Electronic Companion to Input-Output Uncertainty Comparisons for Discrete Optimization via Simulation

EC.1. Proof of Proposition 2

In this appendix, we include the proof of Proposition 2 in Section 5.1.

PROPOSITION 2. For all $i \neq \ell$, the optimal objective function value of

$$\begin{aligned} \min \quad & (\beta_i - \beta_\ell)^\top (\widehat{\theta} - \theta^c) \\ \text{subject to} \quad & (\mathcal{B}_i - \widehat{\mathcal{B}}_i)^\top \mathcal{V}_i^{-1} (\mathcal{B}_i - \widehat{\mathcal{B}}_i) = \chi^2_{(k-1)p,\alpha_{11}}, && \text{(EC.1)} \\ & m(\widehat{\theta} - \theta^c)^\top \Sigma^{-1}(\theta^c)(\widehat{\theta} - \theta^c) = \chi^2_{p,\alpha_{12}}, && \text{(EC.2)} \end{aligned}$$

is the same as the optimal objective function value of $\mathcal{P}_{i\ell}$.

Proof. Let $\mathcal{B}_i^\star = \{(\beta_i^\star - \beta_1^\star)^\top, (\beta_i^\star - \beta_2^\star)^\top, \ldots, (\beta_{i-1}^\star - \beta_1^\star)^\top, (\beta_{i+1}^\star - \beta_1^\star)^\top, \ldots, (\beta_i^\star - \beta_k^\star)^\top\}^\top$ and $(\widehat{\theta} - \theta^c)^\star$ be a feasible solution of $\mathcal{P}_{i\ell}$. By means of contradiction, suppose $(\widehat{\theta} - \theta^c)^\star$ is an interior point of the hyper-ellipsoid defined by (15) in the paper, which satisfies (15) with strict inequality. Then there exists scalar $\delta > 0$ such that $(\widehat{\theta} - \theta^c)^\star - \delta(\beta_k^\star - \beta_\ell^\star)$ satisfies (14) at equality and its objective function value is $(\beta_i^\star - \beta_\ell^\star)^\top (\widehat{\theta} - \theta^c)^\star - \delta(\beta_i^\star - \beta_\ell^\star)^\top (\beta_i^\star - \beta_\ell^\star) \leq (\beta_i^\star - \beta_\ell^\star)^\top (\widehat{\theta} - \theta^c)^\star$. Hence, the new feasible solution, $\mathcal{B}_i^\star$ and $(\widehat{\theta} - \theta^c) = (\widehat{\theta} - \theta^c)^\star - \delta(\beta_k^\star - \beta_\ell^\star)$, has an objective function value no greater than that of $\left(\mathcal{B}_i^\star, (\widehat{\theta} - \theta^c)^\star\right)$. We can show that a similar result holds when $\mathcal{B}_i^\star$ satisfies (14) with strict inequality. Therefore, for any interior point of $\mathcal{P}_{i\ell}$, there exists a point on the boundary of the feasible region whose objective function value is no worse than the interior point. Hence, $\mathcal{P}_{i\ell}$ and the problem in Proposition 2 have the same optimal objective function value. $\square$

EC.2. Random Search Algorithm

This appendix presents the random search algorithm briefly introduced in Section 5.1 to approximate the optimal solution to problem $\mathcal{P}_{i\ell}, i \neq \ell$ and establishes its exponential convergence to the optimal solution.

Random Search Algorithm

1. Sample L values of $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^c$ that satisfy $m(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^c)^\top \Sigma^{-1}(\widehat{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^c) = \chi^2_{p,\alpha_{12}}$.
2. For all $1 \le i \le k$ and $i < \ell \le k$:
   (a) For each sampled $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^c$, compute

$$(\widehat{\beta}_i - \widehat{\beta}_\ell)^\top(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^c) - \left(\chi^2_{(k-1)p,\alpha_{11}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^c)^\top \widehat{V}_i(\ell,\ell)\mathbf{S}_{pp}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^c)\right)^{1/2}. \tag{EC.3}$$

   (b) Set $l_{i\ell}$ to the smallest among the L values obtained from Step 2(a), and set $w^{(1)}_{i\ell} = -l_{i\ell}$.
   (c) Set $w^{(1)}_{\ell i} = w^{(1)}_{i\ell}$.

If we assume sphericity of the simulation error variance-covariance matrix, (EC.3) becomes $(\widehat{\beta}_i - \widehat{\beta}_\ell)^\top(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^c) - \widehat{\tau}\left(2\chi^2_{(k-1)p,\alpha_{11}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^c)^\top \mathbf{S}_{pp}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^c)\right)^{1/2}$.

In Step 1, we need to sample from the surface of a p-dimensional hyper-ellipsoid defined by $m(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^c)^\top \Sigma^{-1}(\widehat{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^c) = \chi^2_{p,\alpha_{12}}$. A uniform sample on the surface of the hyper-ellipsoid lets us search the surface evenly in the algorithm. Sun and Farooq (2002) provide an algorithm to sample random points uniformly distributed within a hyper-ellipsoid. They first obtain a uniform sample within a unit hyper-sphere by sampling each spherical coordinate from a uniform distribution. The spherical coordinates of a p-dimensional point consist of one radial coordinate and $p-1$ angular coordinates. Since the radial coordinate determines the distance to the point from the origin of the hyper-sphere, it is sampled from Uniform$(0,1)$. The sampled points within the unit hyper-sphere are then linearly transformed to generate a uniform sample in the target hyper-ellipsoid. Instead of sampling the radial coordinate of each point, we simply set it to 1 so that we obtain a sample of $\{\mathbf{Z} \in \mathbb{R}^p : \mathbf{Z}^\top \mathbf{Z} = 1\}$ that is uniformly distributed on the surface of the p-dimensional unit hypersphere. For Step 1, $\mathbf{Z}$ is transformed to

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^c = \sqrt{\chi^2_{p,\alpha_{12}}/m}\,\Sigma^{1/2}(\widehat{\boldsymbol{\theta}})\,\mathbf{Z} \tag{EC.4}$$

to generate points on the hyperellipsoid, where $\Sigma^{1/2}(\widehat{\boldsymbol{\theta}})$ is a symmetric positive definite matrix such that $\Sigma(\widehat{\boldsymbol{\theta}}) = \Sigma^{1/2}(\widehat{\boldsymbol{\theta}})\Sigma^{1/2}(\widehat{\boldsymbol{\theta}})$. Due to the transformation, the resulting sample is no longer uniformly distributed on the hyperellipsoidal surface in general.

The following theorem assesses the quality of the solution obtained from the Random Search Algorithm; it provides the probability of finding a solution within ϵ-optimality gap given sample size L. For the statement of the theorem, we define function $g(\mathbf{Z}) = (\widehat{\beta}_i - \widehat{\beta}_\ell)^\top(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^c) - \left(\chi^2_{(k-1)p,\alpha_{11}}\chi^2_{p,\alpha_{12}}\mathbf{Z}\,\Sigma^{1/2}(\widehat{\boldsymbol{\theta}})\,\mathcal{V}_i(\ell,\ell)\,\Sigma^{1/2}(\widehat{\boldsymbol{\theta}})\,\mathbf{Z}/m\right)^{1/2}$ by plugging (EC.4) into (16). Notice that $g(\mathbf{Z})$ is continuously differentiable on its domain, $\{\mathbf{Z} : \mathbf{Z}^\top\mathbf{Z} = 1\}$, and therefore is Lipschitz continouous. Note that $0 < I_{\epsilon^2/\gamma^2(1-\epsilon^2/4\gamma^2)}\left(\frac{n-1}{2}, \frac{1}{2}\right) \le 1$ in Theorem EC.1. Therefore, the probability of finding $(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^c)$ that produces ϵ-level optimality gap via the Random Search Algorithm converges to 1 exponentially in L.

Theorem EC.1. *Let $\mathbf{z}^\star$ be a point on the surface of the p-dimensional unit hypersphere such that $\sqrt{\chi^2_{p,\alpha_{12}}/m}\,\Sigma^{1/2}(\widehat{\boldsymbol{\theta}})\,\mathbf{z}^\star$ is an optimal $(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^c)$ of $\mathcal{P}_{i\ell}$ and define regularized incomplete beta function $I_x(a,b)$ as*

$$I_x(a,b)=\frac{\int_0^x t^{a-1}(1-t)^{b-1}dt}{\int_0^1 t^{a-1}(1-t)^{b-1}dt}, 0\le x\le 1.$$

Let κ the Lipschitz constant of $g(\mathbf{Z})$. Given $0<\epsilon<\kappa$ and size-L uniform sample, $\mathbf{Z}_1,\mathbf{Z}_2,\dots,\mathbf{Z}_L$, from the surface of the unit hypersphere

$$\Pr\{\exists 1\le j\le L: |g(\mathbf{Z}_j)-g(\mathbf{z}^\star)|\le\epsilon\}\ge 1-\left(1-\frac{1}{2}I_{\epsilon^2/\kappa^2(1-\epsilon^2/4\kappa^2)}\left(\frac{n-1}{2},\frac{1}{2}\right)\right)^L. \tag{EC.5}$$

Proof. Let $d=\epsilon/\kappa$. Then, by the Lipschitz continuity of $g(\mathbf{Z})$

$$\begin{aligned}\Pr\{\exists 1\le j\le L: |g(\mathbf{Z}_j)-g(\mathbf{z}^\star)|\le\epsilon\} &\ge \Pr\{\exists 1\le j\le L: \kappa||\mathbf{Z}_j-\mathbf{z}^\star||\le\epsilon\}\\ &=\Pr\{\exists 1\le j\le L: ||\mathbf{Z}_j-\mathbf{z}^\star||\le d\}.\end{aligned}$$

Hence, it suffices to show that $\Pr\{\exists 1\le j\le L, ||\mathbf{Z}_j-\mathbf{z}^\star||\le d\}$ equals the right-hand-side of (EC.5).

Define $A^\star=\{\mathbf{Z}: ||\mathbf{Z}-\mathbf{z}^\star||\le d, \mathbf{Z}^\top\mathbf{Z}=1\}$. In words, $A^\star$ is a locus of points whose Euclidean distances to $\mathbf{Z}$ are less than δ on the surface of the unit hypersphere. Sidiropoulos (2014) shows that $A^\star$ is a hypershperical cap of radius $d\sqrt{1-d^2}$ on the unit hypersphere. A hyperspherical cap is defined as the portion of a hypersphere that is cut by a hyperplane. Note that the intersection of the hypersphere and the hyperplane is a $(p-1)$-dimensional hypersphere and we define its radius as the radius of the hyperspherical cap. By definition, a hyperplane may generate two hyperspherical caps; $A^\star$ is the smaller hyperspherical cap that includes $\mathbf{z}^\star$. Sidiropoulos (2014) proves that the angle between $\mathbf{z}^\star$ and all points on the boundary of hyperspherical cap $A^\star$ is $\phi=\cos^{-1}(1-d^2/2)$ for $A^\star$. Since our size L sample, $\mathbf{Z}_1,\mathbf{Z}_2,\dots,\mathbf{Z}_L$, is uniformly distributed on the surface of the unit hypersphere,

$$\Pr\{\exists 1\le j\le L: \mathbf{Z}_j\in A^\star\}=1-\left(1-\frac{(\text{Surface area of }A^\star)}{(\text{Surface area of the unit hypersphere})}\right)^L. \tag{EC.6}$$

Li (2011) shows the ratio of the surface areas of $A^\star$ and the unit hypersphere is

$$\frac{1}{2}I_{\sin^2\phi}\left(\frac{n-1}{2},\frac{1}{2}\right). \tag{EC.7}$$

By the Pythagorean theorem $\sin^2\phi=1-(1-d^2/2)^2=d^2(1-d^2/4)$. Therefore, combining (EC.6) and (EC.7), (EC.5) follows. ☐

EC.3. Multidimensional Quantile Estimation

In this section, we introduce an estimation procedure for a multidimensional quantile, which we use to find $w_{i\ell}^{(2)}, \forall i \neq \ell$, in the all-in procedure and find $w_{i\ell}^{(1)}, w_{i\ell}^{(2)}, \forall i \neq \ell$, in the plug-in procedure. The following description is for computing $w_{i\ell}^{(1)}, \forall i \neq \ell$, for the plug-in procedure, however, it easily generalizes to finding $w_{i\ell}^{(2)}, \forall i \neq \ell$ in both procedures.

For the plug-in procedure, we approximate the joint distribution of the CID effects by plugging in the estimates $\widehat{\beta}_1, \widehat{\beta}_2, \dots, \widehat{\beta}_k$ and $\Sigma(\widehat{\boldsymbol{\theta}})$ into (7). Using this plug-in distribution, for each i we can find $w_{i\ell}^{(1)}, \ell \neq i$ that satisfy (13) given $1-\alpha_1$. For instance, if $i=1$, the plug-in estimator of the joint distribution of $\left\{(\beta_1-\beta_2)^\top(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^c), (\beta_1-\beta_3)^\top(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^c), \dots, (\beta_1-\beta_k)^\top(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^c)\right\}$ is

$$\mathrm{N}\left(\mathbf{0}, \frac{1}{m}\begin{bmatrix} (\widehat{\beta}_1-\widehat{\beta}_2)^\top\Sigma(\widehat{\boldsymbol{\theta}})(\widehat{\beta}_1-\widehat{\beta}_2) & (\widehat{\beta}_1-\widehat{\beta}_2)^\top\Sigma(\widehat{\boldsymbol{\theta}})(\widehat{\beta}_1-\widehat{\beta}_3) & \cdots & (\widehat{\beta}_1-\widehat{\beta}_2)^\top\Sigma(\widehat{\boldsymbol{\theta}})(\widehat{\beta}_1-\widehat{\beta}_k) \\ (\widehat{\beta}_1-\widehat{\beta}_3)^\top\Sigma(\widehat{\boldsymbol{\theta}})(\widehat{\beta}_1-\widehat{\beta}_2) & (\widehat{\beta}_1-\widehat{\beta}_3)^\top\Sigma(\widehat{\boldsymbol{\theta}})(\widehat{\beta}_1-\widehat{\beta}_3) & \cdots & (\widehat{\beta}_1-\widehat{\beta}_3)^\top\Sigma(\widehat{\boldsymbol{\theta}})(\widehat{\beta}_1-\widehat{\beta}_k) \\ \vdots & \vdots & \ddots & \vdots \\ (\widehat{\beta}_1-\widehat{\beta}_k)^\top\Sigma(\widehat{\boldsymbol{\theta}})(\widehat{\beta}_1-\widehat{\beta}_2) & (\widehat{\beta}_1-\widehat{\beta}_k)^\top\Sigma(\widehat{\boldsymbol{\theta}})(\widehat{\beta}_1-\widehat{\beta}_3) & \cdots & (\widehat{\beta}_1-\widehat{\beta}_k)^\top\Sigma(\widehat{\boldsymbol{\theta}})(\widehat{\beta}_1-\widehat{\beta}_k) \end{bmatrix}\right). \tag{EC.8}$$

Therefore, finding $w_{1\ell}^{(1)}, \ell \neq 1$, is the same as finding $(k-1)$-dimensional α_1 quantile of (EC.8). Since this problem has $(k-2)$ degrees of freedom, many choices are possible for the quantile. One approach is to reduce it to a one-dimensional problem to find $w_i^{(1)}$ such that

$$\Pr\{(\widehat{\beta}_i-\widehat{\beta}_\ell)^\top(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^c) \geq -w_i^{(1)}, \forall \ell \neq i\} \geq 1-\alpha_1 \tag{EC.9}$$

for each i. Namely, $w_{i\ell}^{(1)} = w_i^{(1)}$ for all $\ell \neq i$. However, this method tends to make the quantile estimate too conservative as k increases. In particular, if for some system ℓ', $|\widehat{\beta}_i-\widehat{\beta}_{\ell'}|$ is large and $|\widehat{\beta}_i-\widehat{\beta}_\ell|$ for all $\ell \neq i, \ell'$ are small, then $w_i^{(1)}$ in (EC.9) becomes large. This hurts the IOU comparisons, since it becomes more difficult to separate the best system from the rest. Ideally, we would like to rule out an inferior system easily if the CID affects an inferior system and the best system similarly. As an alternative to (EC.9), we propose the following Monte Carlo estimation algorithm to find an α quantile of D-dimensional multivariate distribution.

Multidimensional Quantile Estimation

1. Choose sample size Q for quantile estimation of D-dimensional random variable $\mathbf{Z}$.
2. For each dimension $d=1,2,\dots,D$, set $q^d(\delta)$ to the δ-level quantile of the marginal distribution of Z^d, where Z^d represents the dth coordinate of $\mathbf{Z}$.
3. Sample $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_Q$.
4. Find $\zeta^\star = \inf\{\zeta \geq 1 : \sum_{q=1}^{Q}\prod_{d=1}^{D}\mathbf{1}\left\{Z_q^d \leq \zeta q^d(\delta)\right\} \geq Q\delta\}$.
   Let $\widehat{\mathbf{q}}(\delta) = \zeta^\star\,(q^1(\delta), q^2(\delta), \dots, q^D(\delta))$.

