Supplemental Appendix for the paper

Inference After Estimation of Breaks

Isaiah Andrews Toru Kitagawa Adam McCloskey

May 28, 2020

This Supplemental Appendix contains proofs and additional results for the paper "Inference After Estimation of Breaks." Section A collects proofs of the formal uniform asymptotic validity statements made in Section 4.2 of the main text. Section B contains a description of and theoretical results for confidence intervals based upon uniformly most accurate unbiased confidence intervals in the conditional norm-maximization problem. Finally, Section C presents additional Monte Carlo simulation results for the confidence intervals discussed both in the main text and in Section B of this Supplemental Appendix.

A Proofs of Uniform Asymptotic Validity Results

To prove uniformity in norm-maximization settings, we rely on some of the lemmas in AKM, along with a few additional results.

Lemma 1

Under Assumptions 2 and 4, for any sequence of confidence sets CS_n, any sequence of sets $\mathcal{C}_n(P)$ indexed by P, $C_n(P)=1\left\{\left(X_n,Y_n,\widehat{\Sigma}_n\right)\in\mathcal{C}_n(P)\right\}$, and any constant α, to show that

$$\limsup_{n\to\infty}\sup_{P\in\mathcal{P}_n}\left|Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS_n|C_n(P)=1\right\}-\alpha\right|Pr_P\{C_n(P)=1\}=0$$

it suffices to show that for all subsequences $\{n_s\}\subseteq\{n\}$, $\{P_{n_s}\}\in\mathcal{P}^\infty=\times_{n=1}^\infty\mathcal{P}_n$ with:

1. $\Sigma(P_{n_s})\to\Sigma^*\in\left\{\Sigma:1/\bar{\lambda}\leq\lambda_{\min}(\Sigma_X)\leq\lambda_{\max}(\Sigma_X)\leq\bar{\lambda},1/\bar{\lambda}\leq\Sigma_Y(\theta)\leq\bar{\lambda}\right\}$
2. $(\mu_{X,n_s}(P_{n_s}),\mu_{Y,n_s}(P_{n_s}))\to(\mu_X^*,\mu_Y^*)$ *for* (μ_X^*,μ_Y^*) *finite*

we have

$$\lim_{s\to\infty}Pr_{P_{n_s}}\left\{\hat{\mu}_{Y,n_s}\left(\hat{\theta}_{n_s};P_{n_s}\right)\in CS_{n_s}|C_{n_s}(P_{n_s})=1\right\}=\alpha.$$

Proof: Follows by the same argument as in the proof of Lemma 5 in AKM. □

To state the next lemma, for $Z_{\tilde{\theta},n,j}(\theta)$ the jth element of $Z_{\tilde{\theta},n}(\theta)$, let us define

$$A_n\left(\tilde{\theta},\theta\right)=\widehat{\Sigma}_{Y,n}\left(\tilde{\theta}\right)^{-2}\sum_{j=1}^{d_X}\left[\widehat{\Sigma}_{XY,n,j}\left(\tilde{\theta}\right)^2-\widehat{\Sigma}_{XY,n,j}\left(\theta,\tilde{\theta}\right)^2\right]$$

$$B_{Z,n}\left(\tilde{\theta},\theta\right)=2\widehat{\Sigma}_{Y,n}\left(\tilde{\theta}\right)^{-2}\sum_{j=1}^{d_X}\left[\widehat{\Sigma}_{XY,n,j}\left(\tilde{\theta}\right)Z_{\tilde{\theta},n,j}\left(\tilde{\theta}\right)-\widehat{\Sigma}_{XY,n,j}\left(\theta,\tilde{\theta}\right)Z_{\tilde{\theta},n,j}(\theta)\right]$$

$$C_{Z,n}\left(\tilde{\theta},\theta\right)=\sum_{j=1}^{d_X}\left[Z_{\tilde{\theta},n,j}\left(\tilde{\theta}\right)^2-Z_{\tilde{\theta},n,j}(\theta)^2\right],$$

$$D_{Z,n}\left(\tilde{\theta},\theta\right)=B_{Z,n}\left(\tilde{\theta},\theta\right)^2-4A_n\left(\tilde{\theta},\theta\right)C_{Z,n}\left(\tilde{\theta},\theta\right),$$

$$G_{Z,n}\left(\tilde{\theta},\theta\right)=\frac{-B_{Z,n}\left(\tilde{\theta},\theta\right)-\sqrt{D_{Z,n}\left(\tilde{\theta},\theta\right)}}{2A_n\left(\tilde{\theta},\theta\right)},K_{Z,n}\left(\tilde{\theta},\theta\right)=\frac{-B_{Z,n}\left(\tilde{\theta},\theta\right)+\sqrt{D_{Z,n}\left(\tilde{\theta},\theta\right)}}{2A_n\left(\tilde{\theta},\theta\right)}$$

and

$$H_{Z,n}\left(\tilde{\theta},\theta\right)=-\frac{C_{Z,n}\left(\tilde{\theta},\theta\right)}{B_{Z,n}\left(\tilde{\theta},\theta\right)}.$$

