n} \mathrm{E}_{P_n}\Big[\Big\| \sum_{\imath\in[N_{\mathbf{0}}]} \psi(W_{\imath,\mathbf{0}};\theta_0,\eta) - \sum_{\imath\in[N_{\mathbf{0}}]} \psi(W_{\imath,\mathbf{0}};\theta_0,\eta_0)\Big\|^2 \Big| I_{\boldsymbol{k}}^c\Big] \\
&= \sup_{\eta\in\mathcal{T}_n} \mathrm{E}_{P_n}\Big[\Big\| \sum_{\imath\in[N_{\mathbf{0}}]} \psi(W_{\imath,\mathbf{0}};\theta_0,\eta) - \sum_{\imath\in[N_{\mathbf{0}}]} \psi(W_{\imath,\mathbf{0}};\theta_0,\eta_0)\Big\|^2\Big] = (r_n')^2,
\end{aligned}
$$

where the first inequality follows from Cauchy-Schwarz's inequality, the second-to-last equality is due to Assumption 1, and the last equality is due to Assumption 3 (iii). Using the similar argument for R, we have $R' = O(\underline{C}^{-1})$.

Hence, $\mathcal{I}_{3,\boldsymbol{k}} = O_{P_n}(r_n')$. To bound $\mathcal{I}_{4,\boldsymbol{k}}$, let

$$
f_{\boldsymbol{k}}(r) := \mathrm{E}_{P_n}\Big[\sum_{\imath\in[N_{\boldsymbol{j}}]} \psi(W_{\imath,\boldsymbol{j}};\theta_0,\eta_0 + r(\widehat{\eta}_{\boldsymbol{k}} - \eta_0))\Big| I_{\boldsymbol{k}}^c\Big] - \mathrm{E}_{P_n}\Big[\sum_{\imath\in[N_{\mathbf{0}}]} \psi(W_{\imath,\mathbf{0}};\theta_0,\eta_0)\Big], \qquad r\in[0,1].
$$

An application of the mean value expansion coordinate-wise gives

$$
f_{\boldsymbol{k}}(1) = f_{\boldsymbol{k}}(0) + f_{\boldsymbol{k}}'(0) + f_{\boldsymbol{k}}''(\widetilde{r})/2,
$$

where $\widetilde{r}\in(0,1)$. Note that $f_{\boldsymbol{k}}(0) = 0$ under Assumption 2 (i), and

$$
\|f_{\boldsymbol{k}}'(0)\| = \Big\|\partial_\eta \mathrm{E}_{P_n}\Big[\sum_{\imath\in[N_{\boldsymbol{j}}]} \psi(W;\theta_0,\eta_0)[\widehat{\eta}_{\boldsymbol{k}} - \eta_0]\Big]\Big\| \leq \lambda_n
$$

under Assumption 2 (iv). Moreover, under Assumption 3 (iii), on the event $\mathcal{E}_n$, we have

$$
\|f_{\boldsymbol{k}}''(\widetilde{r})\| \leq \sup_{r\in(0,1)} \|f_{\boldsymbol{k}}''(r)\| \leq \lambda_n'.
$$

This completes a proof of (D.2).

Step 4. Note that

$$
\mathrm{E}_{P_n}\Big[\Big\|\frac{\sqrt{\underline{C}}}{\prod_C} \sum_{j_1=1}^{C_1} \dots \sum_{j_\ell=1}^{C_\ell} \sum_{\imath\in[N_{\boldsymbol{j}}]} \psi(W_{\imath,\boldsymbol{j}};\theta_0,\eta_0)\Big\|^2\Big]
$$

$$
\begin{aligned}
=&\frac{C}{\prod_C^2}\mathrm{E}_{P_n}\Big[\sum_{m=1}^{d_\theta}\Big(\sum_{j_1=1}^{C_1}\ldots\sum_{j_\ell=1}^{C_\ell}\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi_m(W_{\imath,\boldsymbol{j}};\theta_0,\eta_0)\Big)^2\Big]\\
=&\frac{C}{\prod_C^2}\sum_{\mathbf{e}\in\boldsymbol{\varepsilon}_1}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{A}_{\mathbf{e}}}\mathrm{E}_{P_n}\Big[\sum_{m=1}^{d_\theta}\sum_{\imath\in[N_{\boldsymbol{j}}]}\sum_{\imath'\in[N_{\boldsymbol{j}'}]}\psi_m(W_{\imath,\boldsymbol{j}};\theta_0,\eta_0)\psi_m(W_{\imath',\boldsymbol{j}'};\theta_0,\eta_0)\Big]\\
&+\frac{C}{\prod_C^2}\sum_{r=2}^{\ell}\sum_{\mathbf{e}\in\boldsymbol{\varepsilon}_r}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{A}_{\mathbf{e}}}\mathrm{E}_{P_n}\Big[\sum_{m=1}^{d_\theta}\sum_{\imath\in[N_{\boldsymbol{j}}]}\sum_{\imath'\in[N_{\boldsymbol{j}'}]}\psi_m(W_{\imath,\boldsymbol{j}};\theta_0,\eta_0)\psi_m(W_{\imath',\boldsymbol{j}'};\theta_0,\eta_0)\Big]\\
&+\frac{C}{\prod_C^2}\sum_{\mathbf{e}\in\boldsymbol{\varepsilon}_0}\sum_{(\boldsymbol{j},\boldsymbol{j}')\in\mathcal{A}_{\mathbf{e}}}\mathrm{E}_{P_n}\Big[\sum_{m=1}^{d_\theta}\sum_{\imath\in[N_{\boldsymbol{j}}]}\sum_{\imath'\in[N_{\boldsymbol{j}'}]}\psi_m(W_{\imath,\boldsymbol{j}};\theta_0,\eta_0)\psi_m(W_{\imath',\boldsymbol{j}'};\theta_0,\eta_0)\Big]\\
\lesssim&\mathrm{E}_{P_n}\Big[\Big\|\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi(W_{\imath,\boldsymbol{j}};\theta_0,\eta_0)\Big\|^2\Big]\le c_1^2.
\end{aligned}
$$

Step 5. Note that all singular values of J_0 are bounded from above by c_1 under Assumption 2 (v) and all eigenvalues of Γ are bounded from below by c_0 under Assumption 3 (iv). Therefore, we have $\|\sigma^{-1}\|\le c_1/\sqrt{c_0}$ and thus $\|\sigma^{-1}\| = O_{P_n}(1)$. This completes a proof of (D.4). $\square$