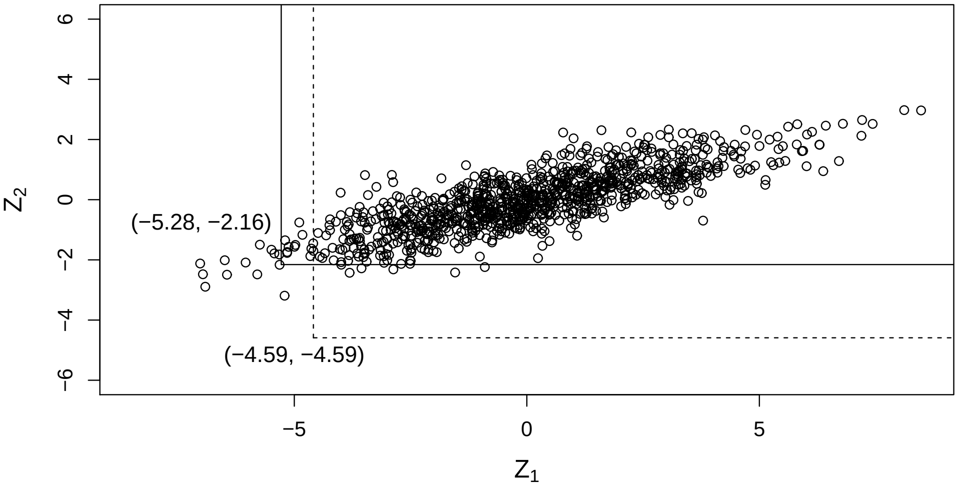

Figure EC.1 An example of the suggested Monte Carlo quantile estimation method for a two-dimensional case for $\alpha_1 = 0.025$.

When we apply this algorithm to obtain the α_1 quantile of (EC.8), it may be infeasible in Step 3 to sample Q random variates from (EC.8) since the variance-covariance matrix can be singular. In such a case, we can sample Q $(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^c)$ variates and convert them to $\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_Q$ using $\widehat{\beta}_1, \widehat{\beta}_2, \ldots, \widehat{\beta}_k$.

Figure EC.1 compares two-dimensional quantiles estimated from the two methods discussed above. A total of $Q = 1{,}000$ points were generated from $\mathrm{N}\left(\mathbf{0}, \begin{bmatrix} 6 & 2 \\ 2 & 1 \end{bmatrix}\right)$, which can be regarded as distribution (EC.8) when $k = 3, Z_1 = (\widehat{\beta}_1 - \widehat{\beta}_2)^\top(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^c)$ and $Z_2 = (\widehat{\beta}_1 - \widehat{\beta}_3)^\top(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^c)$. Notice that $|\widehat{\beta}_1 - \widehat{\beta}_3| < |\widehat{\beta}_1 - \widehat{\beta}_2|$, i.e., system 1 and 3 are affected similarly by the CID. Given $\alpha = 0.025$, $w_1^{(1)}$ satisfying (EC.9) is -4.59 and $(w_{12}^{(1)}, w_{13}^{(1)}) = (-5.05, -2.21)$. If we use $w_1^{(1)} = -4.59$, Z_2 coordinate of all 1,000 points are far above -4.59, which shows that $w_1^{(1)}$ is too conservative for system 3. On the other hand, the proposed quantile estimation algorithm provides a tighter interval width for system 3.

EC.4. Numerical Results for the (s, S) Inventory Problem

Table EC.1 and EC.2 are the detailed results of all-in and plug-in IOU-C and MCB procedures applied to the (s, S) inventory problem in Section 7.

Figure EC.2 and EC.3 show the results of the all-in and plug-in IOU-C as well as the conditional MCB procedure when we assume the simulation error variance-covariance matrix has sphericity and exploit it while all other settings are identical to those of experiments in Section 7. Overall, all three procedures are more conservative under the sphericity assumption; they include more solutions in $\mathcal{S}_0$ and have higher simultaneous coverage probabilities. Detailed results for when $m = 100$ and 400 are presented in Table EC.3 and EC.4, respectively.

Table EC.1 Lower and upper confidence bounds, and the number of times each interval contains 0 for 23 systems in Table 1 when $m = 100, n = 100, B = 159$, and $L = 1{,}000$ from 1,000 runs of all-in IOU-C, plug-in IOU-C, and conditional MCB procedures. Standard errors are in parentheses.

i	$\eta_i(\boldsymbol{\theta}^c) - \max_{\ell \neq i} \eta_\ell(\boldsymbol{\theta}^c)$	All-in IOU-C			Plug-in IOU-C			Conditional MCB		
		LB	UB	Including 0	LB	UB	Including 0	LB	UB	Including 0
1	0.21	−62.78	45.45	1,000	−1.93	8.16	1,000	−0.66	0.93	581
		(0.31)	(0.34)		(0.05)	(0.07)		(0.03)	(0.04)	
2	−0.21	−48.72	22.39	1,000	−8.16	1.93	822	−0.93	0.66	448
		(0.33)	(0.10)		(0.07)	(0.05)		(0.04)	(0.03)	
3	−6.98	−53.09	9.20	993	−14.92	0	0	−7.72	0	0
		(0.33)	(0.12)		(0)	(0)		(0)	(0)	
4	−18.64	−62.64	9.40	982	−26.58	0	0	−19.39	0	0
		(0.24)	(0.13)		(0.04)	(0)		(0.05)	(0)	
5	−34.47	−101.82	9.31	891	−42.34	0	0	−35.21	0	0
		(0.17)	(0.20)		(0.06)	(0)		(0.09)	(0)	
6	−15.76	−73.33	1.42	267	−23.62	0	0	−16.43	0	0
		(0.25)	(0.11)		(0.07)	(0)		(0.04)	(0)	
7	−23.98	−77.77	0.28	80	−31.83	0	0	−24.63	0	0
		(0.25)	(0.04)		(0.06)	(0)		(0.04)	(0)	
8	−36.51	−85.50	0.08	26	−44.38	0	0	−37.19	0	0
		(0.24)	(0.02)		(0.05)	(0)		(0.06)	(0)	
9	−52.46	−116.70	0.04	15	−60.35	0	0	−53.18	0	0
		(0.22)	(0.01)		(0.06)	(0)		(0.09)	(0)	
10	−73.37	−171.50	0.52	119	−81.24	0	0	−74.12	0	0
		(0.26)	(0.06)		(0.10)	(0)		(0.14)	(0)	
11	−34.10	−126.34	9.87	910	−42.01	0	0	−34.83	0	0
		(0.49)	(0.21)		(0.09)	(0)		(0.07)	(0)	
12	−43.66	−136.13	1.70	461	−51.63	0	0	−44.45	0	0
		(0.50)	(0.09)		(0.09)	(0)		(0.08)	(0)	
13	−57.51	−146.38	0.17	47	−65.47	0	0	−58.29	0	0
		(0.49)	(0.04)		(0.09)	(0)		(0.10)	(0)	
14	−74.62	−159.32	0.08	16	−82.60	0	0	−75.42	0	0
		(0.44)	(0.03)		(0.10)	(0)		(0.12)	(0)	
15	−96.18	−188.60	0.00	2	−104.12	0	0	−96.94	0	0
		(0.32)	(0.00)		(0.12)	(0)		(0.15)	(0)	
16	−120.04	−244.50	0.03	10	−127.97	0	0	−120.84	0	0
		(0.34)	(0.01)		(0.16)	(0)		(0.19)	(0)	
17	−58.39	−173.93	3.39	451	−66.26	0	0	−59.06	0	0
		(0.37)	(0.18)		(0.11)	(0)		(0.10)	(0)	
18	−69.84	−175.19	0.06	24	−77.63	0	0	−70.46	0	0
		(0.34)	(0.02)		(0.10)	(0)		(0.10)	(0)	
19	−85.11	−184.46	0	2	−92.90	0	0	−85.72	0	0
		(0.37)	(0)		(0.10)	(0)		(0.11)	(0)	
20	−103.66	−193.62	0	1	−111.45	0	0	−104.29	0	0
		(0.37)	(0)		(0.11)	(0)		(0.12)	(0)	
21	−126.04	−217.91	0	0	−133.96	0	0	−126.79	0	0
		(0.34)	(0)		(0.12)	(0)		(0.15)	(0)	
22	−150.96	−265.31	0	0	−158.80	0	0	−151.64	0	0
		(0.42)	(0)		(0.15)	(0)		(0.18)	(0)	
23	−180.24	−334.13	0	0	−188.09	0	0	−180.99	0	0
		(0.48)	(0)		(0.20)	(0)		(0.23)	(0)	

Table EC.2 Lower and upper confidence bounds, and the number of times each interval contains 0 for 23 systems in Table 1 when $m = 400, n = 100, B = 729$, and $L = 1{,}000$ from 1,000 runs of all-in IOU-C, plug-in IOU-C, and conditional MCB procedures. Standard errors are in parentheses.

i	$\eta_i(\theta^c) - \max_{\ell \neq i} \eta_\ell(\theta^c)$	All-in IOU-C			Plug-in IOU-C			Conditional MCB		
		LB	UB	Including 0	LB	UB	Including 0	LB	UB	Including 0
1	0.21	−23.57 (0.07)	24.02 (0.10)	1,000	−0.95 (0.02)	4.17 (0.03)	1,000	−0.29 (0.01)	0.56 (0.02)	663
2	−0.21	−24.02 (0.10)	15.05 (0.02)	1,000	−4.17 (0.03)	0.95 (0.02)	837	−0.56 (0.02)	0.29 (0.01)	429
3	−6.98	−30.50 (0.09)	1.56 (0.03)	957	−10.92 (0.02)	0 (0)	0	−7.34 (0.02)	0 (0)	0
4	−18.64	−40.42 (0.06)	0.23 (0.01)	317	−22.60 (0.02)	0 (0)	0	−19.01 (0.03)	0 (0)	0
5	−34.47	−68.56 (0.04)	0 (0)	0	−38.39 (0.04)	0 (0)	0	−34.82 (0.05)	0 (0)	0
6	−15.76	−44.67 (0.06)	0 (0)	0	−19.69 (0.03)	0 (0)	0	−16.12 (0.02)	0 (0)	0
7	−23.98	−51.16 (0.06)	0 (0)	0	−27.90 (0.03)	0 (0)	0	−24.31 (0.02)	0 (0)	0
8	−36.51	−60.50 (0.06)	0 (0)	0	−40.45 (0.03)	0 (0)	0	−36.87 (0.03)	0 (0)	0
9	−52.46	−84.79 (0.04)	0 (0)	0	−56.44 (0.04)	0 (0)	0	−52.86 (0.04)	0 (0)	0
10	−73.37	−122.86 (0.06)	0 (0)	0	−77.36 (0.06)	0 (0)	0	−73.77 (0.07)	0 (0)	0
11	−34.10	−80.54 (0.14)	0 (0)	0	−38.00 (0.04)	0 (0)	0	−34.44 (0.04)	0 (0)	0
12	−43.66	−90.70 (0.15)	0 (0)	0	−47.59 (0.04)	0 (0)	0	−44.03 (0.04)	0 (0)	0
13	−57.51	−103.07 (0.14)	0 (0)	0	−61.46 (0.05)	0 (0)	0	−57.89 (0.05)	0 (0)	0
14	−74.62	−117.41 (0.13)	0 (0)	0	−78.61 (0.05)	0 (0)	0	−75.04 (0.06)	0 (0)	0
15	−96.18	−142.13 (0.04)	0 (0)	0	−100.17 (0.07)	0 (0)	0	−96.59 (0.07)	0 (0)	0
16	−120.04	−182.79 (0.05)	0 (0)	0	−124.07 (0.09)	0 (0)	0	−120.49 (0.09)	0 (0)	0
17	−58.39	−116.94 (0.09)	0 (0)	0	−62.28 (0.06)	0 (0)	0	−58.73 (0.05)	0 (0)	0
18	−69.84	−122.57 (0.07)	0 (0)	0	−73.77 (0.05)	0 (0)	0	−70.21 (0.05)	0 (0)	0
19	−85.11	−135.03 (0.09)	0 (0)	0	−88.99 (0.05)	0 (0)	0	−85.43 (0.05)	0 (0)	0
20	−103.66	−148.82 (0.09)	0 (0)	0	−107.56 (0.06)	0 (0)	0	−103.99 (0.06)	0 (0)	0
21	−126.04	−171.26 (0.05)	0 (0)	0	−130.09 (0.07)	0 (0)	0	−126.50 (0.07)	0 (0)	0
22	−150.96	−208.40 (0.05)	0 (0)	0	−154.96 (0.08)	0 (0)	0	−151.37 (0.09)	0 (0)	0
23	−180.24	−257.63 (0.07)	0 (0)	0	−184.29 (0.10)	0 (0)	0	−180.68 (0.11)	0 (0)	0

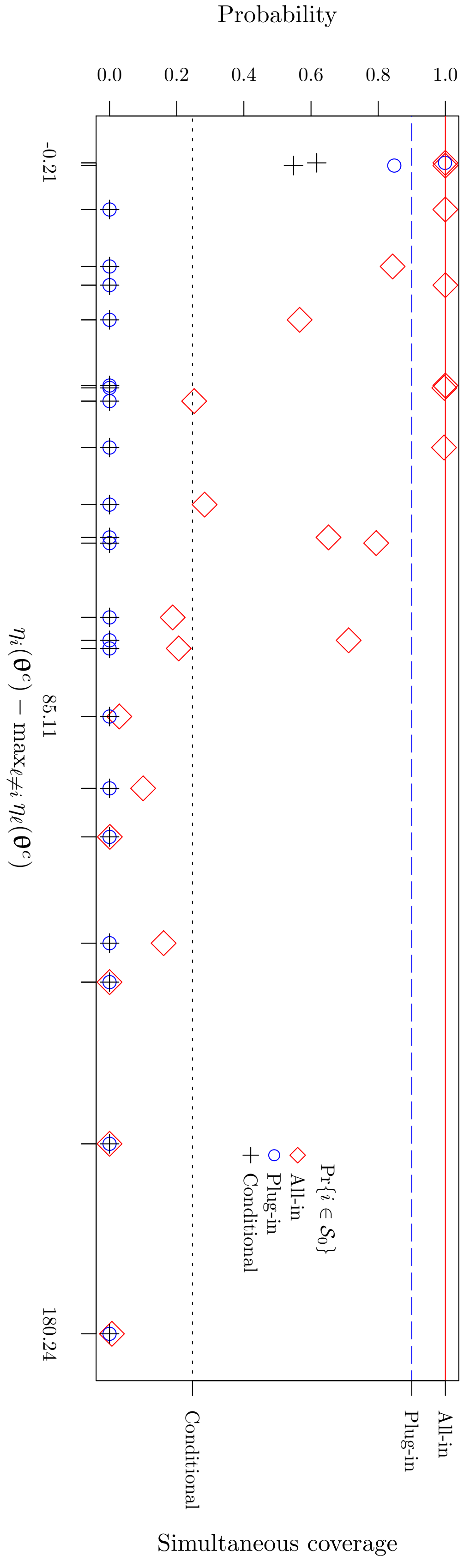

Figure EC.2: Simultaneous coverage probability and $\Pr\{i \in \mathcal{S}_0\}$ for each i when $m = 100$ from 1,000 runs of all-in IOU-C, plug-in IOU-C, and conditional MCB procedures under the sphericity assumption.

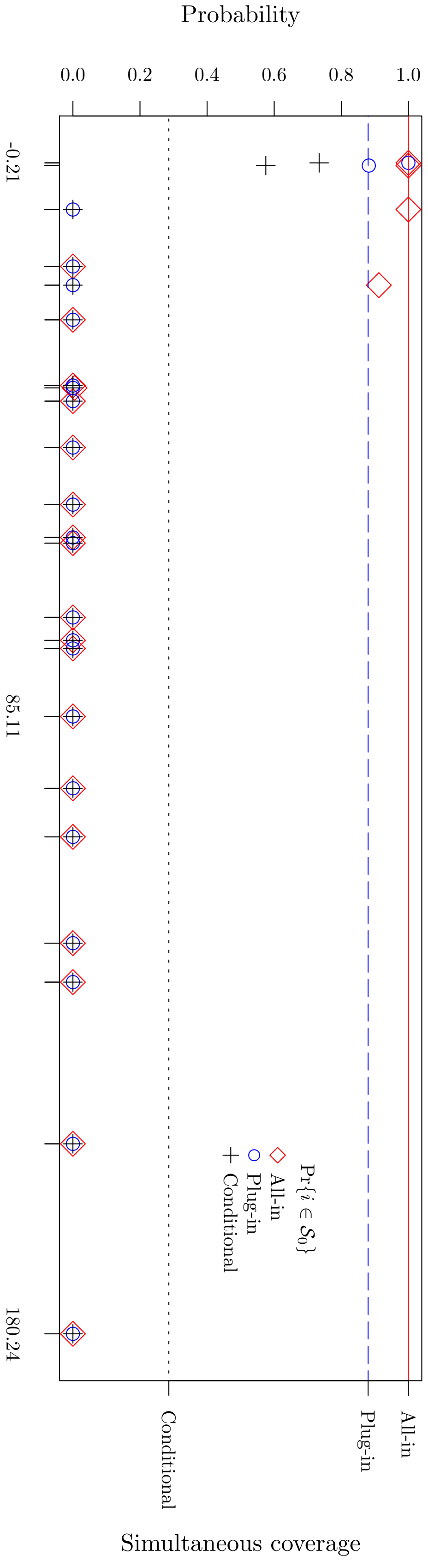

Figure EC.3: Simultaneous coverage probability and $\Pr\{i \in \mathcal{S}_0\}$ for each i when $m = 400$ from 1,000 runs of all-in IOU-C, plug-in IOU-C, and conditional MCB procedures under the sphericity assumption.