Based on these objects, let us further define

$$\ell^1_{Z,n}\left(\tilde{\theta}\right)=\max\left\{\max_{\theta\in\Theta:A_n\left(\tilde{\theta},\theta\right)<0,D_{Z,n}\left(\tilde{\theta},\theta\right)\geq 0}G_{Z,n}\left(\tilde{\theta},\theta\right),\max_{\theta\in\Theta:A_n\left(\tilde{\theta},\theta\right)=0,B_{Z,n}\left(\tilde{\theta},\theta\right)>0}H_{Z,n}\left(\tilde{\theta},\theta\right)\right\}$$

$$\ell^2_{Z,n}\left(\tilde{\theta},\theta\right)=\max\left\{\max_{\theta\in\Theta:A_n\left(\tilde{\theta},\theta\right)<0,D_{Z,n}\left(\tilde{\theta},\theta\right)\geq 0}G_{Z,n}\left(\tilde{\theta},\theta\right),\max_{\theta\in\Theta:A_n\left(\tilde{\theta},\theta\right)=0,B_{Z,n}\left(\tilde{\theta},\theta\right)>0}H_{Z,n}\left(\tilde{\theta},\theta\right),G_{Z,n}\left(\tilde{\theta},\theta\right)\right\}$$

$$u^1_{Z,n}\left(\tilde{\theta},\theta\right)=\min\left\{\min_{\theta\in\Theta:A_n\left(\tilde{\theta},\theta\right)<0,D_{Z,n}\left(\tilde{\theta},\theta\right)\geq 0}K_{Z,n}\left(\tilde{\theta},\theta\right),\min_{\theta\in\Theta:A_n\left(\tilde{\theta},\theta\right)=0,B_{Z,n}\left(\tilde{\theta},\theta\right)<0}H_{Z,n}\left(\tilde{\theta},\theta\right),K_{Z,n}\left(\tilde{\theta},\theta\right)\right\}$$

$$u^2_{Z,n}\left(\tilde{\theta}\right)=\min\left\{\min_{\theta\in\Theta:A_n\left(\tilde{\theta},\theta\right)<0,D_{Z,n}\left(\tilde{\theta},\theta\right)\geq 0}K_{Z,n}\left(\tilde{\theta},\theta\right),\min_{\theta\in\Theta:A_n\left(\tilde{\theta},\theta\right)=0,B_{Z,n}\left(\tilde{\theta},\theta\right)<0}H_{Z,n}\left(\tilde{\theta},\theta\right)\right\}.$$

Lemma 2

Under Assumptions 3 and 1, for any $\{n_s\}$ and $\{P_{n_s}\}$ satisfying conditions (1) and (2) of

Lemma 1,

$$\left(Y_{n_s},\hat{\Sigma}_{n_s},\hat{\theta}_{n_s},\ell^1_{Z,n_s}\left(\tilde{\theta}\right),\ell^2_{Z,n_s}\left(\tilde{\theta},\theta\right),u^1_{Z,n_s}\left(\tilde{\theta},\theta\right),u^2_{Z,n_s}\left(\tilde{\theta}\right)\right)$$
$$\to_d\left(Y^*,\Sigma^*,\hat{\theta},\ell^{1*}_Z\left(\tilde{\theta}\right),\ell^{2*}_Z\left(\tilde{\theta},\theta\right),u^{1*}_Z\left(\tilde{\theta},\theta\right),u^{2*}_Z\left(\tilde{\theta}\right)\right),$$

where the objects on the right hand side are calculated based on (X^*,Y^*,Σ^*) *for*

$$\begin{pmatrix} X^* \\ Y^* \end{pmatrix}\sim N(\mu^*,\Sigma^*).$$

Proof: Note that Assumption 1 along with condition (2) of Lemma 1 imply that

$$\begin{pmatrix} X_{n_s} \\ Y_{n_s} \end{pmatrix}\to_d\begin{pmatrix} X^* \\ Y^* \end{pmatrix}\sim N(\mu^*,\Sigma^*),$$

while Assumption 3 implies that $\widehat{\Sigma}_{n_s}\to_p\Sigma^*$.