D.2 Proof of Theorem 2

Proof. Step 2 of the proof of Theorem 1 proves $\|\widehat{J}-J_0\| = O_p(\underline{C}^{-1/2}+r_n)$ and Assumption 2 (v) implies $\|J_0^{-1}\|\le c_0^{-1}$. Therefore, to prove the claim of the theorem, it suffices to show

$$
\begin{aligned}
\Big\|&\frac{1}{K^\ell}\sum_{\boldsymbol{k}\in[K]^\ell}\frac{|\underline{I_{\boldsymbol{k}}}|}{|I_{\boldsymbol{k}}|^2}\sum_{i=1}^{\ell}\sum_{\substack{\boldsymbol{j},\boldsymbol{j}'\in I_{\boldsymbol{k}}\\ I_i(\boldsymbol{j})=I_i(\boldsymbol{j}')}}\sum_{\imath\in[N_{\boldsymbol{j}}]}\sum_{\imath'\in[N_{\boldsymbol{j}'}]}\psi(W_{\imath,\boldsymbol{j}};\widetilde{\theta},\widehat{\eta}_{\boldsymbol{k}})\psi(W_{\imath',\boldsymbol{j}'};\widetilde{\theta},\widehat{\eta}_{\boldsymbol{k}})'\\
&-\sum_{i=1}^{\ell}\bar{\mu}_i\mathrm{E}_P\Big[\sum_{\imath=1}^{N_{\mathbf{1}}}\sum_{\imath'=1}^{N_{2_i}}\psi(W_{\imath,\mathbf{1}};\theta_0,\eta_0)\psi(W_{\imath',2_i};\theta_0,\eta_0)'\Big]\ \Big\| = O_P(\rho_n).
\end{aligned}
$$

Moreover, since K and d_θ are constants and $\mu_i\to\bar{\mu}_i\le 1$, it suffices to show that for each $\boldsymbol{k}\in[K]^\ell$ and $l,m\in[d_\theta]$, it holds that

$$
\Big|\frac{|\underline{I_{\boldsymbol{k}}}|}{|I_{\boldsymbol{k}}|^2}\sum_{\substack{\boldsymbol{j},\boldsymbol{j}'\in I_{\boldsymbol{k}}\\ I_i(\boldsymbol{j})=I_i(\boldsymbol{j}')}}\sum_{\imath\in[N_{\boldsymbol{j}}]}\sum_{\imath'\in[N_{\boldsymbol{j}'}]}\psi_l(W_{\imath,\boldsymbol{j}};\widetilde{\theta},\widehat{\eta}_{\boldsymbol{k}})\psi_m(W_{\imath',\boldsymbol{j}'};\widetilde{\theta},\widehat{\eta}_{\boldsymbol{k}})-\mu_i\mathrm{E}_P\Big[\sum_{\imath=1}^{N_{\mathbf{1}}}\sum_{\imath'=1}^{N_{2_i}}\psi_l(W_{\imath,\mathbf{1}};\theta_0,\eta_0)\psi_m(W_{\imath',2_i};\theta_0,\eta_0)\Big]\Big|
$$

$$
= O_P(\rho_n).
$$

Denote the left-hand side of the equation as $\mathcal{I}_{\boldsymbol{k},lm}$. First, note that $|\underline{I}|/|I_{k_i}| = \mu_i$. We denote i' for I_{k_i} such that $|I_{k_{i'}}| = |\underline{I_{\boldsymbol{k}}}|$, and apply the triangle inequality to get

$$\mathcal{I}_{\boldsymbol{k},lm} \leq \mathcal{I}_{\boldsymbol{k},lm,1} + \mathcal{I}_{\boldsymbol{k},lm,2},$$

where

$$\begin{aligned}\mathcal{I}_{\boldsymbol{k},lm,1} := \Big| \frac{1}{\prod_{i\neq i'} |I_{k_i}|^2 |I_{k_{i'}}|} \sum_{\substack{\boldsymbol{j},\boldsymbol{j}'\in I_{\boldsymbol{k}} \\ I_i(\boldsymbol{j})=I_i(\boldsymbol{j}')}} \Big\{ \sum_{\imath\in[N_{\boldsymbol{j}}]} \sum_{\imath'\in[N_{\boldsymbol{j}'}]} \psi_l(W_{\imath,\boldsymbol{j}}; \widetilde{\theta}, \widehat{\eta}_{\boldsymbol{k}}) \psi_m(W_{\imath',\boldsymbol{j}'}; \widetilde{\theta}, \widehat{\eta}_{\boldsymbol{k}}) \\ - \sum_{\imath\in[N_{\boldsymbol{j}}]} \sum_{\imath'\in[N_{\boldsymbol{j}'}]} \psi_l(W_{\imath,\boldsymbol{j}}; \theta_0, \eta_0) \psi_m(W_{\imath',\boldsymbol{j}'}; \theta_0, \eta_0) \Big\} \Big|,\end{aligned}$$

$$\begin{aligned}\mathcal{I}_{\boldsymbol{k},lm,2} := \Big| \frac{1}{\prod_{i\neq i'} |I_{k_i}|^2 |I_{k_{i'}}|} \sum_{\substack{\boldsymbol{j},\boldsymbol{j}'\in I_{\boldsymbol{k}} \\ I_i(\boldsymbol{j})=I_i(\boldsymbol{j}')}} \sum_{\imath\in[N_{\boldsymbol{j}}]} \sum_{\imath'\in[N_{\boldsymbol{j}'}]} \psi_l(W_{\imath,\boldsymbol{j}}; \theta_0, \eta_0) \psi_m(W_{\imath',\boldsymbol{j}'}; \theta_0, \eta_0) \\ - \mathrm{E}_P[\sum_{\imath=1}^{N_{\boldsymbol{1}}} \sum_{\imath'=1}^{N_{2_i}} \psi_l(W_{\imath,\boldsymbol{1}}; \theta_0, \eta_0) \psi_m(W_{\imath',2_i}; \theta_0, \eta_0)] \Big|.\end{aligned}$$

We first find a bound for $\mathcal{I}_{\boldsymbol{k},lm,2}$. Since $q > 4$, it holds that