Table EC.3 Results under the sphericity assumption. Lower and upper confidence bounds, and the number of times each interval contains 0 for 23 systems in Table 1 when $m = 100, n = 100, B = 159$, and $L = 1{,}000$ from 1,000 runs of all-in IOU-C, plug-in IOU-C, and conditional MCB procedures when sphericity of the simulation error variance-covariance matrix is assumed. Standard errors are in parentheses.

i	$\eta_i(\boldsymbol{\theta}^c) - \max_{\ell \neq i} \eta_\ell(\boldsymbol{\theta}^c)$	All-in IOU-C			Plug-in IOU-C			Conditional MCB		
		MCB LB	MCB UB	Including 0	MCB LB	MCB UB	Including 0	MCB LB	MCB UB	Including 0
1	0.21	−83.84	60.06	1000	−2.22	8.39	999	−0.81	0.95	617
		(0.19)	(0.31)		(0.06)	(0.07)		(0.04)	(0.04)	
2	−0.21	−63.59	32.20	1000	−8.39	2.22	848	−0.95	0.81	548
		(0.27)	(0.06)		(0.07)	(0.06)		(0.04)	(0.04)	
3	−6.98	−66.68	19.04	1000	−15.16	0	0	−7.90	0	0
		(0.27)	(0.09)		(0.05)	(0)		(0.03)	(0)	
4	−18.64	−75.19	20.52	1000	−26.83	0	0	−19.64	0	0
		(0.18)	(0.11)		(0.04)	(0)		(0.05)	(0)	
5	−34.47	−119.06	22.31	997	−42.66	0	0	−35.57	0	0
		(0.12)	(0.22)		(0.06)	(0)		(0.09)	(0)	
6	−15.76	−88.58	5.97	843	−23.86	0	0	−16.57	0	0
		(0.16)	(0.18)		(0.07)	(0)		(0.04)	(0)	
7	−23.98	−92.62	2.63	566	−32.08	0	0	−24.83	0	0
		(0.17)	(0.12)		(0.05)	(0)		(0.04)	(0)	
8	−36.51	−98.57	0.88	252	−44.62	0	0	−37.44	0	0
		(0.15)	(0.07)		(0.04)	(0)		(0.06)	(0)	
9	−52.46	−133.33	0.87	283	−60.60	0	0	−53.49	0	0
		(0.17)	(0.06)		(0.06)	(0)		(0.09)	(0)	
10	−73.37	−193.45	5.03	712	−81.53	0	0	−74.54	0	0
		(0.23)	(0.17)		(0.10)	(0)		(0.13)	(0)	
11	−34.10	−146.82	23.83	1000	−42.24	0	0	−34.92	0	0
		(0.46)	(0.18)		(0.09)	(0)		(0.07)	(0)	
12	−43.66	−156.67	11.82	996	−51.84	0	0	−44.57	0	0
		(0.46)	(0.12)		(0.08)	(0)		(0.08)	(0)	
13	−57.51	−166.55	3.79	652	−65.69	0	0	−58.46	0	0
		(0.44)	(0.14)		(0.08)	(0)		(0.09)	(0)	
14	−74.62	−177.95	0.90	206	−82.80	0	0	−75.64	0	0
		(0.37)	(0.07)		(0.09)	(0)		(0.11)	(0)	
15	−96.18	−208.77	0.43	100	−104.31	0	0	−97.23	0	0
		(0.23)	(0.06)		(0.12)	(0)		(0.14)	(0)	
16	−120.04	−270.45	0.75	161	−128.20	0	0	−121.23	0	0
		(0.30)	(0.07)		(0.15)	(0)		(0.18)	(0)	
17	−58.39	−196.81	9.88	794	−66.46	0	0	−59.10	0	0
		(0.24)	(0.27)		(0.11)	(0)		(0.10)	(0)	
18	−69.84	−197.81	1.06	188	−77.85	0	0	−70.56	0	0
		(0.20)	(0.09)		(0.09)	(0)		(0.09)	(0)	
19	−85.11	−206.25	0.07	29	−93.09	0	0	−85.87	0	0
		(0.22)	(0.02)		(0.10)	(0)		(0.10)	(0)	
20	−103.66	−213.43	0.00	1	−111.65	0	0	−104.48	0	0
		(0.25)	(0.00)		(0.10)	(0)		(0.12)	(0)	
21	−126.04	−237.56	0	0	−134.14	0	0	−127.05	0	0
		(0.29)	(0)		(0.12)	(0)		(0.14)	(0)	
22	−150.96	−289.63	0	0	−158.98	0	0	−151.99	0	0
		(0.37)	(0)		(0.14)	(0)		(0.17)	(0)	
23	−180.24	−364.70	0.02	7	−188.31	0	0	−181.46	0	0
		(0.44)	(0.01)		(0.19)	(0)		(0.22)	(0)	

Table EC.4 Results under the sphericity assumption. Lower and upper confidence bounds, and the number of times each interval contains 0 for 23 systems in Table 1 when $m = 400, n = 100, B = 729$, and $L = 1{,}000$ from 1,000 runs of all-in IOU-C, plug-in IOU-C, and conditional MCB procedures. Standard errors are in parentheses.

i	$\eta_i(\boldsymbol{\theta}^c) - \max_{\ell\neq i}\eta_\ell(\boldsymbol{\theta}^c)$	All-in IOU-C			Plug-in IOU-C			Conditional MCB		
		MCB LB	MCB UB	Including 0	MCB LB	MCB UB	Including 0	MCB LB	MCB UB	Including 0
1	0.21	−27.43	28.01	1000	−1.18	4.56	1000	−0.44	0.72	734
		(0.08)	(0.11)		(0.03)	(0.03)		(0.02)	(0.02)	
2	−0.21	−28.01	16.81	1000	−4.56	1.18	882	−0.72	0.44	575
		(0.11)	(0.02)		(0.03)	(0.03)		(0.02)	(0.02)	
3	−6.98	−34.40	3.37	1000	−11.28	0	0	−7.58	0.00	0
		(0.10)	(0.03)		(0.02)	(0)		(0.02)	(0)	
4	−18.64	−44.12	1.75	912	−22.90	0	0	−19.22	0	0
		(0.07)	(0.04)		(0.02)	(0)		(0.03)	(0)	
5	−34.47	−73.83	0	5	−38.66	0	0	−35.03	0	0
		(0.04)	(0)		(0.04)	(0)		(0.05)	(0)	
6	−15.76	−49.37	0	0	−20.08	0	0	−16.35	0	0
		(0.07)	(0)		(0.03)	(0)		(0.02)	(0)	
7	−23.98	−55.56	0	0	−28.25	0	0	−24.54	0	0
		(0.07)	(0)		(0.03)	(0)		(0.02)	(0)	
8	−36.51	−64.43	0	0	−40.74	0	0	−37.06	0	0
		(0.07)	(0)		(0.03)	(0)		(0.03)	(0)	
9	−52.46	−89.84	0	0	−56.66	0	0	−53.02	0	0
		(0.04)	(0)		(0.04)	(0)		(0.05)	(0)	
10	−73.37	−130.27	0	0	−77.50	0	0	−73.91	0	0
		(0.06)	(0)		(0.06)	(0)		(0.07)	(0)	
11	−34.10	−87.83	0	0	−38.38	0	0	−34.62	0	0
		(0.17)	(0)		(0.04)	(0)		(0.04)	(0)	
12	−43.66	−97.94	0	0	−47.93	0	0	−44.19	0	0
		(0.17)	(0)		(0.05)	(0)		(0.04)	(0)	
13	−57.51	−110.08	0	0	−61.74	0	0	−58.02	0	0
		(0.16)	(0)		(0.05)	(0)		(0.05)	(0)	
14	−74.62	−123.95	0	0	−78.83	0	0	−75.13	0	0
		(0.15)	(0)		(0.05)	(0)		(0.06)	(0)	
15	−96.18	−149.09	0	0	−100.30	0	0	−96.64	0	0
		(0.05)	(0)		(0.07)	(0)		(0.07)	(0)	
16	−120.04	−192.06	0	0	−124.12	0	0	−120.52	0	0
		(0.06)	(0)		(0.08)	(0)		(0.10)	(0)	
17	−58.39	−125.70	0	0	−62.65	0	0	−58.86	0	0
		(0.10)	(0)		(0.06)	(0)		(0.05)	(0)	
18	−69.84	−130.48	0	0	−74.08	0	0	−70.32	0	0
		(0.08)	(0)		(0.05)	(0)		(0.05)	(0)	
19	−85.11	−142.52	0	0	−89.25	0	0	−85.52	0	0
		(0.09)	(0)		(0.05)	(0)		(0.05)	(0)	
20	−103.66	−155.71	0	0	−107.76	0	0	−104.06	0	0
		(0.10)	(0)		(0.06)	(0)		(0.06)	(0)	
21	−126.04	−178.19	0	0	−130.20	0	0	−126.54	0	0
		(0.06)	(0)		(0.07)	(0)		(0.07)	(0)	
22	−150.96	−216.94	0	0	−154.99	0	0	−151.37	0	0
		(0.07)	(0)		(0.08)	(0)		(0.09)	(0)	
23	−180.24	−268.93	0	0	−184.21	0	0	−180.66	0	0
		(0.08)	(0)		(0.10)	(0)		(0.12)	(0)	

EC.5. Proof of Asymptotic Validity of All-in and Plug-in IOU-C

In this appendix, we provide essential lemmas to prove the asymptotic validity of all-in and plug-in IOU-C procedures. To indicate the dependence of $\widehat{\boldsymbol\theta}$ on the real-world sample size m explicitly, we use $\widehat{\boldsymbol\theta}_m$ to represent $\widehat{\boldsymbol\theta}$ given m real-world observations. Similarly, $\widehat{\boldsymbol\theta}_m^{(b)}$ represents $\widehat{\boldsymbol\theta}^{(b)}$ sampled from $\mathrm{N}(\widehat{\boldsymbol\theta}_\mathrm{m},\Sigma(\widehat{\boldsymbol\theta}_\mathrm{m})/\mathrm{m})$. We use $\mathbf{A}^{1/2}$ to represent a square root of matrix $\mathbf{A}$, i.e., $\mathbf{A}=\mathbf{A}^{1/2}\mathbf{A}^{1/2}$. Koeber and Schäfer (2006) show that if $\mathbf{A}$ is positive definite, then there exists a unique $\mathbf{A}^{1/2}$, which is also positive definite. We use this result in the proofs that follow.

We start by showing that estimator $\widehat{\beta}_i$ in Section 5.1 is consistent and asymptotically normally distributed as $m\to\infty$ in Proposition 3. We first prove auxiliary lemmas, Lemma 1–4 to support Proposition 3.

Lemma 1. Define $\rho_{i\ell}(\boldsymbol\theta)=\mathrm{Corr}(\varepsilon_i(\boldsymbol\theta),\varepsilon_\ell(\boldsymbol\theta)|\boldsymbol\theta)$. As $m\to\infty$,

(a) $\mathrm{E}[\sigma_i^2(\widehat{\boldsymbol\theta}_m^{(b)})|\widehat{\boldsymbol\theta}_m]\xrightarrow{a.s.}\sigma_i^2(\boldsymbol\theta^c)$,

(b) $\mathrm{E}[\sigma_i(\widehat{\boldsymbol\theta}_m^{(b)})\sigma_\ell(\widehat{\boldsymbol\theta}_m^{(b)})\rho_{i\ell}(\widehat{\boldsymbol\theta}_m^{(b)})|\widehat{\boldsymbol\theta}_m]\xrightarrow{a.s.}\sigma_i(\boldsymbol\theta^c)\sigma_\ell(\boldsymbol\theta^c)\rho_{i\ell}(\boldsymbol\theta^c)$,

(c) $\mathrm{E}[\sqrt{m}\sigma_i^2(\widehat{\boldsymbol\theta}_m^{(b)})(\widehat{\boldsymbol\theta}_m^{(b)}-\widehat{\boldsymbol\theta}_m)|\widehat{\boldsymbol\theta}_m]\xrightarrow{a.s.}\mathbf{0}$,

(d) $\mathrm{E}[m\sigma_i^2(\widehat{\boldsymbol\theta}_m^{(b)})(\widehat{\boldsymbol\theta}_m^{(b)}-\widehat{\boldsymbol\theta}_m)(\widehat{\boldsymbol\theta}_m^{(b)}-\widehat{\boldsymbol\theta}_m)^\top|\widehat{\boldsymbol\theta}_m]\xrightarrow{a.s.}\sigma_i^2(\boldsymbol\theta^c)\Sigma(\boldsymbol\theta^c)$,

(e) $\mathrm{Corr}(\varepsilon_i(\widehat{\boldsymbol\theta}_m^{(b)})(\widehat{\boldsymbol\theta}_m^{(b)}-\widehat{\boldsymbol\theta}_m),\varepsilon_i(\widehat{\boldsymbol\theta}_m^{(b)})|\widehat{\boldsymbol\theta}_m)\xrightarrow{a.s.}\mathbf{0}$.

Proof. We first show (d) as (a)–(c) can be shown similarly.

Define $\mathcal{E}=\{\omega|\lim_{m\to\infty}\widehat{\boldsymbol\theta}_m(\omega)=\boldsymbol\theta^c,\lim_{m\to\infty}\Sigma(\widehat{\boldsymbol\theta}_m(\omega))=\Sigma(\boldsymbol\theta^c)\}$. From Assumption 1(ii), $\Pr\{\mathcal{E}\}=1$. Let $\omega\in\mathcal{E}$. We would like to show that

$$\lim_{m\to\infty}\mathrm{E}\left[m\sigma_i^2(\widehat{\boldsymbol\theta}_m^{(b)}(\omega))(\widehat{\boldsymbol\theta}_m^{(b)}(\omega)-\widehat{\boldsymbol\theta}_m(\omega))(\widehat{\boldsymbol\theta}_m^{(b)}(\omega)-\widehat{\boldsymbol\theta}_m(\omega))^\top\middle|\widehat{\boldsymbol\theta}_m(\omega)\right]=\sigma_i^2(\boldsymbol\theta^c)\Sigma(\boldsymbol\theta^c). \quad \text{(EC.10)}$$

Note that, given ω, $\widehat{\boldsymbol\theta}_m(\omega)$ is no longer random, however, $\widehat{\boldsymbol\theta}_m^{(b)}(\omega)\sim\mathrm{N}\left(\widehat{\boldsymbol\theta}_m(\omega),\Sigma(\widehat{\boldsymbol\theta}_m(\omega))/m\right)$. Since $\mathrm{E}[m\sigma_i^2(\widehat{\boldsymbol\theta}_m^{(b)})(\widehat{\boldsymbol\theta}_m^{(b)}-\widehat{\boldsymbol\theta}_m)(\widehat{\boldsymbol\theta}_m^{(b)}-\widehat{\boldsymbol\theta}_m)^\top|\widehat{\boldsymbol\theta}_m]$ is bounded for any $\widehat{\boldsymbol\theta}_m\in\Theta$, we can exchange the order of the limit and the integration in (EC.10). Using that $\sqrt{m}\left(\widehat{\boldsymbol\theta}_m^{(b)}(\omega)-\widehat{\boldsymbol\theta}_m(\omega)\right)\sim\mathrm{N}\left(\mathbf{0},\Sigma(\widehat{\boldsymbol\theta}_m(\omega))\right)$ and $\widehat{\boldsymbol\theta}_m^{(b)}(\omega)=\mathbf{x}/\sqrt{m}+\widehat{\boldsymbol\theta}_m(\omega)$, given $\mathbf{x}=\sqrt{m}(\widehat{\boldsymbol\theta}_m^{(b)}(\omega)-\widehat{\boldsymbol\theta}_m(\omega))$,

$$\begin{aligned}&\mathrm{E}\left[\lim_{m\to\infty}\left\{m\sigma_i^2(\widehat{\boldsymbol\theta}_m^{(b)}(\omega))(\widehat{\boldsymbol\theta}_m^{(b)}(\omega)-\widehat{\boldsymbol\theta}_m(\omega))(\widehat{\boldsymbol\theta}_m^{(b)}(\omega)-\widehat{\boldsymbol\theta}_m(\omega))^\top\right\}\middle|\widehat{\boldsymbol\theta}_m(\omega)\right]\\&=\int_{\mathbb{R}^p}\lim_{m\to\infty}\left\{\sigma_i^2(\mathbf{x}/\sqrt{m}+\widehat{\boldsymbol\theta}_m(\omega))\mathbf{x}\mathbf{x}^\top\frac{1}{(\sqrt{2\pi})^p}\left|\Sigma(\widehat{\boldsymbol\theta}_m(\omega))\right|^{-1/2}\exp\left(-\frac{1}{2}\mathbf{x}^\top\Sigma^{-1}(\widehat{\boldsymbol\theta}_m(\omega))\mathbf{x}\right)\right\}d\mathbf{x}.\end{aligned} \quad \text{(EC.11)}$$