If we define

$$\left(A^*\left(\tilde{\theta},\theta\right),B^*_Z\left(\tilde{\theta},\theta\right),C^*_Z\left(\tilde{\theta},\theta\right),D^*_Z\left(\tilde{\theta},\theta\right),G^*_Z\left(\tilde{\theta},\theta\right),K^*_Z\left(\tilde{\theta},\theta\right),H^*_Z\left(\tilde{\theta},\theta\right)\right)$$

as the analog of

$$\left(A_n\left(\tilde{\theta},\theta\right),B_{Z,n}\left(\tilde{\theta},\theta\right),C_{Z,n}\left(\tilde{\theta},\theta\right),D_{Z,n}\left(\tilde{\theta},\theta\right),G_{Z,n}\left(\tilde{\theta},\theta\right),K_{Z,n}\left(\tilde{\theta},\theta\right),H_{Z,n}\left(\tilde{\theta},\theta\right)\right)$$

based on (X^*,Y^*,Σ^*), the continuous mapping theorem implies that

$$\left(A_{n_s}\left(\tilde{\theta},\theta\right),B_{Z,n_s}\left(\tilde{\theta},\theta\right),C_{Z,n_s}\left(\tilde{\theta},\theta\right)\right)\to_d\left(A^*\left(\tilde{\theta},\theta\right),B^*_Z\left(\tilde{\theta},\theta\right),C^*_Z\left(\tilde{\theta},\theta\right)\right)$$

where this convergence holds jointly over all $\left(\theta,\tilde{\theta}\right)\in\Theta^2$. If $A^*\left(\tilde{\theta},\theta\right)\neq 0$, another application of the continuous mapping theorem implies that[17]

$$\left(D_{Z,n_s}\left(\tilde{\theta},\theta\right),G_{Z,n_s}\left(\tilde{\theta},\theta\right),K_{Z,n_s}\left(\tilde{\theta},\theta\right)\right)\to_d\left(D^*_Z\left(\tilde{\theta},\theta\right),G^*_Z\left(\tilde{\theta},\theta\right),K^*_Z\left(\tilde{\theta},\theta\right)\right).$$

[17] Note that we allow the possibility that $\left(D_{Z,n}\left(\tilde{\theta},\theta\right),D^*_Z\left(\tilde{\theta},\theta\right)\right)$ may be negative, so $\left(G_{Z,n}\left(\tilde{\theta},\theta\right),K_{Z,n}\left(\tilde{\theta},\theta\right)\right)$ and $\left(G^*_Z\left(\tilde{\theta},\theta\right),K^*_Z\left(\tilde{\theta},\theta\right)\right)$ may be complex-valued.

If instead $A^*\left(\tilde{\theta},\theta\right)=0$, note that

$$Z^*_{\tilde{\theta},j}(\theta)=X^*_j(\theta)-\frac{\Sigma^*_{XY,j}\left(\theta,\tilde{\theta}\right)}{\Sigma^*_Y\left(\tilde{\theta}\right)}Y^*\left(\tilde{\theta}\right)=X^*_j(\theta)-\frac{\Sigma^*_{XY,j}\left(\tilde{\theta}\right)}{\Sigma^*_Y\left(\tilde{\theta}\right)}Y^*\left(\tilde{\theta}\right).$$

Hence, in this setting

$$B^*_Z\left(\tilde{\theta},\theta\right)=2\Sigma_Y\left(\tilde{\theta}\right)^{-2}\sum_{j=1}^{d_X}\left[X^*_j\left(\tilde{\theta}\right)-X^*_j(\theta)\right]$$

and condition (1) of Lemma 1 implies that $Pr\left\{B^*_Z\left(\tilde{\theta},\theta\right)=0\right\}=0$ for all $\theta\neq\tilde{\theta}$. Hence, $Pr\left\{D^*_Z\left(\tilde{\theta},\theta\right)>0\right\}=1$. Moreover, note that for $b\neq 0$ and all c

$$\lim_{a\to 0}\frac{-b-\sqrt{b^2-4ac}}{2a}=\begin{cases}-\frac{c}{b} & \text{if } b<0\\ -\infty & \text{if } b>0\end{cases},$$

while

$$\lim_{a\to 0}\frac{-b+\sqrt{b^2-4ac}}{2a}=\begin{cases}\infty & \text{if } b<0\\ -\frac{c}{b} & \text{if } b>0\end{cases}.$$

Hence, if $A^*\left(\theta,\tilde{\theta}\right)=0$,

$$\frac{-B_{Z,n}\left(\tilde{\theta},\theta\right)-\sqrt{D_{Z,n}\left(\tilde{\theta},\theta\right)}}{2A_n\left(\tilde{\theta},\theta\right)}\to_d -\infty\cdot 1\left\{B^*_Z\left(\tilde{\theta},\theta\right)>0\right\}+H^*_Z\left(\tilde{\theta},\theta\right)$$

and

$$\frac{-B_{Z,n}\left(\tilde{\theta},\theta\right)+\sqrt{D_{Z,n}\left(\tilde{\theta},\theta\right)}}{2A_n\left(\tilde{\theta},\theta\right)}\to_d \infty\cdot 1\left\{B^*_Z\left(\tilde{\theta},\theta\right)<0\right\}+H^*_Z\left(\tilde{\theta},\theta\right),$$

with the convention that $\infty\cdot 0=0$. Finally, another application of the continuous mapping theorem shows that when $A^*\left(\tilde{\theta},\theta\right)=0$,

$$H_{Z,n_s}\left(\tilde{\theta},\theta\right)\to_d H^*_Z\left(\tilde{\theta},\theta\right).$$