$$\begin{aligned}\mathrm{E}_P[\mathcal{I}^2_{\boldsymbol{k},lm,2}] =& \frac{1}{\prod_{i\neq i'} |I_{k_i}|^4 |I_{k_{i'}}|^2} \mathrm{E}_P\Big[\Big| \sum_{\substack{\boldsymbol{j},\boldsymbol{j}'\in I_{\boldsymbol{k}} \\ I_i(\boldsymbol{j})=I_i(\boldsymbol{j}')}} \sum_{\imath\in[N_{\boldsymbol{j}}]} \sum_{\imath'\in[N_{\boldsymbol{j}'}]} \psi_l(W_{\imath,\boldsymbol{j}}; \theta_0, \eta_0) \psi_m(W_{\imath',\boldsymbol{j}'}; \theta_0, \eta_0) \\ &- \mathrm{E}_P\Big[\sum_{\imath=1}^{N_{\boldsymbol{1}}} \sum_{\imath'=1}^{N_{2_i}} \psi_l(W_{\imath,\boldsymbol{1}}; \theta_0, \eta_0) \psi_m(W_{\imath',2_i}; \theta_0, \eta_0)\Big]\Big|^2\Big] \\ \leq& \frac{1}{\prod_{i\neq i'} |I_{k_i}|^4 |I_{k_{i'}}|^2} \mathrm{E}_P\Big[\sum_{\substack{\boldsymbol{j},\boldsymbol{j}',\boldsymbol{j}'',\boldsymbol{j}'''\in I_{\boldsymbol{k}} \\ I_i(\boldsymbol{j})=I_i(\boldsymbol{j}'), I_i(\boldsymbol{j}'')=I_i(\boldsymbol{j}''')}} \sum_{\substack{I_s(\boldsymbol{j})=I_s(\boldsymbol{j}'') \\ s\neq i}} \sum_{\imath\in[N_{\boldsymbol{j}}]} \sum_{\imath'\in[N_{\boldsymbol{j}'}]} \sum_{\imath''\in[N_{\boldsymbol{j}''}]} \sum_{\imath'''\in[N_{\boldsymbol{j}'''}]} \\ & \psi_l(W_{\imath,\boldsymbol{j}}; \theta_0, \eta_0) \psi_m(W_{\imath',\boldsymbol{j}'}; \theta_0, \eta_0) \psi_l(W_{\imath'',\boldsymbol{j}''}; \theta_0, \eta_0) \psi_m(W_{\imath''',\boldsymbol{j}'''}; \theta_0, \eta_0)\Big] \\ &+ \frac{1}{\prod_{i\neq i'} |I_{k_i}|^4 |I_{k_{i'}}|^2} \mathrm{E}_P\Big[\sum_{\substack{\boldsymbol{j},\boldsymbol{j}',\boldsymbol{j}'',\boldsymbol{j}'''\in I_{\boldsymbol{k}} \\ I_i(\boldsymbol{j})=I_i(\boldsymbol{j}')=I_i(\boldsymbol{j}'')=I_i(\boldsymbol{j}''')}} \sum_{\imath\in[N_{\boldsymbol{j}}]} \sum_{\imath'\in[N_{\boldsymbol{j}'}]} \sum_{\imath''\in[N_{\boldsymbol{j}''}]} \sum_{\imath'''\in[N_{\boldsymbol{j}'''}]} \\ & \psi_l(W_{\imath,\boldsymbol{j}}; \theta_0, \eta_0) \psi_m(W_{\imath',\boldsymbol{j}'}; \theta_0, \eta_0) \psi_l(W_{\imath'',\boldsymbol{j}''}; \theta_0, \eta_0) \psi_m(W_{\imath''',\boldsymbol{j}'''}; \theta_0, \eta_0)\Big] \\ &+ o(|\underline{I_{\boldsymbol{k}}}|^{-1}) + 0 \\ \lesssim& \frac{1}{|\underline{I_{\boldsymbol{k}}}|} \mathrm{E}_P\Big[\Big\| \sum_{\imath\in[N_{\boldsymbol{0}}]} \psi(W_{\imath,\boldsymbol{0}}; \theta_0, \eta_0) \Big\|^4\Big] \lesssim c_1^4/\underline{C} = O(\underline{C}^{-1}).\end{aligned}$$

Now, to bound $\mathcal{I}_{\boldsymbol{k},lm,1}$, we make use of the following identity coming from the proof of Theorem 3.2 in CCDDHNR (2018): for any numbers a, b, δa, δb such that $|a| \vee |b| \le c$ and $|\delta a| \vee |\delta b| \le r$, it holds that $|(a+\delta a)(b+\delta b) - ab| \le 2r(c+r)$. Denote $\psi_{\boldsymbol{j},h} := \psi_l(W_{\imath,\boldsymbol{j}}; \theta_0, \eta_0)$ and $\widehat{\psi}_{\boldsymbol{j},h} := \psi_l(W_{\imath,\boldsymbol{j}}; \widetilde{\theta}, \widehat{\eta}_{\boldsymbol{k}})$ for $h \in \{l, m\}$ and apply the above identity with $a = \sum_{\imath \in [N_{\boldsymbol{j}}]} \psi_{\boldsymbol{j},l}$, $b = \sum_{\imath' \in [N_{\boldsymbol{j}'}]} \psi_{\boldsymbol{j}',m}$, $a + \delta a = \sum_{\imath \in [N_{\boldsymbol{j}}]} \widehat{\psi}_{\boldsymbol{j},l}$, $b + \delta b = \sum_{\imath' \in [N_{\boldsymbol{j}'}]} \widehat{\psi}_{\boldsymbol{j}',m}$, $r = \Big| \sum_{\imath \in [N_{\boldsymbol{j}}]} \widehat{\psi}_{\boldsymbol{j},l} - \sum_{\imath \in [N_{\boldsymbol{j}}]} \psi_{\boldsymbol{j},l} \Big| \vee \Big| \sum_{\imath' \in [N_{\boldsymbol{j}'}]} \widehat{\psi}_{\boldsymbol{j}',m} - \sum_{\imath' \in [N_{\boldsymbol{j}'}]} \psi_{\boldsymbol{j}',m} \Big|$ and $c = \Big| \sum_{\imath \in [N_{\boldsymbol{j}}]} \psi_{\boldsymbol{j},l} \Big| \vee \Big| \sum_{\imath' \in [N_{\boldsymbol{j}'}]} \psi_{\boldsymbol{j}',m} \Big|$. Then