Since $\sigma_i^2(\cdot)$ is a continuous function by Assumption 1(iv), $\lim_{m\to\infty}\sigma_i^2(\mathbf{x}/\sqrt{m}+\widehat{\boldsymbol\theta}_m(\omega))=\sigma_i^2(\boldsymbol\theta^c)$ by the Continuous Mapping Theorem. Thus, (EC.11) becomes

$$\int_{\mathbb{R}^p}\left\{\sigma_i^2(\boldsymbol\theta^c)\mathbf{x}\mathbf{x}^\top\frac{1}{(\sqrt{2\pi})^p}\left|\Sigma(\boldsymbol\theta^c)\right|^{-1/2}\exp\left(-\frac{1}{2}\mathbf{x}^\top\Sigma^{-1}(\boldsymbol\theta^c)\mathbf{x}\right)\right\}d\mathbf{x}=\sigma_i^2(\boldsymbol\theta^c)\Sigma(\boldsymbol\theta^c).$$

This holds for arbitrary $\omega \in \mathcal{E}$, therefore,

$$\begin{aligned}&\Pr\left\{\omega \,\middle|\, \lim_{m\to\infty} \mathrm{E}\left[m\sigma_i^2(\widehat{\theta}_m^{(b)}(\omega))(\widehat{\theta}_m^{(b)}(\omega)-\widehat{\theta}_m(\omega))(\widehat{\theta}_m^{(b)}(\omega)-\widehat{\theta}_m(\omega))^\top \,\middle|\, \widehat{\theta}_m(\omega)\right] = \sigma_i^2(\theta^c)\Sigma(\theta^c)\right\}\\ &= \Pr\{\mathcal{E}\} = 1.\end{aligned}$$

Next, we show that (e) follows from (a)–(d). Note that

$$\begin{aligned}&\mathrm{Cov}\left(\sqrt{m}\varepsilon_i(\widehat{\theta}_m^{(b)})(\widehat{\theta}_m^{(b)}-\widehat{\theta}_m), \varepsilon_i(\widehat{\theta}_m^{(b)}) \,\middle|\, \widehat{\theta}_m\right)\\ &= \mathrm{E}\left[\sqrt{m}\varepsilon_i^2(\widehat{\theta}_m^{(b)})(\widehat{\theta}_m^{(b)}-\widehat{\theta}_m) \,\middle|\, \widehat{\theta}_m\right] - \mathrm{E}\left[\sqrt{m}\varepsilon_i(\widehat{\theta}_m^{(b)})(\widehat{\theta}_m^{(b)}-\widehat{\theta}_m) \,\middle|\, \widehat{\theta}_m\right] \underbrace{\mathrm{E}\left[\varepsilon_i(\widehat{\theta}_m^{(b)}) \,\middle|\, \widehat{\theta}_m\right]}_{=0}\\ &= \mathrm{E}\left[\mathrm{E}\left[\sqrt{m}\varepsilon_i^2(\widehat{\theta}_m^{(b)})(\widehat{\theta}_m^{(b)}-\widehat{\theta}_m) \,\middle|\, \widehat{\theta}_m^{(b)}\right] \,\middle|\, \widehat{\theta}_m\right]\\ &= \mathrm{E}\left[\sqrt{m}\sigma_i^2(\widehat{\theta}_m^{(b)})(\widehat{\theta}_m^{(b)}-\widehat{\theta}_m) \,\middle|\, \widehat{\theta}_m\right]\end{aligned}$$

and

$$\begin{aligned}&\mathrm{Corr}\left(\varepsilon_i(\widehat{\theta}_m^{(b)})(\widehat{\theta}_m^{(b)}-\widehat{\theta}_m), \varepsilon_i(\widehat{\theta}_m^{(b)}) \,\middle|\, \widehat{\theta}_m\right)\\ &= \mathrm{Cov}\left(\sqrt{m}\varepsilon_i(\widehat{\theta}_m^{(b)})(\widehat{\theta}_m^{(b)}-\widehat{\theta}_m), \varepsilon_i(\widehat{\theta}_m^{(b)}) \,\middle|\, \widehat{\theta}_m\right)\\ &\quad \times \left(\mathrm{E}\left[\sigma_i^2(\widehat{\theta}_m^{(b)}) \,\middle|\, \widehat{\theta}_m\right] \mathrm{E}\left[m\sigma_i^2(\widehat{\theta}_m^{(b)})(\widehat{\theta}_m^{(b)}-\widehat{\theta}_m)(\widehat{\theta}_m^{(b)}-\widehat{\theta}_m)^\top \,\middle|\, \widehat{\theta}_m\right]\right)^{-1/2}.\end{aligned}$$

Therefore, (a)–(d) guarantee (e). ☐

From Lemma 1, Corollary 1 follows directly.

Corollary 1. Define $\boldsymbol{\varepsilon}_{-i,b} = (\varepsilon_1(\widehat{\theta}_m^{(b)}), \varepsilon_2(\widehat{\theta}_m^{(b)}), \cdots, \varepsilon_{i-1}(\widehat{\theta}_m^{(b)}), \varepsilon_{i+1}(\widehat{\theta}_m^{(b)}), \cdots, \varepsilon_k(\widehat{\theta}_m^{(b)}))^\top - \varepsilon_i(\widehat{\theta}_m^{(b)})\mathbf{1}_{k-1}$. Then, as $m \to \infty$

$$\mathrm{Var}(\boldsymbol{\varepsilon}_{-i,b}|\widehat{\theta}_m) \xrightarrow{a.s.} V_i(\theta^c).$$

Proof. The element of $\mathrm{Var}(\boldsymbol{\varepsilon}_{-i,b}|\widehat{\theta}_m)$ corresponding to systems ℓ and ℓ' is

$$\begin{aligned}&\mathrm{Cov}\left(\varepsilon_\ell(\widehat{\theta}_m^{(b)}) - \varepsilon_i(\widehat{\theta}_m^{(b)}), \varepsilon_{\ell'}(\widehat{\theta}_m^{(b)}) - \varepsilon_i(\widehat{\theta}_m^{(b)}) \,\middle|\, \widehat{\theta}_m\right)\\ &= \mathrm{E}\left[\mathrm{Cov}\left(\varepsilon_\ell(\widehat{\theta}_m^{(b)}) - \varepsilon_i(\widehat{\theta}_m^{(b)}), \varepsilon_{\ell'}(\widehat{\theta}_m^{(b)}) - \varepsilon_i(\widehat{\theta}_m^{(b)}) \,\middle|\, \widehat{\theta}_m^{(b)}\right) \,\middle|\, \widehat{\theta}_m\right]\\ &\qquad + \mathrm{Cov}\left(\mathrm{E}\left[\varepsilon_\ell(\widehat{\theta}_m^{(b)}) \,\middle|\, \widehat{\theta}_m^{(b)}\right], \mathrm{E}\left[\varepsilon_{\ell'}(\widehat{\theta}_m^{(b)}) \,\middle|\, \widehat{\theta}_m^{(b)}\right] \,\middle|\, \widehat{\theta}_m\right)\\ &= \mathrm{E}\left[\sigma_i^2(\widehat{\theta}_m^{(b)}) - \sigma_i(\widehat{\theta}_m^{(b)})\sigma_\ell(\widehat{\theta}_m^{(b)})\rho_{i\ell}(\widehat{\theta}_m^{(b)}) - \sigma_i(\widehat{\theta}_m^{(b)})\sigma_{\ell'}(\widehat{\theta}_m^{(b)})\rho_{i\ell'}(\widehat{\theta}_m^{(b)}) + \sigma_\ell(\widehat{\theta}_m^{(b)})\sigma_{\ell'}(\widehat{\theta}_m^{(b)})\rho_{\ell\ell'}(\widehat{\theta}_m^{(b)}) \,\middle|\, \widehat{\theta}_m\right].\end{aligned}$$

Therefore, the result follows from Lemma 1(a)–(b). ☐

The following Lemmas 2–4 are needed to prove Proposition 3 which states the asymptotic normality of $\widehat{\beta}_i$.

Lemma 2. If $B = m^\gamma, \gamma > 0$, then $(B/m)(\mathbf{C}^\top\mathbf{C})^{-1} \xrightarrow{p} \Sigma^{-1}(\theta^c)$ as $m \to \infty$.

Proof. Define linear operator $\mathcal{L}(\cdot)$ that converts a $p\times p$ matrix to a $p^2\times 1$ vector by rearranging its elements. Our strategy is to show $\mathcal{L}\left((m/B)(\mathbf{C}^\top\mathbf{C})\right)\xrightarrow{p}\mathcal{L}(\Sigma(\boldsymbol{\theta}^c))$, which guarantees $(B/m)(\mathbf{C}^\top\mathbf{C})^{-1}\xrightarrow{p}\Sigma^{-1}(\boldsymbol{\theta}^c)$.

Without loss of generality, consider the first element of $\mathcal{L}\left((m/B)(\mathbf{C}^\top\mathbf{C})\right)$, which is $Z_m^B=(m/B)\sum_{b=1}^B(\widehat{\boldsymbol{\theta}}_{m,1}^{(b)}-\widehat{\boldsymbol{\theta}}_{m,1})^2$, where $\widehat{\boldsymbol{\theta}}_{m,1}^{(b)}$ and $\widehat{\boldsymbol{\theta}}_{m,1}$ are the first elements of $\widehat{\boldsymbol{\theta}}_m^{(b)}$ and $\widehat{\boldsymbol{\theta}}_m$, respectively. We further denote the $(1,1)$ elements of $\Sigma(\widehat{\boldsymbol{\theta}}_m)$ and $\Sigma(\boldsymbol{\theta}^c)$ by Z_m^∞ and Z_∞^∞, respectively. Hence, it suffices to show for any $\epsilon>0$, $\Pr\{|Z_m^B-Z_m^\infty|>\epsilon\}\to0$ and $\Pr\{|Z_m^\infty-Z_\infty^\infty|>\epsilon\}\to0$ as $m\to\infty$.

By Assumption 1(ii), $\Sigma(\widehat{\boldsymbol{\theta}}_m)\xrightarrow{a.s.}\Sigma(\boldsymbol{\theta}^c)$, which guarantees $\Pr\{|Z_m^\infty-Z_\infty^\infty|>\epsilon\}\to0$ as $m\to\infty$. Also, $\Pr\{|Z_m^B-Z_m^\infty|>\epsilon\}=\mathrm{E}[\Pr\{|Z_m^B-Z_m^\infty|>\epsilon|\widehat{\boldsymbol{\theta}}_m\}]$ and $\mathrm{E}[Z_m^B|\widehat{\boldsymbol{\theta}}_m]=Z_m^\infty$. By Chebyshev's inequality,

$$\begin{aligned}\Pr\{|Z_m^B-Z_m^\infty|>\epsilon|\widehat{\boldsymbol{\theta}}_m\}&\le\mathrm{Var}(Z_m^B|\widehat{\boldsymbol{\theta}}_m)/\epsilon^2\\&=m^2/(\epsilon^2B)\mathrm{Var}\left((\widehat{\boldsymbol{\theta}}_{m,1}^{(b)}-\widehat{\boldsymbol{\theta}}_{m,1})^2\Big|\widehat{\boldsymbol{\theta}}_m\right)\\&=2m^2/(\epsilon^2B)(Z_m^\infty/m)^2\\&=2(Z_m^\infty)^2/(\epsilon^2B).\end{aligned}$$

The second equality holds as $(\widehat{\boldsymbol{\theta}}_{m,1}^{(b)}-\widehat{\boldsymbol{\theta}}_{m,1})|\widehat{\boldsymbol{\theta}}_m\sim\mathrm{N}(0,Z_m^\infty/m)$. Since the Fisher information matrix $I(\widehat{\boldsymbol{\theta}}_m)$ is continuous in $\widehat{\boldsymbol{\theta}}_m$ by Assumption 1(ii), $\Sigma_{1,1}(\widehat{\boldsymbol{\theta}}_m)$ is also a continuous function of $\widehat{\boldsymbol{\theta}}_m$. As Θ is compact, there exists $K\in\mathbb{R}$ such that $\sup_{\widehat{\boldsymbol{\theta}}_m\in\Theta}\Sigma_{1,1}(\widehat{\boldsymbol{\theta}}_m)<K$. Substituting $B=m^\gamma$, we have

$$\Pr\{|Z_m^{m^\gamma}-Z_m^\infty|>\epsilon|\widehat{\boldsymbol{\theta}}_m\}<2K^2/(\epsilon^2m^\gamma),$$

which holds for any $\widehat{\boldsymbol{\theta}}_m\in\Theta$. Therefore, $\Pr\{|Z_m^{m^\gamma}-Z_m^\infty|>\epsilon\}<2K^2/(\epsilon^2m^\gamma)\to0$ as $m\to\infty$. □

Lemma 3. If $B=m^\gamma,\gamma>0$, then

$$\frac{1}{\sqrt{m}}\left(\frac{\mathbf{X}^\top\mathbf{X}}{B}\right)^{-1}\left.\begin{bmatrix}\mathrm{E}[\sigma_i^2(\widehat{\boldsymbol{\theta}}_m^{(b)})] & \mathbf{0}^\top\\ \mathbf{0} & \mathrm{E}[\sigma_i^2(\widehat{\boldsymbol{\theta}}_m^{(b)})(\widehat{\boldsymbol{\theta}}_m^{(b)}-\widehat{\boldsymbol{\theta}}_m)(\widehat{\boldsymbol{\theta}}_m^{(b)}-\widehat{\boldsymbol{\theta}}_m)^\top]\end{bmatrix}^{1/2}\right|\widehat{\boldsymbol{\theta}}_m \xrightarrow{p}\begin{bmatrix}0 & \mathbf{0}^\top\\ \mathbf{0} & \sigma_i(\boldsymbol{\theta}^c)\Sigma^{-1/2}(\boldsymbol{\theta}^c)\end{bmatrix}$$

as $m\to\infty$.

Proof. For notational convenience, let

$$W(\widehat{\boldsymbol{\theta}}_m)=\begin{bmatrix}\mathrm{E}[\sigma_i^2(\widehat{\boldsymbol{\theta}}_m^{(b)})|\widehat{\boldsymbol{\theta}}_m] & \mathbf{0}^\top\\ \mathbf{0} & \mathrm{E}[\sigma_i^2(\widehat{\boldsymbol{\theta}}_m^{(b)})(\widehat{\boldsymbol{\theta}}_m^{(b)}-\widehat{\boldsymbol{\theta}}_m)(\widehat{\boldsymbol{\theta}}_m^{(b)}-\widehat{\boldsymbol{\theta}}_m)^\top|\widehat{\boldsymbol{\theta}}_m]\end{bmatrix}^{1/2}$$

conditional on $\widehat{\boldsymbol{\theta}}_m$. For sufficiently large m, $W(\widehat{\boldsymbol{\theta}}_m)$ is well-defined for almost all sample paths as $\sigma_i^2(\widehat{\boldsymbol{\theta}}_m^{(b)}|\widehat{\boldsymbol{\theta}}_m) \xrightarrow{a.s.} \sigma_i^2(\boldsymbol{\theta}^c) > 0$ and $\mathrm{E}[\sigma_i^2(\widehat{\boldsymbol{\theta}}_m^{(b)})(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)^\top|\widehat{\boldsymbol{\theta}}_m] \xrightarrow{a.s.} \sigma_i^2(\boldsymbol{\theta}^c)\Sigma(\boldsymbol{\theta}^c)/m$, where $\sigma_i^2(\boldsymbol{\theta}^c)\Sigma(\boldsymbol{\theta}^c)/m$ is positive definite by Lemma 1. Note that

$$W^{-1}(\widehat{\boldsymbol{\theta}}_m) = \begin{bmatrix} \mathrm{E}[\sigma_i^2(\widehat{\boldsymbol{\theta}}_m^{(b)})|\widehat{\boldsymbol{\theta}}_m]^{-1/2} & \mathbf{0}^\top \\ \mathbf{0} & \mathrm{E}[\sigma_i^2(\widehat{\boldsymbol{\theta}}_m^{(b)})(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)^\top|\widehat{\boldsymbol{\theta}}_m]^{-1/2} \end{bmatrix}$$

is also well-defined for sufficiently large m. Furthermore,

$$\frac{1}{\sqrt{m}}\left(\frac{\mathbf{X}^\top\mathbf{X}}{B}\right)^{-1} W(\widehat{\boldsymbol{\theta}}_m) = \frac{1}{\sqrt{m}}\left(W^{-1}(\widehat{\boldsymbol{\theta}}_m)\frac{\mathbf{X}^\top\mathbf{X}}{B}\right)^{-1}.$$