Since all of these convergence results hold jointly over $\left(\theta,\tilde{\theta}\right)\in\Theta^2$, another application of the continuous mapping theorem implies that

$$\left(\ell^1_{Z,n_s}\left(\tilde{\theta}\right),\ell^2_{Z,n_s}\left(\tilde{\theta},\theta\right),u^1_{Z,n_s}\left(\tilde{\theta},\theta\right),u^2_{Z,n_s}\left(\tilde{\theta}\right)\right)\to_d\left(\ell^{1*}_Z\left(\tilde{\theta}\right),\ell^{2*}_Z\left(\tilde{\theta},\theta\right),u^{1*}_Z\left(\tilde{\theta},\theta\right),u^{2*}_Z\left(\tilde{\theta}\right)\right).$$

Moreover, $\hat{\theta}$ is almost everywhere continuous in X^*, so that $(Y_{n_s},\hat{\Sigma}_{n_s},\hat{\theta}_{n_s})\to_d(Y^*,\Sigma^*,\hat{\theta})$, where this convergence occurs jointly with that above. Thus, we have established the desired result. □

To state our next two lemmas, we consider sets that can be written as finite unions of disjoint intervals, $\mathcal{Y}^K=\cup_{k=1}^K\left[\ell^k,u^k\right]$.

Lemma 3

For $F_{TN}\left(\cdot;\mu,\Sigma_Y(\theta),\mathcal{Y}^K\right)$ the distribution function for ζ with

$$\zeta\sim\xi|\xi\in\mathcal{Y}^K,\xi\sim N(\mu,\Sigma_Y(\theta)),$$

$F_{TN}\left(Y(\theta);\mu,\Sigma_Y(\theta),\mathcal{Y}^K\right)$ is continuous on the set

$$\left\{\begin{array}{c}(Y(\theta),\mu,\Sigma_Y(\theta))\in\mathbb{R}^3,\ell^1\in[-\infty,\infty),\\ \left\{\ell^k\right\}_{k=2}^K\in\mathbb{R}^{K-1},\left\{u^k\right\}_{k=1}^{K-1}\in\mathbb{R}^{K-1},u^K\in(-\infty,\infty]\end{array}:\Sigma_Y(\theta)>0,\sum_k\left|u^k-\ell^k\right|>0,u^k\geq\ell^k\geq u^{k-1}\text{ for all }k\right\}.$$

Proof: Note that we can write

$$F_{TN}\left(Y(\theta);\mu,\Sigma_Y(\theta),\mathcal{Y}^K\right)=\frac{\sum_k 1\left\{Y(\theta)\geq\ell^k\right\}\left(F_N\left(\frac{u^k\wedge Y(\theta)-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-F_N\left(\frac{\ell^k-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)\right)}{\sum_k\left(F_N\left(\frac{u^k-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-F_N\left(\frac{\ell^k-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)\right)}.$$

Hence, we trivially obtain continuity for $\Sigma_Y(\theta)>0,Y(\theta)\in\mathbb{R},\mu\in\mathbb{R}$, $0<\sum_k\left|u^k-\ell^k\right|<\infty$. Moreover, as in the proof of Lemma 9 of AKM we retain continuity as we allow $\ell^1\to-\infty$ and/or $u^K\to\infty$, in the sense that for a sequence of sets $\mathcal{Y}^K_m$ with

$$\left\{\ell^k_m,u^k_m\right\}_{k=1}^K\to\left\{\ell^k_\infty,u^k_\infty\right\}_{k=1}^K$$

with $\ell^1_\infty=-\infty$ and/or $u^K_\infty=\infty$ and the other elements finite,

$$F_{TN}\left(Y(\theta);\mu,\Sigma_Y(\theta),\mathcal{Y}^K_m\right)\to F_{TN}\left(Y(\theta);\mu,\Sigma_Y(\theta),\mathcal{Y}^K_\infty\right).\quad\square$$

Lemma 4

Under Assumptions 1-4, for either $C_n = 1\left\{\hat{\theta}_n = \tilde{\theta}\right\}$ *or*

$$C_n = 1\left\{\hat{\theta}_n = \tilde{\theta}, \mu_{Y,n}\left(\hat{\theta}_n, P_n\right) \in CS^{\beta}_{P,n}\right\},$$

there exists $\varepsilon > 0$ *such that*

$$\liminf_{n\to\infty} \inf_{p\in\mathcal{P}_n} Pr_P\{C_n = 1\} \geq \varepsilon.$$

Hence, for any sequence of variables V_n,

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_n} |E_P[V_n|C_n = 1]| Pr_P\{C_n = 1\} = 0$$

if and only if

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_n} |E_P[V_n|C_n = 1]| = 0.$$