$$
\begin{aligned}
\mathcal{I}_{\boldsymbol{k},lm,1} =& \Big| \frac{1}{\prod_{i \neq i'} |I_{k_i}|^2 |I_{k_{i'}}|} \sum_{\substack{\boldsymbol{j},\boldsymbol{j}' \in I_{\boldsymbol{k}} \\ I_i(\boldsymbol{j}) = I_i(\boldsymbol{j}')}} \Big\{ \sum_{\imath \in [N_{\boldsymbol{j}}]} \sum_{\imath' \in [N_{\boldsymbol{j}'}]} \widehat{\psi}_{\boldsymbol{j},l} \widehat{\psi}_{\boldsymbol{j}',m} - \sum_{\imath \in [N_{\boldsymbol{j}}]} \sum_{\imath' \in [N_{\boldsymbol{j}'}]} \psi_{\boldsymbol{j},l} \psi_{\boldsymbol{j}',m} \Big\} \Big| \\
\le& \frac{1}{\prod_{i \neq i'} |I_{k_i}|^2 |I_{k_{i'}}|} \sum_{\substack{\boldsymbol{j},\boldsymbol{j}' \in I_{\boldsymbol{k}} \\ I_i(\boldsymbol{j}) = I_i(\boldsymbol{j}')}} \Big| \sum_{\imath \in [N_{\boldsymbol{j}}]} \sum_{\imath' \in [N_{\boldsymbol{j}'}]} \widehat{\psi}_{\boldsymbol{j},l} \widehat{\psi}_{\boldsymbol{j}',m} - \sum_{\imath \in [N_{\boldsymbol{j}}]} \sum_{\imath' \in [N_{\boldsymbol{j}'}]} \psi_{\boldsymbol{j},l} \psi_{\boldsymbol{j}',m} \Big| \\
\le& \frac{2}{\prod_{i \neq i'} |I_{k_i}|^2 |I_{k_{i'}}|} \sum_{\substack{\boldsymbol{j},\boldsymbol{j}' \in I_{\boldsymbol{k}} \\ I_i(\boldsymbol{j}) = I_i(\boldsymbol{j}')}} \Big(\Big| \sum_{\imath \in [N_{\boldsymbol{j}}]} \widehat{\psi}_{\boldsymbol{j},l} - \sum_{\imath \in [N_{\boldsymbol{j}}]} \psi_{\boldsymbol{j},l} \Big| \vee \Big| \sum_{\imath' \in [N_{\boldsymbol{j}'}]} \widehat{\psi}_{\boldsymbol{j}',m} - \sum_{\imath' \in [N_{\boldsymbol{j}'}]} \psi_{\boldsymbol{j}',m} \Big| \Big) \\
&\times \Big(\Big| \sum_{\imath \in [N_{\boldsymbol{j}}]} \psi_{\boldsymbol{j},l} \Big| \vee \Big| \sum_{\imath' \in [N_{\boldsymbol{j}'}]} \psi_{\boldsymbol{j}',m} \Big| + \Big| \sum_{\imath \in [N_{\boldsymbol{j}}]} \widehat{\psi}_{\boldsymbol{j},l} - \sum_{\imath \in [N_{\boldsymbol{j}}]} \psi_{\boldsymbol{j},l} \Big| \vee \Big| \sum_{\imath' \in [N_{\boldsymbol{j}'}]} \widehat{\psi}_{\boldsymbol{j}',m} - \sum_{\imath' \in [N_{\boldsymbol{j}'}]} \psi_{\boldsymbol{j}',m} \Big| \Big) \\
\le& \Big(\frac{2}{\prod_{i \neq i'} |I_{k_i}|^2 |I_{k_{i'}}|} \sum_{\substack{\boldsymbol{j},\boldsymbol{j}' \in I_{\boldsymbol{k}} \\ I_i(\boldsymbol{j}) = I_i(\boldsymbol{j}')}} \Big| \sum_{\imath \in [N_{\boldsymbol{j}}]} \widehat{\psi}_{\boldsymbol{j},l} - \sum_{\imath \in [N_{\boldsymbol{j}}]} \psi_{\boldsymbol{j},l} \Big|^2 \vee \Big| \sum_{\imath' \in [N_{\boldsymbol{j}'}]} \widehat{\psi}_{\boldsymbol{j}',m} - \sum_{\imath' \in [N_{\boldsymbol{j}'}]} \psi_{\boldsymbol{j}',m} \Big|^2 \Big)^{1/2} \\
&\times \Big(\frac{2}{\prod_{i \neq i'} |I_{k_i}|^2 |I_{k_{i'}}|} \sum_{\substack{\boldsymbol{j},\boldsymbol{j}' \in I_{\boldsymbol{k}} \\ I_i(\boldsymbol{j}) = I_i(\boldsymbol{j}')}} \Big\{ \Big| \sum_{\imath \in [N_{\boldsymbol{j}}]} \psi_{\boldsymbol{j},l} \Big| \vee \Big| \sum_{\imath' \in [N_{\boldsymbol{j}'}]} \psi_{\boldsymbol{j}',m} \Big| \\
&\qquad\qquad + \Big| \sum_{\imath \in [N_{\boldsymbol{j}}]} \widehat{\psi}_{\boldsymbol{j},l} - \sum_{\imath \in [N_{\boldsymbol{j}}]} \psi_{\boldsymbol{j},l} \Big| \vee \Big| \sum_{\imath' \in [N_{\boldsymbol{j}'}]} \widehat{\psi}_{\boldsymbol{j}',m} - \sum_{\imath' \in [N_{\boldsymbol{j}'}]} \psi_{\boldsymbol{j}',m} \Big| \Big\}^2 \Big)^{1/2} \\
\le& \Big(\frac{2}{\prod_{i \neq i'} |I_{k_i}|^2 |I_{k_{i'}}|} \sum_{\substack{\boldsymbol{j},\boldsymbol{j}' \in I_{\boldsymbol{k}} \\ I_i(\boldsymbol{j}) = I_i(\boldsymbol{j}')}} \Big| \sum_{\imath \in [N_{\boldsymbol{j}}]} \widehat{\psi}_{\boldsymbol{j},l} - \sum_{\imath \in [N_{\boldsymbol{j}}]} \psi_{\boldsymbol{j},l} \Big|^2 \vee \Big| \sum_{\imath' \in [N_{\boldsymbol{j}'}]} \widehat{\psi}_{\boldsymbol{j}',m} - \sum_{\imath' \in [N_{\boldsymbol{j}'}]} \psi_{\boldsymbol{j}',m} \Big|^2 \Big)^{1/2} \\
&\times \Big\{ \Big(\frac{2}{\prod_{i \neq i'} |I_{k_i}|^2 |I_{k_{i'}}|} \sum_{\substack{\boldsymbol{j},\boldsymbol{j}' \in I_{\boldsymbol{k}} \\ I_i(\boldsymbol{j}) = I_i(\boldsymbol{j}')}} \Big| \sum_{\imath \in [N_{\boldsymbol{j}}]} \psi_{\boldsymbol{j},l} \Big|^2 \vee \Big| \sum_{\imath' \in [N_{\boldsymbol{j}'}]} \psi_{\boldsymbol{j}',m} \Big|^2 \Big)^{1/2}
\end{aligned}
$$

$$+\Big(\frac{2}{\prod_{i\neq i'}|I_{k_i}|^2|I_{k_{i'}}|}\sum_{\substack{\boldsymbol{j},\boldsymbol{j}'\in I_{\boldsymbol{k}}\\ I_i(\boldsymbol{j})=I_i(\boldsymbol{j}')}}\Big|\sum_{\imath\in[N_{\boldsymbol{j}}]}\widehat{\psi}_{\boldsymbol{j},l}-\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi_{\boldsymbol{j},l}\Big|^2\vee\Big|\sum_{\imath'\in[N_{\boldsymbol{j}'}]}\widehat{\psi}_{\boldsymbol{j}',m}-\sum_{\imath'\in[N_{\boldsymbol{j}'}]}\psi_{\boldsymbol{j}',m}\Big|^2\Big)^{1/2}\Big\},$$

where the second to the last inequality follows from Cauchy-Schwarz's inequality and Minkowski's inequality. Notice that

$$\sum_{\substack{\boldsymbol{j},\boldsymbol{j}'\in I_{\boldsymbol{k}}\\ I_i(\boldsymbol{j})=I_i(\boldsymbol{j}')}}\Big|\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi_{\boldsymbol{j},l}\Big|^2\vee\Big|\sum_{\imath'\in[N_{\boldsymbol{j}'}]}\psi_{\boldsymbol{j}',m}\Big|^2\leq\max_{1\leq i\leq\ell}\{|I_{k_i}|\}\sum_{j_1=1}^{C_1}...\sum_{j_\ell=1}^{C_\ell}\Big\|\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi(W_{\imath,\boldsymbol{j}};\theta_0,\eta_0)\Big\|^2,$$