Since

$$\frac{\mathbf{X}^\top\mathbf{X}}{B} = \begin{bmatrix} 1 & \sum_{b=1}^B(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)^\top/B \\ \sum_{b=1}^B(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)/B & \sum_{b=1}^B(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)^\top/B \end{bmatrix},$$

$$W^{-1}(\widehat{\boldsymbol{\theta}}_m)\frac{\mathbf{X}^\top\mathbf{X}}{B} = \begin{bmatrix} \mathrm{E}[\sigma_i^2(\widehat{\boldsymbol{\theta}}^{(b)})|\widehat{\boldsymbol{\theta}}_m]^{-1/2} & \mathrm{E}[\sigma_i^2(\widehat{\boldsymbol{\theta}}^{(b)})|\widehat{\boldsymbol{\theta}}_m]^{-1/2}\sum_{b=1}^B(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)^\top/B \\ \mathrm{E}[\sigma_i^2(\widehat{\boldsymbol{\theta}}_m^{(b)})(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)^\top|\widehat{\boldsymbol{\theta}}_m]^{-1/2}\sum_{b=1}^B(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)/B & \mathrm{E}[\sigma_i^2(\widehat{\boldsymbol{\theta}}_m^{(b)})(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)^\top|\widehat{\boldsymbol{\theta}}_m]^{-1/2}\sum_{b=1}^B(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)^\top/B \end{bmatrix}.$$

Using

$$\frac{1}{\sqrt{m}}\mathbf{I}_{p+1} = \begin{bmatrix} 1/\sqrt{m} & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{I}_p \end{bmatrix}\begin{bmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{I}_p/\sqrt{m} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{m} & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{I}_p \end{bmatrix}\begin{bmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & \sqrt{m}\mathbf{I}_p \end{bmatrix}^{-1},$$

we can rewrite

$$\frac{1}{\sqrt{m}}\left(W^{-1}(\widehat{\boldsymbol{\theta}}_m)\frac{\mathbf{X}^\top\mathbf{X}}{B}\right)^{-1} = \begin{bmatrix} 1/\sqrt{m} & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{I}_p \end{bmatrix}\left(W^{-1}(\widehat{\boldsymbol{\theta}}_m)\frac{\mathbf{X}^\top\mathbf{X}}{B}\begin{bmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & \sqrt{m}\mathbf{I}_p \end{bmatrix}\right)^{-1}. \tag{EC.12}$$

We first show that as $m \to \infty$

$$W^{-1}(\widehat{\boldsymbol{\theta}}_m)\frac{\mathbf{X}^\top\mathbf{X}}{B}\begin{bmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & \sqrt{m}\mathbf{I}_p \end{bmatrix}\Bigg|\widehat{\boldsymbol{\theta}}_m \xrightarrow{p} \begin{bmatrix} \sigma_i^{-1}(\boldsymbol{\theta}^c) & \mathbf{0}^\top \\ \mathbf{0} & \sigma_i^{-1}(\boldsymbol{\theta}^c)\Sigma^{1/2}(\boldsymbol{\theta}^c) \end{bmatrix}.$$

Multiplying $\begin{bmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & \sqrt{m}\mathbf{I}_p \end{bmatrix}$ to the right-hand-side of $W^{-1}(\widehat{\boldsymbol{\theta}}_m)\mathbf{X}^\top\mathbf{X}/B$ does not change the first column of $W^{-1}(\widehat{\boldsymbol{\theta}}_m)\mathbf{X}^\top\mathbf{X}/B$, but latter p columns are multiplied by $\sqrt{m}$. Now, we consider each block matrix of $W^{-1}(\widehat{\boldsymbol{\theta}}_m)\dfrac{\mathbf{X}^\top\mathbf{X}}{B}\begin{bmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & \sqrt{m}\mathbf{I}_p \end{bmatrix}\Bigg|\widehat{\boldsymbol{\theta}}_m$ separately.

(i) $\mathrm{E}[\sigma_i^2(\widehat{\boldsymbol{\theta}}_m^{(b)})|\widehat{\boldsymbol{\theta}}_m]^{-1/2} \xrightarrow{p} \sigma_i^{-1}(\boldsymbol{\theta}^c)$ follows directly from Lemma 1.(a).

(ii) $\mathrm{E}[\sigma_i^2(\widehat{\boldsymbol{\theta}}_m^{(b)})(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)^\top|\widehat{\boldsymbol{\theta}}_m]^{-1/2}\sum_{b=1}^B(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)/B\Big|\widehat{\boldsymbol{\theta}}_m \xrightarrow{p} \mathbf{0}$.

From Lemma 1.(c),

$$\mathrm{E}[m\sigma_i^2(\widehat{\boldsymbol{\theta}}_m^{(b)})(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)^\top|\widehat{\boldsymbol{\theta}}_m]^{-1/2} \xrightarrow{a.s.} \sigma_i^{-1}(\boldsymbol{\theta}^c)\Sigma^{-1/2}(\boldsymbol{\theta}^c).$$

Therefore, we need to show that $\sqrt{m}\sum_{b=1}^{B}(\widehat{\boldsymbol{\theta}}_m^{(b)}-\widehat{\boldsymbol{\theta}}_m)/B\Big|\widehat{\boldsymbol{\theta}}_m \xrightarrow{p} \mathbf{0}$. Since $\sqrt{m}\sum_{b=1}^{B}(\widehat{\boldsymbol{\theta}}_m^{(b)}-\widehat{\boldsymbol{\theta}}_m)/B\Big|\widehat{\boldsymbol{\theta}}_m \sim \mathrm{N}(\mathbf{0},\Sigma(\widehat{\boldsymbol{\theta}}_m)/B)$, for any constant vector $\mathbf{t}\in\mathbb{R}^p$ and $\epsilon>0$

$$\Pr\left\{\left|\mathbf{t}^\top\sqrt{m}\sum_{b=1}^{B}(\widehat{\boldsymbol{\theta}}_m^{(b)}-\widehat{\boldsymbol{\theta}}_m)/B\right|>\epsilon\,\middle|\,\widehat{\boldsymbol{\theta}}_m\right\}\le\left(\sum_{j=1}^{p}t_j^2\Sigma_{j,j}(\widehat{\boldsymbol{\theta}}_m)+\sum_{j\neq\ell}t_jt_\ell\Sigma_{j,\ell}(\widehat{\boldsymbol{\theta}}_m)\right)\Big/(\epsilon^2B),$$

where $\Sigma_{j,\ell}$ indicates the (j,ℓ)th element of the matrix Σ. Note that $\sum_{j=1}^{p}t_j^2\Sigma_{j,j}(\widehat{\boldsymbol{\theta}}_m)+\sum_{j\neq\ell}t_jt_\ell\Sigma_{j,\ell}(\widehat{\boldsymbol{\theta}}_m)$ is a linear combination of the elements of $\Sigma(\widehat{\boldsymbol{\theta}}_m)$, and therefore is bounded for $\widehat{\boldsymbol{\theta}}_m\in\Theta$ by Assumption 1(i)–(ii). Thus, we conclude $\sqrt{m}\sum_{b=1}^{B}(\widehat{\boldsymbol{\theta}}_m^{(b)}-\widehat{\boldsymbol{\theta}}_m)/B\Big|\widehat{\boldsymbol{\theta}}_m\xrightarrow{p}\mathbf{0}$ by the Cramér-Wold device (Athreya and Lahiri 2006).

(iii) $\sqrt{m}\,\mathrm{E}[\sigma_i^2(\widehat{\boldsymbol{\theta}}^{(b)})|\widehat{\boldsymbol{\theta}}_m]^{-1/2}\sum_{b=1}^{B}(\widehat{\boldsymbol{\theta}}_m^{(b)}-\widehat{\boldsymbol{\theta}}_m)/B\Big|\widehat{\boldsymbol{\theta}}_m\xrightarrow{p}\mathbf{0}$ follows from (i)–(ii).

(iv) $\sqrt{m}\,\mathrm{E}[\sigma_i^2(\widehat{\boldsymbol{\theta}}_m^{(b)})(\widehat{\boldsymbol{\theta}}_m^{(b)}-\widehat{\boldsymbol{\theta}}_m)(\widehat{\boldsymbol{\theta}}_m^{(b)}-\widehat{\boldsymbol{\theta}}_m)^\top|\widehat{\boldsymbol{\theta}}_m]^{-1/2}\sum_{b=1}^{B}(\widehat{\boldsymbol{\theta}}_m^{(b)}-\widehat{\boldsymbol{\theta}}_m)(\widehat{\boldsymbol{\theta}}_m^{(b)}-\widehat{\boldsymbol{\theta}}_m)^\top/B\Big|\widehat{\boldsymbol{\theta}}_m\xrightarrow{p}\sigma_i^{-1}(\boldsymbol{\theta}^c)\Sigma^{1/2}(\boldsymbol{\theta}^c)$.

Similar to (ii), we only need to show that $m\sum_{b=1}^{B}(\widehat{\boldsymbol{\theta}}_m^{(b)}-\widehat{\boldsymbol{\theta}}_m)(\widehat{\boldsymbol{\theta}}_m^{(b)}-\widehat{\boldsymbol{\theta}}_m)^\top/B\Big|\widehat{\boldsymbol{\theta}}_m\xrightarrow{p}\Sigma(\boldsymbol{\theta}^c)$, which is shown in the proof of Lemma 2.

From (i)–(iv), we can conclude $W^{-1}(\widehat{\boldsymbol{\theta}}_m)\dfrac{\mathbf{X}^\top\mathbf{X}}{B}\begin{bmatrix}1&\mathbf{0}^\top\\\mathbf{0}&\sqrt{m}\mathbf{I}_p\end{bmatrix}\Bigg|\widehat{\boldsymbol{\theta}}_m\xrightarrow{p}\begin{bmatrix}\sigma_i^{-1}(\boldsymbol{\theta}^c)&\mathbf{0}^\top\\\mathbf{0}&\sigma_i^{-1}(\boldsymbol{\theta}^c)\Sigma^{1/2}(\boldsymbol{\theta}^c)\end{bmatrix}$. Since the limiting matrix is invertible, and $\begin{bmatrix}1/\sqrt{m}&\mathbf{0}^\top\\\mathbf{0}&\mathbf{I}_p\end{bmatrix}\to\begin{bmatrix}0&\mathbf{0}^\top\\\mathbf{0}&\mathbf{I}_p\end{bmatrix}$ as $m\to\infty$,

$$\begin{bmatrix}1/\sqrt{m}&\mathbf{0}^\top\\\mathbf{0}&\mathbf{I}_p\end{bmatrix}\left(W^{-1}(\widehat{\boldsymbol{\theta}}_m)\frac{\mathbf{X}^\top\mathbf{X}}{B}\begin{bmatrix}1&\mathbf{0}^\top\\\mathbf{0}&\sqrt{m}\mathbf{I}_p\end{bmatrix}\right)^{-1}\Bigg|\widehat{\boldsymbol{\theta}}_m\xrightarrow{p}\begin{bmatrix}0&\mathbf{0}^\top\\\mathbf{0}&\sigma_i(\boldsymbol{\theta}^c)\Sigma^{-1/2}(\boldsymbol{\theta}^c)\end{bmatrix}.\quad\square$$

The following lemma shows a distribution conditional on $\widehat{\boldsymbol{\theta}}$ converges *weakly uniformly* to a normal distribution *in probability* as $m\to\infty$. In general, a distribution $F_m(\cdot)$ is said to converge weakly uniformly to $F^c(\cdot)$ if $\sup_{\mathbf{x}}|F_m(\mathbf{x})-F^c(\mathbf{x})|\to0$. Suppose we replace $F_m(\cdot)$ with conditional distribution $F(\cdot|\widehat{\boldsymbol{\theta}}_m)$, which is a random function of $\widehat{\boldsymbol{\theta}}_m$. If $\sup_{\mathbf{x}}|F(\mathbf{x}|\widehat{\boldsymbol{\theta}}_m)-F^c(\mathbf{x})|\xrightarrow{p}0$, then $F(\cdot|\widehat{\boldsymbol{\theta}}_m)$ is said to converge weakly uniformly to $F^c(\cdot)$ in probability. Weak uniform convergence with probability 1 is also defined similarly.

LEMMA 4. Let $\boldsymbol{\varepsilon}_i=\left\{\varepsilon_i(\widehat{\boldsymbol{\theta}}_m^{(1)}),\varepsilon_i(\widehat{\boldsymbol{\theta}}_m^{(2)}),\ldots,\varepsilon_i(\widehat{\boldsymbol{\theta}}_m^{(B)})\right\}^\top$. If $B=m^\gamma,\gamma>0$, then $\dfrac{\sqrt{B}}{\sqrt{m}}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\boldsymbol{\varepsilon}_i\Bigg|\widehat{\boldsymbol{\theta}}_m$ converges weakly uniformly to $\mathrm{N}\left(\mathbf{0},\begin{bmatrix}0&\mathbf{0}^\top\\\mathbf{0}&\sigma_i^2(\boldsymbol{\theta}^c)\Sigma^{-1}(\boldsymbol{\theta}^c)\end{bmatrix}\right)$ in probability as $m\to\infty$.

Proof. We continue to use $W^{-1}(\widehat{\boldsymbol{\theta}}_m)$ defined in Lemma 3 in this proof. Note that

$$\frac{\sqrt{B}}{\sqrt{m}}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\boldsymbol{\varepsilon}_i=\frac{1}{\sqrt{m}}\left(\frac{\mathbf{X}^\top\mathbf{X}}{B}\right)^{-1}W(\widehat{\boldsymbol{\theta}}_m)W^{-1}(\widehat{\boldsymbol{\theta}}_m)\sqrt{B}\left(\frac{\mathbf{X}^\top\boldsymbol{\varepsilon}_i}{B}\right).$$

Lemma 3 shows $\dfrac{1}{\sqrt{m}}\left(\frac{\mathbf{x}^\top\mathbf{x}}{B}\right)^{-1}W(\widehat{\boldsymbol{\theta}}_m)\Bigg|\widehat{\boldsymbol{\theta}}_m\xrightarrow{p}\begin{bmatrix}0&\mathbf{0}^\top\\\mathbf{0}&\sigma_i(\boldsymbol{\theta}^c)\Sigma^{-1/2}(\boldsymbol{\theta}^c)\end{bmatrix}$. Hence, it remains to show $W^{-1}(\widehat{\boldsymbol{\theta}}_m)\sqrt{B}\left(\mathbf{X}^\top\boldsymbol{\varepsilon}_i/B\right)\Big|\widehat{\boldsymbol{\theta}}_m\xrightarrow{D}\mathrm{N}\left(\mathbf{0},\mathbf{I}_{p+1}\right)$.

First, note that

$$W^{-1}(\widehat{\theta}_m)\sqrt{B}\left(\mathbf{X}^\top\boldsymbol{\varepsilon}_i/B\right)=\sqrt{B}\sum_{b=1}^{B}W^{-1}(\widehat{\theta}_m)\begin{bmatrix}1\\ \widehat{\theta}_m^{(b)}-\widehat{\theta}_m\end{bmatrix}\varepsilon_i(\widehat{\theta}_m^{(b)})/B$$

and

$$W^{-1}(\widehat{\theta}_m)\begin{bmatrix}1\\ \widehat{\theta}_m^{(b)}-\widehat{\theta}_m\end{bmatrix}\varepsilon_i(\widehat{\theta}_m^{(b)})=\begin{bmatrix}\varepsilon_i(\widehat{\theta}_m^{(b)})/\mathrm{Var}\left(\varepsilon_i(\widehat{\theta}_m^{(b)})\Big|\widehat{\theta}_m\right)^{1/2}\\ \mathrm{Cov}\left(\varepsilon_i(\widehat{\theta}_m^{(b)})(\widehat{\theta}^{(b)}-\widehat{\theta}_m)\Big|\widehat{\theta}_m\right)^{-1/2}\varepsilon_i(\widehat{\theta}_m^{(b)})(\widehat{\theta}^{(b)}-\widehat{\theta}_m)\end{bmatrix}.$$

Therefore,

$$\mathrm{Var}\left(W^{-1}(\widehat{\theta}_m)\begin{bmatrix}1\\ \widehat{\theta}_m^{(b)}-\widehat{\theta}_m\end{bmatrix}\varepsilon_i(\widehat{\theta}_m^{(b)})\Bigg|\widehat{\theta}_m\right)=$$
$$\begin{bmatrix}1 & \mathrm{Corr}\left(\varepsilon_i(\widehat{\theta}_m^{(b)}),(\widehat{\theta}_m^{(b)}-\widehat{\theta}_m)\varepsilon_i(\widehat{\theta}_m^{(b)})\Big|\widehat{\theta}_m\right)^\top\\ \mathrm{Corr}\left(\varepsilon_i(\widehat{\theta}_m^{(b)}),(\widehat{\theta}_m^{(b)}-\widehat{\theta}_m)\varepsilon_i(\widehat{\theta}_m^{(b)})\Big|\widehat{\theta}_m\right) & \mathbf{I}_p\end{bmatrix}.$$

From Lemma 1(d), for sufficiently large m, $\left|\left|\mathrm{Corr}\left((\widehat{\theta}_m^{(b)}-\widehat{\theta}_m)\varepsilon_i(\widehat{\theta}_m^{(b)}),\varepsilon_i(\widehat{\theta}_m^{(b)})\Big|\widehat{\theta}_m\right)\right|\right|<1$ for almost all sample paths. This guarantees $\mathrm{Var}\left(W^{-1}(\widehat{\theta}_m)\begin{bmatrix}1\\ \widehat{\theta}_m^{(b)}-\widehat{\theta}_m\end{bmatrix}\varepsilon_i(\widehat{\theta}_m^{(b)})\Bigg|\widehat{\theta}_m\right)$ to be invertible for sufficiently large m. Let $U(\widehat{\theta}_m)=\mathrm{Var}\left(W^{-1}(\widehat{\theta}_m)\begin{bmatrix}1\\ \widehat{\theta}_m^{(b)}-\widehat{\theta}_m\end{bmatrix}\varepsilon_i(\widehat{\theta}_m^{(b)})\Bigg|\widehat{\theta}_m\right)^{1/2}$, then $U(\widehat{\theta}_m)$ is also invertible for almost all sample paths for sufficiently large m. Thus, we can rewrite

$$W^{-1}(\widehat{\theta}_m)\sqrt{B}\left(\mathbf{X}^\top\boldsymbol{\varepsilon}_i/B\right)=U(\widehat{\theta}_m)\sqrt{B}U^{-1}(\widehat{\theta}_m)W^{-1}(\widehat{\theta}_m)\left(\mathbf{X}^\top\boldsymbol{\varepsilon}_i/B\right).$$

Clearly, $U(\widehat{\theta}_m)\xrightarrow{a.s.}\mathbf{I}_{p+1}$ by Lemma 1(d). Hence, the remaining piece is

$$\sqrt{B}U^{-1}(\widehat{\theta}_m)W^{-1}(\widehat{\theta}_m)\left(\mathbf{X}^\top\boldsymbol{\varepsilon}_i/B\right)\Big|\widehat{\theta}_m\xrightarrow{D}\mathrm{N}(0,\mathbf{I}_{p+1})$$

in probability, which we show in the following using the Berry-Esseen theorem (Athreya and Lahiri 2006).