Proof of Lemma 4 By the same argument as in the proof of Lemma 5 in AKM, it suffices to consider sequences as in Lemma 1, where by Assumption 4,

$$\|\mu_X^*\| + \|\mu_Y^*\| \leq C.$$

Note, next, that for $(X^{*\prime}, Y^{*\prime})' \sim N((\mu_X^{*\prime}, \mu_Y^{*\prime})', \Sigma)$, Σ_X full-rank, and $\hat{\theta}^* = \text{argmax}_{\theta\in\Theta} \|X^*(\theta)\|$, $\hat{\theta}^*$ has full support. Moreover, $\hat{\theta}^*$ is almost everywhere continuous in X^*, so by the continuous mapping theorem, $\hat{\theta}_{n_s} \to_d \hat{\theta}^*$ under $\{n_s\}$, $\{P_{n_s}\}$. Moreover, $Pr\left\{\hat{\theta}^* = \tilde{\theta}\right\}$ is continuous in μ_X^* and Σ_X, and the set of μ_X^*, Σ_X values we consider is compact. Hence, $Pr\left\{\hat{\theta}^* = \tilde{\theta}\right\}$ is bounded away from zero, from which the bound for $C_n = 1\left\{\hat{\theta}_n = \tilde{\theta}\right\}$ follows. The claim for

$$C_n = 1\left\{\hat{\theta}_n = \tilde{\theta}, \mu_{Y,n}\left(\hat{\theta}_n, P_n\right) \in CS^{\beta}_{P,n}\right\},$$

follows by the same argument, using almost everywhere continuity of CS_P^{β} in the limit problem. The final claim is then immediate. □

Proof of Proposition 3 As in the proof of Proposition 9 of AKM, note that

$$\hat{\mu}_{\alpha,n} \geq \mu_{Y,n}\left(\hat{\theta}_n; P\right) \Longleftrightarrow \mu_{Y,n}\left(\hat{\theta}_n; P\right) \in CS_{U,-,n}$$

for $CS_{U,-,n}=(-\infty,\hat{\mu}_{\alpha,n}]$. Hence, by Lemmas 1 and 4, to prove that (11) holds it suffices to show that for all $\{n_s\}$ and $\{P_{n_s}\}$ such that conditions (1) and (2) of Lemma 1 hold with $C_n=1\left\{\hat{\theta}_n=\tilde{\theta}\right\}$, we have

$$\lim_{s\to\infty} Pr_{P_{n_s}}\left\{\hat{\mu}_{Y,n_s}\left(\hat{\theta}_{n_s};P_{n_s}\right)\in CS_{U,-,n_s}|\hat{\theta}_{n_s}=\tilde{\theta}\right\}=\alpha. \tag{16}$$

To this end, note that for $F_{TN}\left(Y(\theta);\mu,\Sigma_Y(\theta),\mathcal{Y}^K\right)$ as defined in the statement of Lemma 3, the estimator $\hat{\mu}_{\alpha,n}$ solves

$$F_{TN}\left(Y_n\left(\hat{\theta}_n\right);\mu,\widehat{\Sigma}_{Y,n}\left(\hat{\theta}_n\right),\mathcal{Y}_n\right)=1-\alpha,$$

for

$$\mathcal{Y}_n=\bigcap_{\theta\in\Theta:A_n\left(\tilde{\theta},\theta\right)>0,D_{Z,n}\left(\tilde{\theta},\theta\right)\geq 0}\left[\ell^1_{Z,n}\left(\tilde{\theta}\right),u^1_{Z,n}\left(\tilde{\theta},\theta\right)\right]\cap\left[\ell^2_{Z,n}\left(\tilde{\theta},\theta\right),u^2_{Z,n}\left(\tilde{\theta}\right)\right] \tag{17}$$

(see Proposition 1). The set $\mathcal{Y}_n$ can be written as a finite union of disjoint intervals by DeMorgan's Laws.

The cdf $F_{TN}\left(Y_n\left(\hat{\theta}_n\right);\mu,\widehat{\Sigma}_{Y,n}\left(\hat{\theta}_n\right),\mathcal{Y}_n\right)$ is strictly decreasing in μ as argued in the proof of Proposition 8 of AKM, and is increasing in $Y_n\left(\hat{\theta}\right)$. Hence, $\hat{\mu}_{\alpha,n}\geq\mu_{Y,n}\left(\hat{\theta}_n;P\right)$ if and only if

$$F_{TN}\left(Y_n\left(\hat{\theta}_n\right);\mu_{Y,n}\left(\hat{\theta}_n;P\right),\widehat{\Sigma}_{Y,n}\left(\hat{\theta}_n\right),\mathcal{Y}_n\right)\geq 1-\alpha.$$