$$\begin{aligned}&\sum_{\substack{\boldsymbol{j},\boldsymbol{j}'\in I_{\boldsymbol{k}}\\ I_i(\boldsymbol{j})=I_i(\boldsymbol{j}')}}\Big|\sum_{\imath\in[N_{\boldsymbol{j}}]}\widehat{\psi}_{\boldsymbol{j},l}-\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi_{\boldsymbol{j},l}\Big|^2\vee\Big|\sum_{\imath'\in[N_{\boldsymbol{j}'}]}\widehat{\psi}_{\boldsymbol{j}',m}-\sum_{\imath'\in[N_{\boldsymbol{j}'}]}\psi_{\boldsymbol{j}',m}\Big|^2\\ &\qquad\leq\max_{1\leq i\leq\ell}\{|I_{k_i}|\}\sum_{j_1=1}^{C_1}...\sum_{j_\ell=1}^{C_\ell}\Big\|\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi(W_{\imath,\boldsymbol{j}};\widetilde{\theta},\widehat{\eta}_{\boldsymbol{k}})-\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi(W_{\imath,\boldsymbol{j}};\theta_0,\eta_0)\Big\|^2.\end{aligned}$$

Thus, the above bound for $\mathcal{I}_{\boldsymbol{k},lm,1}$ implies that

$$\mathcal{I}^2_{\boldsymbol{k},lm,1}\lesssim R_n\times\Big(\frac{1}{|I_{\boldsymbol{k}}|}\sum_{\boldsymbol{j}\in I_{\boldsymbol{k}}}\Big\|\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi(W_{\imath,\boldsymbol{j}};\theta_0,\eta_0)\Big\|^2+R_n\Big),$$

where

$$R_n:=\frac{1}{|I_{\boldsymbol{k}}|}\sum_{\boldsymbol{j}\in I_{\boldsymbol{k}}}\Big\|\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi(W_{\imath,\boldsymbol{j}};\widetilde{\theta},\widehat{\eta}_{\boldsymbol{k}})-\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi(W_{\imath,\boldsymbol{j}};\theta_0,\eta_0)\Big\|^2.$$

Notice that

$$\frac{1}{|I_{\boldsymbol{k}}|}\sum_{\boldsymbol{j}\in I_{\boldsymbol{k}}}\Big\|\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi(W_{\imath,\boldsymbol{j}};\theta_0,\eta_0)\Big\|^2=O_P(1),$$

which is implied by Markov's inequality and the calculations

$$\mathrm{E}_P\Big[\frac{1}{|I_{\boldsymbol{k}}|}\sum_{\boldsymbol{j}\in I_{\boldsymbol{k}}}\Big\|\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi(W_{\imath,\boldsymbol{j}};\theta_0,\eta_0)\Big\|^2\Big]=\mathrm{E}_P\Big[\Big\|\sum_{\imath=1}^{N_{\boldsymbol{0}}}\psi(W_{\imath,\boldsymbol{0}};\theta_0,\eta_0)\Big\|^2\Big]\leq c_1^2$$

under Assumptions 1 and 3 (ii). Finally, to bound R_n, using Assumption 2 (ii),

$$\begin{aligned}R_n\lesssim&\frac{1}{|I_{\boldsymbol{k}}|}\sum_{\boldsymbol{j}\in I_{\boldsymbol{k}}}\Big\|\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi^a(W_{\imath,\boldsymbol{j}};\widehat{\eta}_{\boldsymbol{k}})(\widetilde{\theta}-\theta_0)\Big\|^2\\ &+\frac{1}{|I_{\boldsymbol{k}}|}\sum_{\boldsymbol{j}\in I_{\boldsymbol{k}}}\Big\|\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi(W_{\imath,\boldsymbol{j}};\theta_0,\widehat{\eta}_{\boldsymbol{k}})-\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi(W_{\imath,\boldsymbol{j}};\theta_0,\eta_0)\Big\|^2.\end{aligned}$$

The first term on RHS is bounded by

$$\Big(\frac{1}{|I_{\boldsymbol{k}}|}\sum_{\boldsymbol{j}\in I_{\boldsymbol{k}}}\Big\|\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi^a(W_{\imath,\boldsymbol{j}};\widehat{\eta}_{\boldsymbol{k}})\Big\|^2\Big)\times\|\widetilde{\theta}-\theta_0\|^2=O_P(1)\times O_P(\underline{C}^{-1})=O_P(\underline{C}^{-1})$$

due to Assumption 3 (ii), Markov's inequality, and Theorem 1. Furthermore, given that $(W_{\imath,\boldsymbol{j}})_{\boldsymbol{j}\in I^c_{\boldsymbol{k}}}$ satisfies $\widehat{\eta}_{\boldsymbol{k}}\in\mathcal{T}_n$,

$$\begin{aligned}&\mathrm{E}_P\Big[\Big\|\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi(W_{\imath,\boldsymbol{j}};\theta_0,\widehat{\eta}_{\boldsymbol{k}})-\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi(W_{\imath,\boldsymbol{j}};\theta_0,\eta_0)\Big\|^2\Big|I^c_{\boldsymbol{k}}\Big]\\&\le\sup_{\eta\in\mathcal{T}_n}\mathrm{E}_P\Big[\Big\|\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi(W_{\imath,\boldsymbol{j}};\theta_0,\eta)-\sum_{\imath\in[N_{\boldsymbol{j}}]}\psi(W_{\imath,\boldsymbol{j}};\theta_0,\eta_0)\Big\|^2\Big|I^c_{\boldsymbol{k}}\Big]\le(r'_n)^2\end{aligned}$$

due to Assumptions 1 and 3 (iii). Also, the event $\widehat{\eta}_{\boldsymbol{k}}\in\mathcal{T}_n$ happens with probability $1-o(1)$, we have $R_n=O_P(\underline{C}^{-1}+(r'_n)^2)$. Thus we conclude that

$$\mathcal{I}_{\boldsymbol{k},lm,1}=O_P(\underline{C}^{-1/2}+r'_n).$$

This completes the proof. $\square$

E Additional Lemma

In this section, we establish a multiway generalization of Lemma 1, which is a minor modification of Lemma D.2 in Davezies et al. (2018). We include its proof for completeness purpose. For any $r=1,...,\ell$,we let $\mathcal{I}_r(\mathbf{C})=\Big\{\mathbf{c}=\boldsymbol{j}\odot\mathbf{e}:\mathbf{e}\in\mathcal{E}_r,\mathbf{1}\le\boldsymbol{j}\le\mathbf{C}\Big\}$ and $\boldsymbol{\varepsilon}_m=\{\mathbf{e}\in\{0;1\}^\ell:\sum_{i=1}^\ell e_i=m\}$, with $\odot$ the Hadamard product on $\mathbb{R}^\ell$.