Let $F_{m,B}(\cdot|\widehat{\theta}_m)$ denote the conditional cumulative distribution function (cdf) of $\sqrt{B}U^{-1}(\widehat{\theta}_m)W^{-1}(\widehat{\theta}_m)\left(\mathbf{X}^\top\boldsymbol{\varepsilon}_i/B\right)\Big|\widehat{\theta}_m$ and $\mathbf{t}\in\mathbb{R}^{p+1}$ be an arbitrary constant vector. We further use $F^{\mathbf{t}}_{m,B}(\cdot|\widehat{\theta}_m)$ to denote the cdf of $\mathbf{t}^\top\mathbf{Z}_{m,B}$, where $\mathbf{Z}_{m,B}\sim F_{m,B}(\cdot|\widehat{\theta}_m)$. Similarly, let $F^{\mathbf{t}}_\infty(\cdot)$ denote the cdf of $\mathbf{t}^\top\mathbf{Z}_\infty$, where $\mathbf{Z}_\infty\sim\mathrm{N}(0,\mathbf{I}_{p+1})$. Note that

$$\begin{aligned}&\mathrm{E}\left[\left|\left|\mathbf{t}^\top U^{-1}(\widehat{\theta}_m)W^{-1}(\widehat{\theta}_m)\begin{bmatrix}1\\ \widehat{\theta}_m^{(b)}-\widehat{\theta}_m\end{bmatrix}\varepsilon_i(\widehat{\theta}_m^{(b)})\right|\right|^3\Bigg|\widehat{\theta}_m\right]\\ &\le\mathrm{E}\left[||\mathbf{t}||^3||U^{-1}(\widehat{\theta}_m)||^3\left|\left|W^{-1}(\widehat{\theta}_m)\begin{bmatrix}1\\ \widehat{\theta}_m^{(b)}-\widehat{\theta}_m\end{bmatrix}\varepsilon_i(\widehat{\theta}_m^{(b)})\right|\right|^3\Bigg|\widehat{\theta}_m\right]\end{aligned}$$

$$
\begin{aligned}
&= ||\mathbf{t}||^3 ||U^{-1}(\widehat{\boldsymbol{\theta}}_m)||^3 \mathrm{E}\left[\left|\left| W^{-1}(\widehat{\boldsymbol{\theta}}_m) \begin{bmatrix} 1 \\ \widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m \end{bmatrix} \varepsilon_i(\widehat{\boldsymbol{\theta}}_m^{(b)}) \right|\right|^3 \middle| \widehat{\boldsymbol{\theta}}_m \right] \\
&= ||\mathbf{t}||^3 ||U^{-1}(\widehat{\boldsymbol{\theta}}_m)||^3 \mathrm{E}\left[\left|\left| \begin{bmatrix} \varepsilon_i(\widehat{\boldsymbol{\theta}}_m^{(b)}) / \mathrm{Var}\left(\varepsilon_i(\widehat{\boldsymbol{\theta}}_m^{(b)})\right)^{1/2} \\ \mathrm{Var}\left(\varepsilon_i(\widehat{\boldsymbol{\theta}}_m^{(b)})(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)\right)^{-1/2} \varepsilon_i(\widehat{\boldsymbol{\theta}}_m^{(b)})(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m) \end{bmatrix} \right|\right|^3 \middle| \widehat{\boldsymbol{\theta}}_m \right] \\
&= ||\mathbf{t}||^3 ||U^{-1}(\widehat{\boldsymbol{\theta}}_m)||^3 \\
&\quad \times \mathrm{E}\left[\left\{ \frac{\varepsilon_i^2(\widehat{\boldsymbol{\theta}}_m^{(b)})}{\mathrm{Var}\left(\varepsilon_i(\widehat{\boldsymbol{\theta}}_m^{(b)})\right)} + \varepsilon_i^2(\widehat{\boldsymbol{\theta}}_m^{(b)})(\widehat{\boldsymbol{\theta}}^{(b)} - \widehat{\boldsymbol{\theta}}_m)^\top \mathrm{Var}\left(\varepsilon_i(\widehat{\boldsymbol{\theta}}_m^{(b)})(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)\right)^{-1} (\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m) \right\}^{3/2} \middle| \widehat{\boldsymbol{\theta}}_m \right] \\
&= ||\mathbf{t}||^3 ||U^{-1}(\widehat{\boldsymbol{\theta}}_m)||^3 \\
&\quad \times \mathrm{E}\left[\varepsilon_i^3(\widehat{\boldsymbol{\theta}}_m^{(b)}) \left\{ \mathrm{Var}\left(\varepsilon_i(\widehat{\boldsymbol{\theta}}_m^{(b)})\right)^{-1} + (\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)^\top \mathrm{Var}\left(\varepsilon_i(\widehat{\boldsymbol{\theta}}_m^{(b)})(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)\right)^{-1} (\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m) \right\}^{3/2} \middle| \widehat{\boldsymbol{\theta}}_m \right] \\
&< ||\mathbf{t}||^3 ||U^{-1}(\widehat{\boldsymbol{\theta}}_m)||^3 u^\star,
\end{aligned}
$$

where the last inequality holds for constant $u^\star$ from Assumption 1(v)(d). Also, $||U^{-1}(\widehat{\boldsymbol{\theta}}_m)|| \xrightarrow{a.s.} 1$ because $U(\widehat{\boldsymbol{\theta}}_m) \xrightarrow{a.s.} \mathbf{I}_{p+1}$, which implies that for sufficiently large m, $||U^{-1}(\widehat{\boldsymbol{\theta}}_m)|| < 2$ for almost all sample paths. Therefore, by the Berry-Esseen theorem,

$$
\sup_{x \in \mathbb{R}} \left| F_{m,B}^{\mathbf{t}}(x|\widehat{\boldsymbol{\theta}}_m) - F_\infty^{\mathbf{t}}(x|\widehat{\boldsymbol{\theta}}_m) \right| < C ||\mathbf{t}||^3 2u^\star / \sqrt{B}
$$

for some positive constant C. By requiring $B = m^\gamma$, we have

$$
\sup_{x \in \mathbb{R}} \left| F_{m,B}^{\mathbf{t}}(x|\widehat{\boldsymbol{\theta}}_m) - F_\infty^{\mathbf{t}}(x|\widehat{\boldsymbol{\theta}}_m) \right| < C ||\mathbf{t}||^3 2u^\star m^{-\gamma/2}.
$$

Therefore, $\sup_{x \in \mathbb{R}} \left| F_{m,B}^{\mathbf{t}}(x|\widehat{\boldsymbol{\theta}}_m) - F_\infty^{\mathbf{t}}(x|\widehat{\boldsymbol{\theta}}_m) \right| \xrightarrow{a.s.} 0$. Since this result holds for any $\mathbf{t} \in \mathbb{R}^{p+1}$, by the Cramér-Wold device, we conclude $F_{m,B}(\cdot|\widehat{\boldsymbol{\theta}}_m)$ converges weakly uniformly to $N(\mathbf{0}, \mathbf{I}_{p+1})$, almost surely. Finally, by Slutsky's theorem (Athreya and Lahiri 2006),

$$
\begin{aligned}
&\frac{\sqrt{B}}{\sqrt{m}} \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \boldsymbol{\varepsilon}_i \bigg| \widehat{\boldsymbol{\theta}}_m \\
&= \underbrace{\frac{1}{\sqrt{m}} \left(\frac{\mathbf{X}^\top \mathbf{X}}{B}\right)^{-1} W(\widehat{\boldsymbol{\theta}}_m)}_{\xrightarrow{p} \begin{bmatrix} 0 & \mathbf{0}^\top \\ \mathbf{0} & \sigma_i(\boldsymbol{\theta}^c)\Sigma^{-1/2}(\boldsymbol{\theta}^c) \end{bmatrix}} \underbrace{U(\widehat{\boldsymbol{\theta}}_m)}_{\xrightarrow{a.s.} \mathbf{I}_{p+1}} \underbrace{\sqrt{B} U^{-1}(\widehat{\boldsymbol{\theta}}_m) W^{-1}(\widehat{\boldsymbol{\theta}}_m) \left(\mathbf{X}^\top \boldsymbol{\varepsilon}_i / B\right)}_{\xrightarrow{\text{weakly uniformly}} N(\mathbf{0}, \mathbf{I}_{p+1}) \text{ a.s.}} \bigg| \widehat{\boldsymbol{\theta}}_m \\
&\xrightarrow{\text{weakly uniformly}} \mathrm{N}\left(\mathbf{0}, \begin{bmatrix} 0 & \mathbf{0}^\top \\ \mathbf{0} & \sigma_i^2(\boldsymbol{\theta}^c)\Sigma^{-1}(\boldsymbol{\theta}^c) \end{bmatrix}\right)
\end{aligned}
$$

in probability. $\square$

PROPOSITION 3. Let $\boldsymbol{\beta}_i = (\eta_i(\boldsymbol{\theta}^c), \nabla\eta_i(\boldsymbol{\theta}^c)^\top)^\top$. For each i, if $B = m^\gamma, 0 < \gamma < 2$, then $\sqrt{\frac{B}{m}}((\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}_i - \boldsymbol{\beta}_i)\Big|\widehat{\boldsymbol{\theta}}_m$ converges weakly uniformly to $\mathrm{N}\left(\mathbf{0}, \begin{bmatrix}0 & \mathbf{0}^\top \\ \mathbf{0} & \sigma_i^2(\boldsymbol{\theta}^c)\Sigma^{-1}(\boldsymbol{\theta}^c))\end{bmatrix}\right)$ in probability as $m \to \infty$.

Proof. From Assumption 1(iii), for almost all sample paths $\eta_i(\cdot)$ is twice differentiable at $\widehat{\boldsymbol{\theta}}_m$ for sufficiently large m. Define $\tilde{\boldsymbol{\beta}}_i = (\eta_i(\widehat{\boldsymbol{\theta}}_m), \nabla\eta_i(\widehat{\boldsymbol{\theta}}_m)^\top)^\top$. Then, for some $\ddot{\boldsymbol{\theta}}^{(b)} \in [\widehat{\boldsymbol{\theta}}_m^{(b)}, \widehat{\boldsymbol{\theta}}_m]$

$$Y_i(\widehat{\boldsymbol{\theta}}_m^{(b)}) = \left(1, (\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)^\top\right)\tilde{\boldsymbol{\beta}}_i + \frac{1}{2}(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)^\top\nabla^2\eta_i(\ddot{\boldsymbol{\theta}}^{(b)})(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m) + \varepsilon_i(\widehat{\boldsymbol{\theta}}_m^{(b)}).$$

Let $\mathbf{h}_i$ be a $B \times 1$ vector whose bth element is $(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)^\top\nabla^2\eta_i(\ddot{\boldsymbol{\theta}}^{(b)})(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)$. Then,

$$\begin{aligned}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}_i - \boldsymbol{\beta}_i &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top(\mathbf{X}\tilde{\boldsymbol{\beta}}_i + \mathbf{h}_i + \boldsymbol{\varepsilon}_i) - \boldsymbol{\beta}_i \\ &= \tilde{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i + (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{h}_i + (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\varepsilon}_i.\end{aligned}$$

Hence, we need to show the following:

(i) $\sqrt{B}/\sqrt{m}(\tilde{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i) \xrightarrow{p} \mathbf{0}$.

(ii) $\sqrt{B}/\sqrt{m}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\varepsilon}_i\Big|\widehat{\boldsymbol{\theta}}_m$ converges weakly uniformly to $\mathrm{N}\left(\mathbf{0}, \begin{bmatrix}0 & \mathbf{0}^\top \\ \mathbf{0} & \sigma_i^2(\boldsymbol{\theta}^c)\Sigma^{-1}(\boldsymbol{\theta}^c)\end{bmatrix}\right)$ in probability.

(iii) $\sqrt{B}/\sqrt{m}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{h}_i\Big|\widehat{\boldsymbol{\theta}}_m \xrightarrow{p} \mathbf{0}$.

Lemma 4 shows (ii). To show (i), note that

$$\sqrt{B}/\sqrt{m}(\tilde{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i) = m^{(\gamma-1)/2}\begin{bmatrix}\eta_i(\widehat{\boldsymbol{\theta}}_m) - \eta_i(\boldsymbol{\theta}^c) \\ \nabla\eta_i(\widehat{\boldsymbol{\theta}}_m) - \nabla\eta_i(\boldsymbol{\theta}^c)\end{bmatrix}.$$

For sufficiently large m,

$$m^{(\gamma-1)/2}\left(\eta_i(\widehat{\boldsymbol{\theta}}_m) - \eta_i(\boldsymbol{\theta}^c)\right) = m^{(\gamma-1)/2}\nabla\eta_i\left(\bar{\boldsymbol{\theta}}\right)^\top(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}^c)$$

for some $\bar{\boldsymbol{\theta}} \in [\widehat{\boldsymbol{\theta}}_m, \boldsymbol{\theta}^c]$. Since $(\gamma-1)/2 < 1/2, m^{(\gamma-1)/2}(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}^c) \xrightarrow{p} \mathbf{0}$ and $m^{(\gamma-1)/2}\nabla\eta_i\left(\bar{\boldsymbol{\theta}}\right)^\top(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}^c) \xrightarrow{p} 0$. Similarly,

$$m^{(\gamma-1)/2}\left(\nabla\eta_i(\widehat{\boldsymbol{\theta}}_m) - \nabla\eta_i(\boldsymbol{\theta}^c)\right) = m^{(\gamma-1)/2}\nabla^2\eta_i(\bar{\bar{\boldsymbol{\theta}}})^\top(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}^c)$$

for some $\bar{\bar{\boldsymbol{\theta}}} \in [\widehat{\boldsymbol{\theta}}_m, \boldsymbol{\theta}^c]$ as $\eta_i(\cdot)$ is twice continuously differentiable at $\boldsymbol{\theta}^c$. Therefore, $m^{(\gamma-1)/2}\nabla^2\eta_i\left(\bar{\bar{\boldsymbol{\theta}}}\right)^\top(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}^c) \xrightarrow{p} \mathbf{0}$.