Lemma 2 shows that $\left(Y_n\left(\hat{\theta}_{n_s}\right),\widehat{\Sigma}_{Y,n_s}\left(\hat{\theta}_{n_s}\right),\mathcal{Y}_{n_s},\hat{\theta}_{n_s}\right)$ converges in distribution as $s\to\infty$,[18] so since F_{TN} is continuous by Lemma 3 while $\arg\max_\theta\|X^*(\theta)\|$ is almost everywhere continuous for X^*, the continuous mapping theorem implies that

$$\begin{array}{c}\left(F_{TN}\left(Y_{n_s}\left(\hat{\theta}_{n_s}\right);\mu_{Y,n_s}\left(\tilde{\theta};P_{n_s}\right),\widehat{\Sigma}_{Y,n_s}\left(\hat{\theta}_{n_s}\right),\mathcal{Y}_{n_s}\right),1\left\{\hat{\theta}_{n_s}=\tilde{\theta}\right\}\right)\\ \to_d\left(F_{TN}\left(Y^*\left(\hat{\theta}\right);\mu_{Y,n_s}\left(\tilde{\theta};P_{n_s}\right),\Sigma^*_Y\left(\hat{\theta}\right),\mathcal{Y}^*\right),1\left\{\hat{\theta}=\tilde{\theta}\right\}\right)\end{array},$$

where $\mathcal{Y}^*$ is the analog of $\mathcal{Y}_n$ calculated based on (X^*,Y^*,Σ^*).

Since we can write

$$Pr_{P_{n_s}}\left\{F_{TN}\left(Y_{n_s}\left(\hat{\theta}_{n_s}\right);\mu_{Y,n_s}\left(\tilde{\theta};P_{n_s}\right),\widehat{\Sigma}_{Y,n_s}\left(\hat{\theta}_{n_s}\right),\mathcal{Y}_{n_s}\right)\geq 1-\alpha|\hat{\theta}_{n_s}=\tilde{\theta}\right\}$$

[18] Since $\mathcal{Y}_n$ can be represented as a finite union of intervals, we use $\mathcal{Y}_n\to_d\mathcal{Y}^*$ to denote joint convergence in distribution of (i) the number of intervals and (ii) the endpoints of the intervals.

$$=\frac{E_{P_{n_s}}\left[1\left\{F_{TN}\left(Y_{n_s}\left(\hat{\theta}_{n_s}\right);\mu_{Y,n_s}\left(\tilde{\theta};P_{n_s}\right),\widehat{\Sigma}_{Y,n_s}\left(\hat{\theta}_{n_s}\right),\mathcal{Y}_{n_s}\right)\geq 1-\alpha\right\}1\left\{\hat{\theta}_{n_s}=\tilde{\theta}\right\}\right]}{E_{P_{n_s}}\left[1\left\{\hat{\theta}_{n_s}=\tilde{\theta}\right\}\right]},$$

and by construction

$$F_{TN}\left(Y^*\left(\hat{\theta}\right);\mu_{Y,n_s}\left(\tilde{\theta};P_{n_s}\right),\Sigma_Y^*\left(\hat{\theta}\right),\mathcal{Y}^*,\hat{\theta}\right)|\hat{\theta}=\tilde{\theta}\sim U[0,1],$$

and $Pr\left\{\hat{\theta}=\tilde{\theta}\right\}=p^*>0$ by Assumption 4, we thus have that

$$Pr_{P_{n_s}}\left\{F_{TN}\left(Y_{n_s}\left(\hat{\theta}_{n_s}\right);\mu_{Y,n_s}\left(\tilde{\theta};P_{n_s}\right),\widehat{\Sigma}_{Y,n_s}\left(\hat{\theta}_{n_s}\right),\mathcal{Y}_{n_s}\right)\geq 1-\alpha|\hat{\theta}_{n_s}=\tilde{\theta}\right\}$$

$$\rightarrow Pr\left\{F_{TN}\left(Y^*\left(\hat{\theta}\right);\mu_Y^*\left(\tilde{\theta}\right),\Sigma_Y^*\left(\hat{\theta}\right),\mathcal{Y}^*\right)\geq 1-\alpha|\hat{\theta}=\tilde{\theta}\right\}=\alpha,$$

which verifies (16).

Since this argument holds for all $\tilde{\theta}\in\Theta$, and Assumptions 1 and 2 imply that for all $\theta,\tilde{\theta}\in\Theta$ with $\theta\neq\tilde{\theta}$,

$$\lim_{n\to\infty}\sup_{P\in\mathcal{P}_n}Pr_P\left\{\|X_n(\theta)\|=\left\|X_n\left(\tilde{\theta}\right)\right\|\right\}=0,$$

Lemma 6 of AKM implies (12). □

Proof of Corollary 2 Follows from Proposition 3 by the same argument used to prove Corollary 1 of AKM. □

Proof of Proposition 4 Follows by the same argument as in the proof of Proposition 11 of AKM. □

Proof of Proposition 5 Follows by an argument along the same lines as in the proof of Proposition 12 of AKM, using Lemmas 1, 2, 3, and 4 in place of Lemmas 5, 8, and 9 in AKM, and using the conditioning event $\{Y_n(\hat{\theta}_n)\in\mathcal{Y}_n^H\}=\{Y_n(\hat{\theta}_n)\in\mathcal{Y}_n\}\cap\left\{\mu_{Y,n}\left(\hat{\theta}_n,P_n\right)\in CS_{P,n}^{\beta}\right\}$. □

Proof of Corollary 3 Follows by the same argument as in the proof of Corollary 2 in AKM. □

B Uniformly Most Accurate Unbiased Confidence Intervals

This section provides uniform asymptotic results for the uniformly most accurate unbiased confidence sets of AKM, and related procedures, in the norm-maximization setting.