For each $n\in\mathbb{N}$, let $(N^n_{\boldsymbol{j}},(W^n_{\imath,\boldsymbol{j}})_{1\le\iota\le\overline{N}})_{\boldsymbol{j}\ge\mathbf{1}}$ be a set of random variables. For any $f:\mathrm{supp}(W^n)\to\mathbb{R}^d$ for a fixed $d\in\mathbb{N}$, let us define the multiway empirical process

$$\mathbb{G}_nf:=\sqrt{\underline{C}}\Big\{\frac{1}{\prod_C}\sum_{i=1}^{\ell}\sum_{j_i=1}^{C_i}\sum_{\iota\in N^n_{\mathbf{j}}}f(W^n_{\imath,\boldsymbol{j}})-\mathrm{E}_P[\sum_{\imath\in[N^n_{\mathbf{1}}]}f(W^n_{\imath,\mathbf{1}})]\Big\}.$$

Lemma 3 (Independentization via Hájek Projections)**.** *For each $n\in\mathbb{N}$, suppose that $(N^n_{\mathbf{j}},(W^n_{\imath,\mathbf{j}})_{1\le\iota\le\overline{N}})_{\mathbf{j}\ge\boldsymbol{1}}$ satisfies Assumption 1. Let $\mathcal{F}_n$, $|\mathcal{F}_n|=d$, be a family of functions $f:\mathrm{supp}(W^n)\to\mathbb{R}$ that* *satisfies* $\mathbb{E}\Big[\Big(\sum_{\imath\in[N^n_{\boldsymbol{1}}]} f\Big(W^n_{\imath,\boldsymbol{1}}\Big)\Big)^2\Big] < K < \infty$ *for some* K *independent of* n*. In addition, assume that* $\underline{C} \to \infty$ *and for every* $\boldsymbol{e} \in \boldsymbol{\varepsilon}_1$, $\frac{\underline{C}}{\prod_C} \to \overline{\mu}_i \geq 0$, *where* i *is the nonzero coordinate of* $\boldsymbol{e}$*. Then there exists a family of mutually independent standard uniform r.v.'s* $(U_{\boldsymbol{c}})_{\boldsymbol{c}>0}$ *such that the* $H_n f$*, the Hájek projection of* $G_n f$ *on the set of statistics of the form* $\sum_{\boldsymbol{c}\in\mathcal{I}_r(\boldsymbol{C})} g_{\boldsymbol{c}}(U_{\boldsymbol{c}})$ *(with* $g_{\boldsymbol{c}}(U_{\boldsymbol{c}})$ *square integrable, satisfies*

$$H_n f = \sum_{\boldsymbol{c}\in\mathcal{I}_1(\boldsymbol{C})} \frac{\sqrt{\underline{C}}}{\prod_{i:\boldsymbol{c}_i\neq 0} C_i}\left(\mathbb{E}\left[\sum_{\imath=1}^{N^n_{\boldsymbol{c}\vee\boldsymbol{1}}} f\left(W^n_{\imath,\boldsymbol{c}\vee\boldsymbol{1}}\right)\Big|U_{\boldsymbol{c}}\right] - \mathbb{E}\left[\sum_{\imath\in[N^n_{\boldsymbol{1}}]} f\left(W^n_{\imath,\boldsymbol{1}}\right)\right]\right). \tag{E.1}$$

In addition, it holds uniformly over $\mathcal{F}_n$ *that*

$$V(\mathbb{G}_n f) = V(H_n f) + O(\underline{C}^{-1}) = \sum_{\boldsymbol{e}\in\boldsymbol{\varepsilon}_1} \bar{\mu}_i Cov(\sum_{\imath=1}^{N^n_{\boldsymbol{1}}} f(W^n_{\imath,\boldsymbol{1}}), \sum_{\imath=1}^{N^n_{\boldsymbol{2}-\boldsymbol{e}}} f(W^n_{\imath,\boldsymbol{2}-\boldsymbol{e}})) + O(\underline{C}^{-1}).$$

Proof. Throughout the proof, we drop the superscript n for simplicity. Under Assumption 1(i) and (ii), for each n, one can apply Lemma 7.35 of Kallenberg (2006) and obtain a measurable function τ_n such that

$$(N_{\boldsymbol{j}}, (W_{\imath,\boldsymbol{j}})_{1\leq\imath\leq\overline{N}})_{\boldsymbol{j}\geq\mathbf{1}} = \left(\tau_n(U_{\boldsymbol{j}\odot\mathbf{e}})_{\mathbf{1}\prec\mathbf{e}\preceq\mathbf{1}}\right)_{\boldsymbol{j}\geq\mathbf{1}} \tag{E.2}$$

where $(U_{\mathbf{c}})_{\mathbf{c}\geq\mathbf{0}}$ denote a family of mutually independent uniform random variables on $[0,1]$.

The rest of our proof closely follows that of Lemma D.2 in Davezies et al. (2018) with $r = \underline{r} = 1$. The Hájek projection $H_n f$ is characterized by

$$\mathrm{E}\Big[(\mathbb{G}_n f - H_n f) \times \sum_{\mathbf{c}\in\mathcal{I}_1(\mathbf{C})} g_{\mathbf{c}}\left(U_{\mathbf{c}}\right)\Big] = 0 \text{ for any } \left(g_{\mathbf{c}}\right)_{\mathbf{c}\in\mathcal{I}_1(\mathbf{C})} \in \left(L^{\ell}([0;1])\right)^{|\mathcal{I}_1(\mathbf{C})|}.$$

As a result,we have

$$\mathrm{E}\left[\mathbb{G}_n f|U_{\mathbf{c}}\right] = \mathrm{E}\left[H_n f|U_{\mathbf{c}}\right] \text{ for any } \mathbf{c}\in\mathcal{I}_1(\mathbf{C}).$$

Because the range H_n is closed subspace of square integrable random variables,

$$H_n f = \sum_{\mathbf{c}\in\mathcal{I}_1(\mathbf{C})} \mathbb{E}\left(H_n f|U_{\mathbf{c}}\right).$$

Next

$$H_n f = \sum_{\mathbf{c} \in \mathcal{I}_1(\mathbf{C})} \mathbb{E}\left(\mathbb{G}_n f | U_{\mathbf{c}}\right).$$