Now, we show (iii) by first rewriting

$$\sqrt{B}/\sqrt{m}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{h}_i = 1/\sqrt{m}(\mathbf{X}^\top\mathbf{X}/B)^{-1}W(\widehat{\boldsymbol{\theta}}_m)W^{-1}(\widehat{\boldsymbol{\theta}}_m)\sqrt{B}\mathbf{X}^\top\mathbf{h}_i/B,$$

where $W(\widehat{\boldsymbol{\theta}}_m)$ is defined in the proof of Lemma 3. Lemma 4 shows $1/\sqrt{m}(\mathbf{X}^\top\mathbf{X}/B)^{-1}W(\widehat{\boldsymbol{\theta}}_m)\Big|\widehat{\boldsymbol{\theta}}_m \xrightarrow{p} \begin{bmatrix}0 & \mathbf{0}^\top \\ \mathbf{0} & \sigma_i(\boldsymbol{\theta}^c)\Sigma^{-1/2}(\boldsymbol{\theta}^c)\end{bmatrix}$. So, we focus on showing $W^{-1}(\widehat{\boldsymbol{\theta}}_m)\sqrt{B}\mathbf{X}^\top\mathbf{h}_i/B \xrightarrow{p} \mathbf{0}$. Note that

$$W^{-1}(\widehat{\boldsymbol{\theta}}_m)\sqrt{B}\mathbf{X}^\top\mathbf{h}_i/B = \frac{\sqrt{B}}{B}\sum_{b=1}^{B}W^{-1}(\widehat{\boldsymbol{\theta}}_m)\begin{bmatrix}1 \\ \widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m\end{bmatrix}(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)^\top\nabla^2\eta_i(\ddot{\boldsymbol{\theta}}^{(b)})(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)/2,$$

where

$$\begin{aligned}
&\sqrt{B}W^{-1}(\widehat{\boldsymbol{\theta}}_m)\begin{bmatrix}1\\ \widehat{\boldsymbol{\theta}}_m^{(b)}-\widehat{\boldsymbol{\theta}}_m\end{bmatrix}(\widehat{\boldsymbol{\theta}}_m^{(b)}-\widehat{\boldsymbol{\theta}}_m)^\top\nabla^2\eta_i(\ddot{\boldsymbol{\theta}}^{(b)})(\widehat{\boldsymbol{\theta}}_m^{(b)}-\widehat{\boldsymbol{\theta}}_m)/2\\
&=\sqrt{B}\begin{bmatrix}\mathrm{E}[\sigma_i^2(\widehat{\boldsymbol{\theta}}_m^{(b)})|\widehat{\boldsymbol{\theta}}_m]^{-1/2}(\widehat{\boldsymbol{\theta}}_m^{(b)}-\widehat{\boldsymbol{\theta}}_m)^\top\\ \mathrm{E}[\sigma_i^2(\widehat{\boldsymbol{\theta}}_m^{(b)})(\widehat{\boldsymbol{\theta}}_m^{(b)}-\widehat{\boldsymbol{\theta}}_m)(\widehat{\boldsymbol{\theta}}_m^{(b)}-\widehat{\boldsymbol{\theta}}_m)^\top|\widehat{\boldsymbol{\theta}}_m]^{-1/2}(\widehat{\boldsymbol{\theta}}_m^{(b)}-\widehat{\boldsymbol{\theta}}_m)(\widehat{\boldsymbol{\theta}}_m^{(b)}-\widehat{\boldsymbol{\theta}}_m)^\top\end{bmatrix}\nabla^2\eta_i(\ddot{\boldsymbol{\theta}}^{(b)})(\widehat{\boldsymbol{\theta}}_m^{(b)}-\widehat{\boldsymbol{\theta}}_m)/2\\
&=\sqrt{B}\mathbf{O}_p(m^{-1/2})\nabla^2\eta_i\left(\ddot{\boldsymbol{\theta}}^{(b)}\right)\mathbf{O}_p(m^{-1/2})/2.
\end{aligned}\tag{EC.13}$$

Since $\widehat{\boldsymbol{\theta}}_m^{(b)}\xrightarrow{p}\widehat{\boldsymbol{\theta}}_m$ and $\widehat{\boldsymbol{\theta}}_m\to\boldsymbol{\theta}^c$ as $m\to\infty$, $\ddot{\boldsymbol{\theta}}^{(b)}$ is included in the neighborhood of $\boldsymbol{\theta}^c$ where $\nabla^2\eta_i(\ddot{\boldsymbol{\theta}}^{(b)})$ is continuous, in probability. This makes $\nabla^2\eta_i(\ddot{\boldsymbol{\theta}}^{(b)})$ bounded in probability as m increases. As $W^{-1}(\widehat{\boldsymbol{\theta}}_m)\sqrt{B}\mathbf{X}^\top\mathbf{h}_i/B$ is a sample average of B i.i.d. observations of (EC.13) $W(\widehat{\boldsymbol{\theta}}_m)\sqrt{B}\mathbf{X}^\top\mathbf{h}_i/B\Big|\widehat{\boldsymbol{\theta}}_m\xrightarrow{p}\mathbf{0}$. Thus, from (i)–(iii), we obtain the resulting limiting distribution in the proposition. $\square$

Proposition 3 only requires $0<\gamma<2$ for asymptotic normality of the gradient estimator. In particular, the upper bound, $\gamma<2$, is only needed to ensure Part (i) in the proof of Proposition 3. For $\widehat{\beta}_i\xrightarrow{p}\beta_i$, we need γ to be greater than 1. Although consistency of $\widehat{\beta}_i$ is not needed for IOU-C procedures, this result is stated in Proposition 1 in Section 5.1.1 (restated below) for its independent interest.

PROPOSITION 1. *As $m\to\infty$, for all i 1) $\sqrt{B/m}(\widehat{\mathcal{B}}_i-\mathcal{B}_i)\Big|\widehat{\boldsymbol{\theta}}$ converges weakly uniformly to $\mathrm{N}\left(\mathbf{0},V_i(\boldsymbol{\theta}^c)\otimes\Sigma^{-1}(\boldsymbol{\theta}^c)\right)$ in probability, if $B=m^\gamma$ for $0<\gamma<2$, and 2) $\widehat{\beta}_i\xrightarrow{p}\beta_i$, if $B=m^\gamma$ for $\gamma>1$.*

Proof. The proof of Part 1) follows by replacing $\boldsymbol{\beta}_i,\tilde{\boldsymbol{\beta}}_i$, and $(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}_i$ in the proof of Proposition 3 with

$$\begin{pmatrix}\eta_i(\boldsymbol{\theta}^c)-\eta_1(\boldsymbol{\theta}^c)\\ \vdots\\ \eta_i(\boldsymbol{\theta}^c)-\eta_{i-1}(\boldsymbol{\theta}^c)\\ \eta_i(\boldsymbol{\theta}^c)-\eta_{i+1}(\boldsymbol{\theta}^c)\\ \vdots\\ \eta_i(\boldsymbol{\theta}^c)-\eta_k(\boldsymbol{\theta}^c)\\ \nabla\eta_i(\boldsymbol{\theta}^c)-\nabla\eta_1(\boldsymbol{\theta}^c)\\ \vdots\\ \nabla\eta_i(\boldsymbol{\theta}^c)-\nabla\eta_{i-1}(\boldsymbol{\theta}^c)\\ \nabla\eta_i(\boldsymbol{\theta}^c)-\nabla\eta_{i+1}(\boldsymbol{\theta}^c)\\ \vdots\\ \nabla\eta_i(\boldsymbol{\theta}^c)-\nabla\eta_k(\boldsymbol{\theta}^c)\end{pmatrix},\begin{pmatrix}\eta_i(\widehat{\boldsymbol{\theta}}_m)-\eta_1(\widehat{\boldsymbol{\theta}}_m)\\ \vdots\\ \eta_i(\widehat{\boldsymbol{\theta}}_m)-\eta_{i-1}(\widehat{\boldsymbol{\theta}}_m)\\ \eta_i(\widehat{\boldsymbol{\theta}}_m)-\eta_{i+1}(\widehat{\boldsymbol{\theta}}_m)\\ \vdots\\ \eta_i(\widehat{\boldsymbol{\theta}}_m)-\eta_k(\widehat{\boldsymbol{\theta}}_m)\\ \nabla\eta_i(\widehat{\boldsymbol{\theta}}_m)-\nabla\eta_1(\widehat{\boldsymbol{\theta}}_m)\\ \vdots\\ \nabla\eta_i(\widehat{\boldsymbol{\theta}}_m)-\nabla\eta_{i-1}(\widehat{\boldsymbol{\theta}}_m)\\ \nabla\eta_i(\widehat{\boldsymbol{\theta}}_m)-\nabla\eta_{i+1}(\widehat{\boldsymbol{\theta}}_m)\\ \vdots\\ \nabla\eta_i(\widehat{\boldsymbol{\theta}}_m)-\nabla\eta_k(\widehat{\boldsymbol{\theta}}_m)\end{pmatrix},\text{ and }\mathbf{I}_{k-1}\otimes(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\begin{pmatrix}\mathbf{Y}_i-\mathbf{Y}_1\\ \vdots\\ \mathbf{Y}_i-\mathbf{Y}_{i-1}\\ \mathbf{Y}_i-\mathbf{Y}_{i+1}\\ \vdots\\ \mathbf{Y}_i-\mathbf{Y}_k\end{pmatrix},$$

respsectively.

From Proposition 3, it is easy to see that we need $\gamma>1$ for $\widehat{\beta}_i\xrightarrow{p}\beta_i$. For consistency, we need to show (i′) $(\tilde{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i) \xrightarrow{p} \mathbf{0}$.

(ii′) $(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\varepsilon}_i \big| \widehat{\boldsymbol{\theta}}_m \xrightarrow{p} \mathbf{0}$.

(iii′) $(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{h}_i \big| \widehat{\boldsymbol{\theta}}_m \xrightarrow{p} \mathbf{0}$.

As $m \to \infty$, (i′) holds and (ii′) and (iii′) hold when $\gamma > 1$ directly from (ii) and (iii) in Proposition 3. □

Next, we use the asymptotic normality of $\widehat{\mathcal{B}}_i$ established above to obtain the joint asymptotic coverage probability of $\mathrm{CR}_{1,\alpha_{11}}$ and $\mathrm{CR}_{2,\alpha_{12}}$ in $\mathcal{P}_{i\ell}$. We start with showing consistency of $\widehat{\tau}^2(\widehat{\boldsymbol{\theta}}_m)$.

Lemma 5. If $B = m^\gamma, 0 < \gamma < 2$, then $\widehat{V}_i(\widehat{\boldsymbol{\theta}}_m)|\widehat{\boldsymbol{\theta}}_m \xrightarrow{p} V_i(\boldsymbol{\theta}^c)$.

Proof. Recall that

$$\widehat{V}_i = \sum_{b=1}^{B} (\mathbf{e}_{-i,b} - \bar{\mathbf{e}}_{-i})(\mathbf{e}_{-i,b} - \bar{\mathbf{e}}_{-i})^\top/(B-1),$$

where $\mathbf{e}_{-i,b} = (e_{1b}, e_{2b}, \cdots, e_{(i-1)b}, e_{(i+1)b}, \cdots, e_{kb})^\top - e_{ib}\mathbf{1}_{k-1}$ and $\bar{\mathbf{e}}_{-i} = \sum_{b=1}^{B}\mathbf{e}_{-ib}/B$. Note that

$$\begin{aligned} e_{ib} &= Y_i(\widehat{\boldsymbol{\theta}}_m^{(b)}) - \widehat{\boldsymbol{\beta}}_i^\top \begin{bmatrix} 1 \\ \widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m \end{bmatrix} \\ &= \varepsilon_i(\widehat{\boldsymbol{\theta}}_m^{(b)}) + \eta_i(\widehat{\boldsymbol{\theta}}_m^{(b)}) - \tilde{\boldsymbol{\beta}}_i^\top \begin{bmatrix} 1 \\ \widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m \end{bmatrix} + \tilde{\boldsymbol{\beta}}_i^\top \begin{bmatrix} 1 \\ \widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m \end{bmatrix} - \widehat{\boldsymbol{\beta}}_i^\top \begin{bmatrix} 1 \\ \widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m \end{bmatrix}, \end{aligned}$$

where $\widehat{\boldsymbol{\beta}}_i$ and $\tilde{\boldsymbol{\beta}}_i$ are defined in Proposition 3. For sufficiently large m, $\eta_i(\widehat{\boldsymbol{\theta}}_m^{(b)}) = \eta_i(\widehat{\boldsymbol{\theta}}_m) + \nabla\eta_i(\widehat{\boldsymbol{\theta}}_m)^\top(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m) + o(||\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m||)$ from Assumption 1(iii). Therefore,

$$\eta_i(\widehat{\boldsymbol{\theta}}_m^{(b)}) - \tilde{\boldsymbol{\beta}}_i^\top \begin{bmatrix} 1 \\ \widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m \end{bmatrix} = o(||\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m||),$$

and

$$e_{ib} = \varepsilon_i(\widehat{\boldsymbol{\theta}}_m^{(b)}) + o(||\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m||) + \left(\tilde{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}}_i\right)^\top \begin{bmatrix} 1 \\ \widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m \end{bmatrix}. \tag{EC.14}$$

Define $\bar{e}_i = \sum_{b=1}^{B} e_{ib}/B$ and $\bar{\varepsilon}_i = \sum_{b=1}^{B}\varepsilon_{ib}(\widehat{\boldsymbol{\theta}}_m^{(b)})/B$ for $i = 1, 2, \ldots, k$. Using (EC.14) for $i \neq \ell$

$$\begin{aligned} e_{ib} - e_{\ell b} - (\bar{e}_i - \bar{e}_\ell) &= \varepsilon_i(\widehat{\boldsymbol{\theta}}_m^{(b)}) - \varepsilon_\ell(\widehat{\boldsymbol{\theta}}_m^{(b)}) - (\bar{\varepsilon}_i - \bar{\varepsilon}_\ell) + o(||\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m||) + \sum_{b=1}^{B} o(||\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m||)/B \\ &\quad + \left(\tilde{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}}_i - (\tilde{\boldsymbol{\beta}}_\ell - \widehat{\boldsymbol{\beta}}_\ell)\right)^\top \left(\begin{bmatrix} 1 \\ \widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m \end{bmatrix} - \frac{1}{B}\sum_{b=1}^{B} \begin{bmatrix} 1 \\ \widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m \end{bmatrix} \right). \end{aligned} \tag{EC.15}$$

Suppose we rewrite $\widehat{V}_i$ as $\widehat{V}_i = \sum_{b=1}^{B}(\boldsymbol{\varepsilon}_{-i,b} - \bar{\boldsymbol{\varepsilon}}_{-i})(\boldsymbol{\varepsilon}_{-i,b} - \bar{\boldsymbol{\varepsilon}}_{-i})^\top/(B-1) + \mathcal{R}$ using (EC.15), where $\boldsymbol{\varepsilon}_{-i,b}$ and $\bar{\boldsymbol{\varepsilon}}_{-i}$ are defined similarly as $\mathbf{e}_{-i,b}$ and $\bar{\mathbf{e}}_{-i}$, respectively. Then, $\sum_{b=1}^{B}(\boldsymbol{\varepsilon}_{-i,b} - \bar{\boldsymbol{\varepsilon}}_{-i})(\boldsymbol{\varepsilon}_{-i,b} - \bar{\boldsymbol{\varepsilon}}_{-i})^\top/(B-1)$ is an unbiased estimator of $\mathrm{Var}(\boldsymbol{\varepsilon}_{-i,b}|\widehat{\boldsymbol{\theta}}_m)$. Since $\mathrm{E}[\varepsilon_i^4(\widehat{\boldsymbol{\theta}}_m^{(b)})|\widehat{\boldsymbol{\theta}}_m]$ is bounded over $\widehat{\boldsymbol{\theta}}_m \in \Theta$ by Assumption 1(iii)

$$\left.\sum_{b=1}^{B}(\boldsymbol{\varepsilon}_{-i,b} - \bar{\boldsymbol{\varepsilon}}_{-i})(\boldsymbol{\varepsilon}_{-i,b} - \bar{\boldsymbol{\varepsilon}}_{-i})^\top/(B-1)\right| \widehat{\boldsymbol{\theta}}_m \xrightarrow{p} \mathrm{Var}(\boldsymbol{\varepsilon}_{-i,b}|\widehat{\boldsymbol{\theta}}_m).$$

as $B \to \infty$. Corollary 1 shows that $\mathrm{Var}(\boldsymbol{\varepsilon}_{-i,b}|\widehat{\boldsymbol{\theta}}_m) \xrightarrow{a.s.} V_i(\boldsymbol{\theta}^c)$ as $m \to \infty$, which implies $\sum_{b=1}^{B}(\boldsymbol{\varepsilon}_{-i,b} - \bar{\boldsymbol{\varepsilon}}_{-i})(\boldsymbol{\varepsilon}_{-i,b} - \bar{\boldsymbol{\varepsilon}}_{-i})^\top/(B-1)\Big|\,\widehat{\boldsymbol{\theta}}_m \xrightarrow{p} V_i(\boldsymbol{\theta}^c)$. Also, since $o(||\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m||^2)|\widehat{\boldsymbol{\theta}}_m \xrightarrow{p} 0$ from the distribution of $\widehat{\boldsymbol{\theta}}_m^{(b)}$ and $\left((\nabla\eta_i(\widehat{\boldsymbol{\theta}}_m) - \widehat{\beta}_i)^\top(\widehat{\boldsymbol{\theta}}_m^{(b)} - \widehat{\boldsymbol{\theta}}_m)\right)^2\Big|\,\widehat{\boldsymbol{\theta}}_m \xrightarrow{p} 0$ from the proof of Proposition 3, we can guarantee $\mathcal{R}|\widehat{\boldsymbol{\theta}}_m \xrightarrow{p} 0$. Therefore, we have $\widehat{V}_i(\widehat{\boldsymbol{\theta}}_m)|\widehat{\boldsymbol{\theta}}_m \xrightarrow{p} V_i(\boldsymbol{\theta}^c)$. $\square$

LEMMA 6. For each i, if $B = m^\gamma, 0 < \gamma < 2$,

$$\Pr\Big\{(\mathcal{B}_i - \widehat{\mathcal{B}}_i)^\top \mathcal{V}_i^{-1}(\widehat{\boldsymbol{\theta}}_m)(\mathcal{B}_i - \widehat{\mathcal{B}}_i) \le \chi^2_{(k-1)p,\alpha_{11}},\\ m(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}^c)^\top \Sigma(\widehat{\boldsymbol{\theta}}_m)^{-1}(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}^c) \le \chi^2_{p,\alpha_{12}}\Big\} \to (1-\alpha_{11})(1-\alpha_{12}),$$

as $m \to \infty$.