In the inference problem that conditions on $\hat{\theta}$ and $\hat{\gamma}$, rather than considering equal-tailed intervals, we can alternatively consider confidence intervals that are unbiased, in the sense that the probability of covering any given false parameter value is bounded above by $1-\alpha$. AKM develop a confidence interval, CS_U, that is uniformly most accurate unbiased in the conditional problem, in the sense that it has a weakly lower probability of covering any given incorrect parameter value than does any other unbiased confidence set while still maintaing correct conditional coverage $1-\alpha$:

$$Pr_{\mu}\left\{\mu_Y(\tilde{\theta})\in CS_U|\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma}\right\}=1-\alpha \text{ for all } \mu,\ \tilde{\theta},\ \tilde{\gamma}.$$

See AKM for the construction of these confidence sets.

The feasible versions of these intervals $CS_{U,n}$ are defined identically to CS_U after replacing $\hat{\theta}$ by $\hat{\theta}_n$, Y by Y_n, Σ by $\hat{\Sigma}_n$, and $Z_{\tilde{\theta}}$ by $Z_{\tilde{\theta},n}$. Here we establish that these feasible intervals have correct coverage both conditionally and unconditionally.

Proposition 6

Under Assumptions 1-4,

$$\lim_{n\to\infty}\sup_{P\in\mathcal{P}_n}\left|Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS_{U,n}|\hat{\theta}_n=\tilde{\theta}\right\}-(1-\alpha)\right|=0, \tag{18}$$

for all $\tilde{\theta}\in\Theta$, *and*

$$\lim_{n\to\infty}\sup_{P\in\mathcal{P}_n}\left|Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS_{U,n}\right\}-(1-\alpha)\right|=0. \tag{19}$$

Proof: Note that by the definition of $CS_{U,n}$

$$\begin{gathered}\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS_{U,n}\\ \Longleftrightarrow Y_n\left(\hat{\theta}_n\right)\in\left[c_l\left(\mu_{Y,n}\left(\hat{\theta}_n;P\right),\widehat{\Sigma}_{Y,n}\left(\hat{\theta}_n\right),\mathcal{Y}_n\right),c_u\left(\mu_{Y,n}\left(\hat{\theta}_n;P\right),\widehat{\Sigma}_{Y,n}\left(\hat{\theta}_n\right),\mathcal{Y}_n\right)\right]\end{gathered}$$

where $\mathcal{Y}_n$ is as defined in (17) while $(c_l(\mu,\Sigma_Y(\theta),\mathcal{Y}_n),c_u(\mu,\Sigma_Y(\theta),\mathcal{Y}_n))$ are as defined immediately before Lemma 5 below, after replacing $\mathcal{Y}^K$ with $\mathcal{Y}_n$.

By Lemmas 1 and 4, to prove that (18) holds it suffices to show that for all $\{n_s\}$ and $\{P_{n_s}\}$ satisfying conditions (1) and (2) of Lemma 1,

$$\lim_{s\to\infty}Pr_{P_{n_s}}\left\{\mu_{Y,n_s}\left(\hat{\theta}_{n_s}\right)\in CS_{U,n_s}|\hat{\theta}_{n_s}=\tilde{\theta}\right\}=1-\alpha.$$

Thus, it suffices to show that

$$\lim_{s\to\infty} Pr_{P_{n_s}}\left\{ Y_{n_s}\left(\hat{\theta}_{n_s}\right)\in \begin{array}{l}\left[c_l\left(\mu_{Y,n_s}\left(\hat{\theta},P_{n_s}\right),\widehat{\Sigma}_{Y,n_s}\left(\hat{\theta}_{n_s}\right),\mathcal{Y}_{n_s}\right),\right.\\ \left. c_u\left(\mu_{Y,n_s}\left(\hat{\theta},P_{n_s}\right),\widehat{\Sigma}_{Y,n_s}\left(\hat{\theta}_{n_s}\right),\mathcal{Y}_{n_s}\right)\right]\end{array} \middle| \hat{\theta}_{n_s}=\tilde{\theta}\right\} = 1-\alpha.$$