Note that for any $\mathbf{c} \in \mathcal{I}_1(\mathbf{C})$, $\mathbf{c} \wedge \mathbf{1}$ is the unique element $\boldsymbol{\varepsilon}_1$ such that $\mathbf{c} = \boldsymbol{j} \odot \mathbf{e}$ for some $\boldsymbol{j}$ (note that $\boldsymbol{j}$ is not unique). Moreover, for any $\mathbf{c} \in \mathcal{I}_1(\mathbf{C})$ independence between the U' s ensures that $\sum_{\imath \in [N_{\boldsymbol{j}}]} f\left(W_{\imath, \boldsymbol{j}}\right) \perp\!\!\!\perp U_{\mathbf{c}}$ if $\boldsymbol{j} \odot \mathbf{e} \neq \mathbf{c}$. This implies

$$\begin{aligned}
\mathbb{E}\left(\mathbb{G}_n f | U_{\mathbf{c}}\right) &= \frac{\sqrt{C}}{\Pi_C} \sum_{\mathbf{1} \leq \boldsymbol{j} \leq \mathbf{C}} \mathbb{E}\left[\sum_{\imath \in [N_{\boldsymbol{j}}]} f\left(W_{\imath, \boldsymbol{j}}\right) - \mathbb{E}\left[\sum_{\imath \in [N_{\mathbf{1}}]} f\left(W_{\imath, \mathbf{1}}\right)\right] \Big| U_{\mathbf{c}}\right] \\
&= \frac{\sqrt{C}}{\Pi_C} \sum_{\mathbf{1} \leq \boldsymbol{j} \leq \mathbf{C}} \mathbb{1}\{\boldsymbol{j} \odot \mathbf{e} = \mathbf{c}\} \mathbb{E}\left[\sum_{\imath \in [N_{\boldsymbol{j}}]} f\left(W_{\imath, \boldsymbol{j}}\right) - \mathbb{E}\left[\sum_{\imath \in [N_{\mathbf{1}}]} f\left(W_{\imath, \mathbf{1}}\right)\right] \Big| U_{\mathbf{c}}\right].
\end{aligned}$$

The representation of $(N_{\boldsymbol{j}}, (W_{\imath, \boldsymbol{j}})_{1 \leq \iota \leq \overline{N}})_{\boldsymbol{j} \geq \mathbf{1}}$ in terms of the U's implies that

$$\mathbb{E}\left[\sum_{\imath=1}^{N_{\boldsymbol{j}}} f\left(W_{\imath, \boldsymbol{j}}\right) - \mathbb{E}\left[\sum_{\imath \in [N_{\mathbf{1}}]} f\left(N_{\mathbf{1}}, W_{\imath, \mathbf{1}}\right)\right] \Big| U_{\mathbf{c}}\right] = \mathbb{E}\left[\sum_{\imath=1}^{N_{\mathbf{c} \vee \mathbf{1}}} f\left(W_{\imath, \mathbf{c} \vee \mathbf{1}}\right) - \mathbb{E}\left[\sum_{\imath \in [N_{\mathbf{1}}]} f\left(W_{\imath, \mathbf{1}}\right)\right] \Big| U_{\mathbf{c}}\right]$$

for any $\boldsymbol{j}$ such that $\boldsymbol{j} \odot \mathbf{e} = \mathbf{c}$. Moreover,

$$\begin{aligned}
\mathbb{E}\left(\mathbb{G}_n f | U_{\mathbf{c}}\right) &= \frac{\sqrt{C}}{\Pi_C} \sum_{\mathbf{1} \leq \boldsymbol{j} \leq \mathbf{C}} \mathbb{1}\{\boldsymbol{j} \odot \mathbf{e} = \mathbf{c}\} \mathbb{E}\left[\sum_{\imath=1}^{N_{\mathbf{c} \vee \mathbf{1}}} f\left(W_{\imath, \mathbf{c} \vee \mathbf{1}}\right) - \mathbb{E}\left[\sum_{\imath \in [N_{\mathbf{1}}]} f\left(W_{\imath, \mathbf{1}}\right)\right] \Big| U_{\mathbf{c}}\right] \\
&= \frac{\sqrt{C} \prod_{i: \mathbf{c}_i = 0} C_i}{\Pi_C} \mathbb{E}\left[\sum_{\imath=1}^{N_{\mathbf{c} \vee \mathbf{1}}} f\left(W_{\imath, \mathbf{c} \vee \mathbf{1}}\right) - \mathbb{E}\left[\sum_{\imath \in [N_{\mathbf{1}}]} f\left(W_{\imath, \mathbf{1}}\right)\right] \Big| U_{\mathbf{c}}\right] \\
&= \frac{\sqrt{C}}{\prod_{i: \mathbf{c}_i \neq 0} C_i} \left(\mathbb{E}\left[\sum_{\imath=1}^{N_{\mathbf{c} \vee \mathbf{1}}} f\left(W_{\imath, \mathbf{c} \vee \mathbf{1}}\right) \Big| U_{\mathbf{c}}\right] - \mathbb{E}\left[\sum_{\imath \in [N_{\mathbf{1}}]} f\left(W_{\imath, \mathbf{1}}\right)\right]\right).
\end{aligned}$$

It follows that

$$H_n f = \sum_{\mathbf{c} \in \mathcal{I}_1(\mathbf{C})} \frac{\sqrt{C}}{\prod_{i: \mathbf{c}_i \neq 0} C_i} \left(\mathbb{E}\left[\sum_{\imath=1}^{N_{\mathbf{c} \vee \mathbf{1}}} f\left(W_{\imath, \mathbf{c} \vee \mathbf{1}}\right) \Big| U_{\mathbf{c}}\right] - \mathbb{E}\left[\sum_{\imath \in [N_{\mathbf{1}}]} f\left(W_{\imath, \mathbf{1}}\right)\right]\right).$$

This shows the first claim of the lemma.

Since $\mathcal{F}_n$ is a finite family, we are left to prove that for each $f \in \mathcal{F}_n$,

$$V(\mathbb{G}_n f) = V(H_n f) + O(\underline{C}^{-1}) = \sum_{i=1}^{\ell} \bar{\mu}_i Cov(\sum_{\imath=1}^{N_{\mathbf{1}}} f(W_{\imath,\mathbf{1}}), \sum_{\imath=1}^{N_{2_i}} f(W_{\imath,2_i})) + O(\underline{C}^{-1}),$$

where 2_i denotes the $\ell-$tuple vector with 2 in each entry but for 1 in the $i-$th entry. Note that

$$\mathbb{V}\left(H_n f\right) = \sum_{e\in\boldsymbol{\varepsilon}_1} \frac{\underline{C}}{\prod_{i:\mathbf{e}_i=1} C_i} \mathbb{V}\left(\mathbb{E}\left[\sum_{\imath\in[N_{\mathbf{1}}]} f\left(W_{\imath,\mathbf{1}}\right) \middle| U_{\mathbf{e}}\right]\right). \tag{E.3}$$

To conclude, it suffices to show that for each $\mathbf{e} \in \boldsymbol{\varepsilon}_1$,

$$\mathbb{V}\left(\mathbb{E}\left[\sum_{\imath\in[N_{\mathbf{1}}]} f\left(W_{\imath,\mathbf{1}}\right) \middle| U_{\mathbf{e}}\right]\right) = Cov\left(\sum_{\imath\in[N_{\mathbf{1}}]} f\left(W_{\imath,\mathbf{1}}\right), \sum_{\imath=1}^{N_{\mathbf{2-e}}} f\left(W_{\imath,\mathbf{2-e}}\right)\right).$$