Proof. First note that the given probability in the lemma can be rewritten as

$$\begin{aligned}
&\mathrm{E}\Big[\Pr\Big\{(\mathcal{B}_i - \widehat{\mathcal{B}}_i)^\top \mathcal{V}_i^{-1}(\widehat{\boldsymbol{\theta}}_m)(\mathcal{B}_i - \widehat{\mathcal{B}}_i) \le \chi^2_{(k-1)p,\alpha_{11}},\\
&\qquad\qquad m(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}^c)^\top \Sigma(\widehat{\boldsymbol{\theta}}_m)^{-1}(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}^c) \le \chi^2_{p,\alpha_{12}}\Big|\widehat{\boldsymbol{\theta}}_m\Big\}\Big]\\
&= \mathrm{E}\Big[\Pr\Big\{(\mathcal{B}_i - \widehat{\mathcal{B}}_i)^\top \mathcal{V}^{-1}(\widehat{\boldsymbol{\theta}}_m)(\mathcal{B}_i - \widehat{\mathcal{B}}_i) \le \chi^2_{(k-1)p,\alpha_{11}}\Big|\widehat{\boldsymbol{\theta}}_m\Big\}\\
&\qquad\qquad \times \mathbf{1}\{m(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}^c)^\top \Sigma(\widehat{\boldsymbol{\theta}}_m)^{-1}(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}^c) \le \chi^2_{p,\alpha_{12}}\}\Big]
\end{aligned} \tag{EC.16}$$

The equality holds because conditional on $\widehat{\boldsymbol{\theta}}_m$, the two events are independent. From Lemma 2,

$$B/m\mathbf{S}_{pp} = B/m(\mathbf{C}^\top(\mathbf{I}_B - \mathbf{1}\mathbf{1}^\top/B)\mathbf{C})^{-1} \xrightarrow{p} \Sigma^{-1}(\boldsymbol{\theta}^c)$$

because $\mathbf{I}_B - \mathbf{1}\mathbf{1}^\top/B \to \mathbf{I}_B$ as $B \to \infty$. Since $\mathcal{V}_i(\widehat{\boldsymbol{\theta}}_m) \equiv \widehat{V}_i(\widehat{\boldsymbol{\theta}}_m) \otimes \mathbf{S}_{pp}$,

$$\frac{B}{m}\mathcal{V}_i(\widehat{\boldsymbol{\theta}}_m) \xrightarrow{p} V_i(\boldsymbol{\theta}^c) \otimes \Sigma^{-1}(\boldsymbol{\theta}^c) \tag{EC.17}$$

by Lemma 5. Combining (EC.17) with Proposition 1 and by Slutsky's theorem

$$\frac{B}{m}(\mathcal{B}_i - \widehat{\mathcal{B}}_i)^\top \left(\frac{B}{m}\mathcal{V}_i(\widehat{\boldsymbol{\theta}}_m)\right)^{-1}(\mathcal{B}_i - \widehat{\mathcal{B}}_i)\Big|\,\widehat{\boldsymbol{\theta}}_m \xrightarrow{D} \chi^2_{(k-1)p}$$

in probability over $\widehat{\boldsymbol{\theta}}_m$. Therefore,

$$\lim_{m\to\infty}\Pr\Big\{(\mathcal{B}_i - \widehat{\mathcal{B}}_i)^\top \mathcal{V}_i^{-1}(\widehat{\boldsymbol{\theta}}_m)(\mathcal{B}_i - \widehat{\mathcal{B}}_i) \le \chi^2_{(k-1)p,\alpha_{11}}\Big|\widehat{\boldsymbol{\theta}}_m\Big\} = 1 - \alpha_{11}.$$

Note that we implicitly exchanged the order of the limit and the expectation (probability) above, which is allowed by the dominated convergence theorem since the probability is bounded by 1. As a result, the limit of (EC.16) becomes

$$(1-\alpha_{11})\Pr\{m(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}^c)^\top \Sigma(\widehat{\boldsymbol{\theta}}_m)^{-1}(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}^c) \le \chi^2_{p,\alpha_{12}}\} \to (1-\alpha_{11})(1-\alpha_{12})$$

by Assumption 1(ii) as $m \to \infty$. $\square$

Lemma 6 implies that $w_{i\ell}^{(1)}, i\neq\ell$ obtained by solving $\mathcal{P}_{i\ell}, i\neq\ell$ asymptotically provide $1-\alpha_1$ MCB CIs of the CID effects represented by Model (5). The following lemma connects the coverage probability of $w_{i\ell}^{(1)}, i\neq\ell$ with that of $w_{i\ell}^{(2)}, i\neq\ell$.

LEMMA 7. Given $B=m^\gamma, 0<\gamma<2$, and $w_{i\ell}^{(1)}$ and $w_{i\ell}^{(2)}, i\neq\ell$, from the all-in IOU-C, if for each i,

$$\Pr\{\beta_i^\top(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^c)-\beta_\ell^\top(\widehat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}^c)\geq -w_{i\ell}^{(1)},\forall\ell\neq i\}\to 1-\alpha_1$$

as $m\to\infty$, then

$$\Pr\{\bar{Y}_i(\widehat{\boldsymbol{\theta}}_m)-\bar{Y}_\ell(\widehat{\boldsymbol{\theta}}_m)-(\eta_i(\boldsymbol{\theta}^c)-\eta_\ell(\boldsymbol{\theta}^c))\geq -(w_{i\ell}^{(1)}+w_{i\ell}^{(2)}),\forall\ell\neq i\}\to(1-\alpha_1)(1-\alpha_2).$$

as $n\to\infty, m\to\infty$.

Proof. For each i

$$\begin{aligned}
&\Pr\left\{\bar{Y}_i(\widehat{\boldsymbol{\theta}}_m)-\bar{Y}_\ell(\widehat{\boldsymbol{\theta}}_m)-(\eta_i(\boldsymbol{\theta}^c)-\eta_\ell(\boldsymbol{\theta}^c))\geq -(w_{i\ell}^{(1)}+w_{i\ell}^{(2)}),\forall\ell\neq i\right\}\\
&\geq\Pr\left\{\eta_i(\widehat{\boldsymbol{\theta}}_m)-\eta_\ell(\widehat{\boldsymbol{\theta}}_m)-(\eta_i(\boldsymbol{\theta}^c)-\eta_\ell(\boldsymbol{\theta}^c))\geq -w_{i\ell}^{(1)},\bar{\varepsilon}_i(\widehat{\boldsymbol{\theta}}_m)-\bar{\varepsilon}_\ell(\widehat{\boldsymbol{\theta}}_m)\geq -w_{i\ell}^{(2)},\forall\ell\neq i\right\}\\
&=\mathrm{E}\left[\mathrm{E}\left[\mathbf{1}\left\{\eta_i(\widehat{\boldsymbol{\theta}}_m)-\eta_\ell(\widehat{\boldsymbol{\theta}}_m)-(\eta_i(\boldsymbol{\theta}^c)-\eta_\ell(\boldsymbol{\theta}^c))\geq -w_{i\ell}^{(1)},\bar{\varepsilon}_i(\widehat{\boldsymbol{\theta}})-\bar{\varepsilon}_\ell(\widehat{\boldsymbol{\theta}}_m)\geq -w_{i\ell}^{(2)},\forall\ell\neq i\right\}\middle|\widehat{\boldsymbol{\theta}}_m\right]\right]\\
&=\mathrm{E}\Bigg[\mathbf{1}\left\{\eta_i(\widehat{\boldsymbol{\theta}}_m)-\eta_\ell(\widehat{\boldsymbol{\theta}}_m)-(\eta_i(\boldsymbol{\theta}^c)-\eta_\ell(\boldsymbol{\theta}^c))\geq -w_{i\ell}^{(1)},\forall\ell\neq i\right\}\\
&\qquad\times\mathrm{E}\left[\mathbf{1}\left\{\bar{\varepsilon}_i(\widehat{\boldsymbol{\theta}}_m)-\bar{\varepsilon}_\ell(\widehat{\boldsymbol{\theta}}_m)\geq -w_{i\ell}^{(2)},\forall\ell\neq i\right\}\middle|\widehat{\boldsymbol{\theta}}_m\right]\Bigg]\\
&=\mathrm{E}\Bigg[\mathbf{1}\left\{\eta_i(\widehat{\boldsymbol{\theta}}_m)-\eta_\ell(\widehat{\boldsymbol{\theta}}_m)-(\eta_i(\boldsymbol{\theta}^c)-\eta_\ell(\boldsymbol{\theta}^c))\geq -w_{i\ell}^{(1)},\forall\ell\neq i\right\}\\
&\qquad\times\Pr\left\{\bar{\varepsilon}_i(\widehat{\boldsymbol{\theta}}_m)-\bar{\varepsilon}_\ell(\widehat{\boldsymbol{\theta}}_m)\geq -w_{i\ell}^{(2)},\forall\ell\neq i\middle|\widehat{\boldsymbol{\theta}}_m\right\}\Bigg]
\end{aligned}$$

where the second equality holds, since conditional on $\widehat{\boldsymbol{\theta}}_m$, $\eta_i(\widehat{\boldsymbol{\theta}}_m),\forall i$, are constants. We first show the following

$$\begin{aligned}
&\lim_{m\to\infty}\lim_{n\to\infty}\mathrm{E}\Bigg[\mathbf{1}\left\{\eta_i(\widehat{\boldsymbol{\theta}}_m)-\eta_\ell(\widehat{\boldsymbol{\theta}}_m)-(\eta_i(\boldsymbol{\theta}^c)-\eta_\ell(\boldsymbol{\theta}^c))\geq -w_{i\ell}^{(1)},\forall\ell\neq i\right\}\\
&\qquad\times\Pr\left\{\bar{\varepsilon}_i(\widehat{\boldsymbol{\theta}}_m)-\bar{\varepsilon}_\ell(\widehat{\boldsymbol{\theta}}_m)\geq -w_{i\ell}^{(2)},\forall\ell\neq i\middle|\widehat{\boldsymbol{\theta}}_m\right\}\Bigg]\\
&=\lim_{m\to\infty}\mathrm{E}\Bigg[\mathbf{1}\left\{\eta_i(\widehat{\boldsymbol{\theta}}_m)-\eta_\ell(\widehat{\boldsymbol{\theta}}_m)-(\eta_i(\boldsymbol{\theta}^c)-\eta_\ell(\boldsymbol{\theta}^c))\geq -w_{i\ell}^{(1)},\forall\ell\neq i\right\}\\
&\qquad\times\lim_{n\to\infty}\Pr\left\{\bar{\varepsilon}_i(\widehat{\boldsymbol{\theta}}_m)-\bar{\varepsilon}_\ell(\widehat{\boldsymbol{\theta}}_m)\geq -w_{i\ell}^{(2)},\forall\ell\neq i\middle|\widehat{\boldsymbol{\theta}}_m\right\}\Bigg] &&\text{(EC.18)}\\
&=(1-\alpha_2)\lim_{m\to\infty}\Pr\left\{\eta_i(\widehat{\boldsymbol{\theta}}_m)-\eta_\ell(\widehat{\boldsymbol{\theta}}_m)-(\eta_i(\boldsymbol{\theta}^c)-\eta_\ell(\boldsymbol{\theta}^c))\geq -w_{i\ell}^{(1)},\forall\ell\neq i\right\}\\
&=(1-\alpha_2)\lim_{m\to\infty}\Pr\left\{(\nabla\eta_i(\boldsymbol{\theta}^c)-\nabla\eta_\ell(\boldsymbol{\theta}^c))^\top(\widehat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}^c)-o(||\widehat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}^c||)\geq -w_{i\ell}^{(1)},\forall\ell\neq i\right\} &&\text{(EC.19)}\\
&=(1-\alpha_2)\lim_{m\to\infty}\Pr\left\{(\beta_i-\beta_\ell)^\top(\widehat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}^c)-o(||\widehat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}^c||)\geq -w_{i\ell}^{(1)},\forall\ell\neq i\right\}\\
&=(1-\alpha_1)(1-\alpha_2).
\end{aligned}$$

In (EC.18), we exchanged the order of the limit and the expectation, which is justified by the dominated convergence theorem. Also, (EC.19) holds since $\eta_i(\cdot)$ is twice differentiable at $\boldsymbol{\theta}^c$.

Now, we show that the same limiting probability is obtained when we change the order of limits.

$$\begin{aligned}
&\lim_{n\to\infty}\lim_{m\to\infty}\mathrm{E}\Bigg[\mathbf{1}\left\{\eta_i(\widehat{\boldsymbol{\theta}}_m)-\eta_\ell(\widehat{\boldsymbol{\theta}}_m)-(\eta_i(\boldsymbol{\theta}^c)-\eta_\ell(\boldsymbol{\theta}^c))\geq -w_{i\ell}^{(1)},\forall\ell\neq i\right\}\\
&\qquad\qquad\times\Pr\left\{\bar{\varepsilon}_i(\widehat{\boldsymbol{\theta}}_m)-\bar{\varepsilon}_\ell(\widehat{\boldsymbol{\theta}}_m)\geq -w_{i\ell}^{(2)},\forall\ell\neq i\,\middle|\,\widehat{\boldsymbol{\theta}}_m\right\}\Bigg]\\
&=\lim_{n\to\infty}\mathrm{E}\Bigg[\lim_{m\to\infty}\mathbf{1}\left\{\eta_i(\widehat{\boldsymbol{\theta}}_m)-\eta_\ell(\widehat{\boldsymbol{\theta}}_m)-(\eta_i(\boldsymbol{\theta}^c)-\eta_\ell(\boldsymbol{\theta}^c))\geq -w_{i\ell}^{(1)},\forall\ell\neq i\right\}\\
&\qquad\qquad\times\Pr\left\{\bar{\varepsilon}_i(\widehat{\boldsymbol{\theta}}_m)-\bar{\varepsilon}_\ell(\widehat{\boldsymbol{\theta}}_m)\geq -w_{i\ell}^{(2)},\forall\ell\neq i\,\middle|\,\widehat{\boldsymbol{\theta}}_m\right\}\Bigg]\\
&=(1-\alpha_1)\lim_{n\to\infty}\Pr\left\{\bar{\varepsilon}_i(\boldsymbol{\theta}^c)-\bar{\varepsilon}_\ell(\boldsymbol{\theta}^c)\geq -w_{i\ell}^{(2)},\forall\ell\neq i\right\} \qquad\qquad \text{(EC.20)}\\
&=(1-\alpha_1)(1-\alpha_2).
\end{aligned}$$

In (EC.20), we used that $\varepsilon_i(\widehat{\boldsymbol{\theta}}_m)|\widehat{\boldsymbol{\theta}}_m\sim\mathrm{N}(0,\sigma_i^2(\widehat{\boldsymbol{\theta}}_m))$ and $\sigma_i^2(\cdot)$ is a continuous function by Assumption 1(iii), therefore, $\varepsilon_i(\widehat{\boldsymbol{\theta}}_m)|\widehat{\boldsymbol{\theta}}_m\xrightarrow{D}\varepsilon_i(\boldsymbol{\theta}^c)$ almost surely. □

Lastly, we present the proof of Theorem 2 below.

THEOREM 2. Under Assumption 1, if $B=m^\gamma$, $0<\gamma<2$, then as $m\to\infty, n\to\infty$ $\Pr\left\{\bar{Y}_i(\widehat{\boldsymbol{\theta}})-\bar{Y}_\ell(\widehat{\boldsymbol{\theta}})-(\eta_i(\boldsymbol{\theta}^c)-\eta_\ell(\boldsymbol{\theta}^c))\geq -w_{i\ell},\forall\ell\neq i\right\}\to(1-\alpha_1)(1-\alpha_2)$ given 1) $w_{i\ell}, i\neq\ell$, from the plug-in IOU-C procedure, or 2) $w_{i\ell}, i\neq\ell$, from the all-in IOU-C procedure.

Proof. From Proposition 1, $\widehat{\beta}_i\xrightarrow{p}\beta_i$. Therefore, the plug-in distribution of the CID effects converges to the asymptotic distribution of the CID effects by the continuous mapping theorem. Therefore, Part 1) follows from Assumption 1(vii). Part 2) follows directly by combining Lemma 6 and Lemma 7. □

References

Athreya K, Lahiri S (2006) *Measure Theory and Probability Theory* (New York, NY, USA: Springer).

Koeber M, Schäfer U (2006) The unique square root of a positive semidefinite matrix. *International Journal of Mathematical Education in Science & Technology* 37(8):990–992.

Li S (2011) Concise formulas for the area and volume of a hyperspherical cap. *Asian Journal of Mathematics and Statistics* 66–70.

Sidiropoulos P (2014) N-sphere chord length distribution, preprint, `https://arxiv.org/pdf/1411.5639.pdf`.

Sun H, Farooq M (2002) Note on the generation of random points uniformly distributed in hyper-ellipsoids. *Information Fusion, 2002. Proceedings of the Fifth International Conference on*, volume 1, 489–496.