As $(N_{\boldsymbol{j}}, (W_{\imath,\boldsymbol{j}})_{1\leq\iota\leq\overline{N}})_{\boldsymbol{j}\geq\mathbf{1}} = \left(\tau\left(\left(U_{\boldsymbol{j}\odot\mathbf{e}}\right)_{\mathbf{e}\in\cup_{r=1}^{\ell}\boldsymbol{\varepsilon}_r}\right)\right)_{\boldsymbol{j}\geq\mathbf{1}}$ with i.i.d. U's, we have $\mathbb{E}\left[\sum_{\imath\in[N_{\mathbf{1}}]} f\left(W_{\imath,\mathbf{1}}\right) \middle| U_{\mathbf{e}}\right] = \mathbb{E}\left[\sum_{\imath\in[N_{\boldsymbol{j}}]} f\left(W_{\imath,\boldsymbol{j}}\right) \middle| U_{\mathbf{e}}\right]$ for any $\boldsymbol{j}$ such that $\boldsymbol{j}\odot\mathbf{e} = \mathbf{1}\odot\mathbf{e} = \mathbf{e}$. Because $\mathbf{2}-\mathbf{e}\odot\mathbf{e} = \mathbf{e}$, we have $\mathbb{V}\left(\mathbb{E}\left[\sum_{\imath\in[N_{\mathbf{1}}]} f\left(W_{\imath,\mathbf{1}}\right) \middle| U_{\mathbf{e}}\right]\right) = Cov\left(\mathbb{E}\left[\sum_{\imath\in[N_{\mathbf{1}}]} f\left(W_{\imath,\mathbf{1}}\right) \middle| U_{\mathbf{e}}\right], \mathbb{E}\left[\sum_{\imath=1}^{N_{\mathbf{2-e}}} f\left(W_{\imath,\mathbf{2-e}}\right) \middle| U_{\mathbf{e}}\right]\right)$. For any $\mathbf{e} \in \boldsymbol{\varepsilon}_1$, we have $\mathbf{2}-\mathbf{e} \neq \mathbf{1}$. The independence of the U's ensures

$$\left(U_{\mathbf{1}\odot\mathbf{e}'}\right)_{\mathbf{e}'\in\cup_{r=1}^{\ell}\boldsymbol{\varepsilon}_r\setminus\mathbf{e}} \perp \left(U_{(\mathbf{2-e})\odot\mathbf{e}'}\right)_{\mathbf{e}'\in\cup_{r=1}^{\ell}\boldsymbol{\varepsilon}_r\setminus\mathbf{e}} | U_{\mathbf{e}}$$

and thus $\sum_{\imath=1}^{N_{\mathbf{1}}} f\left(W_{\imath,\mathbf{1}}\right) \perp\!\!\!\perp \sum_{\imath=1}^{N_{\mathbf{2-e}}} f\left(W_{\imath,\mathbf{2-e}}\right) | U_{\mathbf{e}}$. Hence, for $\mathbf{e} \in \boldsymbol{\varepsilon}_1$

$$\mathbb{E}\left[Cov\left(\sum_{\imath\in[N_{\mathbf{1}}]} f_1\left(W_{\imath,\mathbf{1}}\right), \sum_{\imath=1}^{N_{\mathbf{2-e}}} f_2\left(W_{\imath,\mathbf{2-e}}\right) \middle| U_{\mathbf{e}}\right)\right] = 0.$$

By the law of total covariance, we obtain

$$\mathbb{V}\left(\mathbb{E}\left[\sum_{\imath\in[N_{\mathbf{1}}]} f(W_{\imath,\mathbf{1}}) \middle| U_{\mathbf{e}}\right]\right) = Cov\left(\sum_{\imath\in[N_{\mathbf{1}}]} f\left(W_{\imath,\mathbf{1}}\right), \sum_{\imath=1}^{N_{\mathbf{2-e}}} f\left(W_{\imath,\mathbf{2-e}}\right)\right).$$

This establishes the second claim of the lemma. $\square$

F Additional Details on Discussions in Section 3.1.1

In this section, we provide additional details on (3.1) in the discussion in Section 3.1.1. Letting $\mathbb{E}_n = n^{-1}\sum_{i=1}^n$, we have the following concrete expression for the influence function representation (3.1) in the main text.

$$
\begin{aligned}
&\sqrt{n}(\widehat{\theta} - \theta_0) \\
=& \underbrace{E[(D - g_{20}(X))(Z - m_0(X))]^{-1}\sqrt{n}(\mathbb{E}_n - E)[(Y - (D - g_{20}(X))\theta_0 - g_{10}(X))(Z - m_0(X))]}_{A^*} \\
&+ \underbrace{E[(D - g_{20}(X))(Z - m_0(X))]^{-1}\sqrt{n}\mathbb{E}_n[(\widehat{m}(X) - m_0(X))(\widehat{g}_1(X) - g_{10}(X))]}_{B_1^*} \\
&- \underbrace{E[(D - g_{20}(X))(Z - m_0(X))]^{-1}\sqrt{n}\mathbb{E}_n[(\widehat{m}(X) - m_0(X))(\widehat{g}_2(X) - g_{20}(X))]\theta_0}_{B_2^*} \\
&- \underbrace{E[(D - g_{20}(X))(Z - m_0(X))]^{-1}\sqrt{n}\mathbb{E}_n[(\widehat{m}(X) - m_0(X))(Y - (D - g_{20}(X))\theta_0 - g_{10}(X)]}_{C_1^*} \\
&- \underbrace{E[(D - g_{20}(X))(Z - m_0(X))]^{-1}\sqrt{n}\mathbb{E}_n[(\widehat{g}_1(X) - g_{10}(X))(Z - m_0(X))]}_{C_2^*} \\
&+ \underbrace{E[(D - g_{20}(X))(Z - m_0(X))]^{-1}\sqrt{n}\mathbb{E}_n[(\widehat{g}_2(X) - g_{20}(X))(Z - m_0(X))\theta_0]}_{C_3^*} + o_p(1).
\end{aligned}
$$

Term A^* above is the part that is asymptotically normal as mentioned in the main text. Terms B_1^* and $-B_2^*$ above consist B^* in the main text. Terms $-C_1^*$, $-C_2^*$ and C_3^* consist C^* in the main text.

References

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018): "Double/debiased machine learning for treatment and structural parameters," *Econometrics Journal*, 21, C1 – C68.

Chiang, H. and Y. Sasaki (2019): "Lasso under Multi-way Clustering: Estimation and Post-selection Inference," ArXiv:1905.02107.

Davezies, L., X. D'Haultfoeuille, and Y. Guyonvarch (2018): "Asymptotic Results under Multiway Clustering," ArXiv:1807.07925.

Kallenberg, O. (2006): *Probabilistic symmetries and invariance principles*, Springer Science & Business Media.

van der Vaart, A. (1998): *Asymptotic Statistics*, Cambridge University Press.