To this end, note that by Lemma 2,

$$\left(Y_{n_s},\mathcal{Y}_{n_s},\widehat{\Sigma}_{n_s},1\left\{\hat{\theta}_{n_s}=\tilde{\theta}\right\}\right)\to_d \left(Y^*,\mathcal{Y}^*,\Sigma^*,1\left\{\hat{\theta}=\tilde{\theta}\right\}\right),$$

and thus, by Lemma 5 and the continuous mapping theorem,[19]

$$\begin{aligned}&\left(Y_{n_s}\left(\tilde{\theta}\right),c_l\left(\mu_{Y,n_s}\left(\tilde{\theta},P_{n_s}\right),\widehat{\Sigma}_{Y,n_s}\left(\tilde{\theta}\right),\mathcal{Y}_{n_s}\right),c_u\left(\mu_{Y,n_s}\left(\tilde{\theta},P_{n_s}\right),\widehat{\Sigma}_{Y,n_s}\left(\tilde{\theta}\right),\mathcal{Y}_{n_s}\right),1\left\{\hat{\theta}_{n_s}=\tilde{\theta}\right\}\right)\\ &\quad\to_d \left(Y^*\left(\tilde{\theta}\right),c_l\left(\mu_Y^*\left(\tilde{\theta}\right),\Sigma_Y^*\left(\tilde{\theta}\right),\mathcal{Y}^*\right),c_u\left(\mu_Y^*\left(\tilde{\theta}\right),\Sigma_Y^*\left(\tilde{\theta}\right),\mathcal{Y}^*\right),1\left\{\hat{\theta}=\tilde{\theta}\right\}\right).\end{aligned}$$

By construction,

$$Pr\left\{Y^*\left(\tilde{\theta}\right)\in\left[c_l\left(\mu_Y^*\left(\tilde{\theta}\right),\mathcal{Y}^*,\Sigma_Y^*\left(\tilde{\theta}\right)\right),c_u\left(\mu_Y^*\left(\tilde{\theta}\right),\mathcal{Y}^*,\Sigma_Y^*\left(\tilde{\theta}\right)\right)\right]|\hat{\theta}=\tilde{\theta}\right\}=1-\alpha,$$

and $Y^*\left(\tilde{\theta}\right)|\hat{\theta}=\tilde{\theta},Y^*\left(\tilde{\theta}\right)\in\mathcal{Y}^*$ follows a truncated normal distribution, so

$$Pr\left\{Y^*\left(\tilde{\theta}\right)=c_l\left(\mu_Y^*\left(\tilde{\theta}\right),\Sigma_Y^*\left(\tilde{\theta}\right),\mathcal{Y}^*\right)\right\}=Pr\left\{Y^*\left(\tilde{\theta}\right)=c_u\left(\mu_Y^*\left(\tilde{\theta}\right),\Sigma_Y^*\left(\tilde{\theta}\right),\mathcal{Y}^*\right)\right\}=0.$$

Hence,

$$\begin{aligned}&Pr_{P_{n_s}}\left\{ Y_{n_s}\left(\hat{\theta}_{n_s}\right)\in \begin{array}{l}\left[c_l\left(\mu_{Y,n_s}\left(\tilde{\theta},P_{n_s}\right),\widehat{\Sigma}_{Y,n_s}\left(\hat{\theta}_{n_s}\right),\mathcal{Y}_{n_s}\right),\right.\\ \left. c_u\left(\mu_{Y,n_s}\left(\tilde{\theta},P_{n_s}\right),\widehat{\Sigma}_{Y,n_s}\left(\hat{\theta}_{n_s}\right),\mathcal{Y}_{n_s}\right)\right]\end{array} \middle| \hat{\theta}_{n_s}=\tilde{\theta}\right\}\\ &=\frac{E_{P_{n_s}}\left[1\left\{Y_{n_s}\left(\hat{\theta}_{n_s}\right)\in\left[c_l\left(\mu_{Y,n_s}\left(\tilde{\theta},P_{n_s}\right),\widehat{\Sigma}_{Y,n_s}\left(\hat{\theta}_{n_s}\right),\mathcal{Y}_{n_s}\right),c_u\left(\mu_{Y,n_s}\left(\tilde{\theta},P_{n_s}\right),\widehat{\Sigma}_{Y,n_s}\left(\hat{\theta}_{n_s}\right),\mathcal{Y}_{n_s}\right)\right]\right\}1\left\{\hat{\theta}_{n_s}=\tilde{\theta}\right\}\right]}{E_{P_{n_s}}\left[1\left\{\hat{\theta}_{n_s}=\tilde{\theta}\right\}\right]}\\ &\to\frac{E\left[1\left\{Y^*\left(\hat{\theta}\right)\in\left[c_l\left(\mu_Y^*\left(\tilde{\theta}\right),\Sigma_Y^*\left(\hat{\theta}\right),\mathcal{Y}^*\right),c_u\left(\mu_Y^*\left(\tilde{\theta}\right),\Sigma_Y^*\left(\hat{\theta}\right),\mathcal{Y}^*\right)\right]\right\}1\left\{\hat{\theta}=\tilde{\theta}\right\}\right]}{E\left[1\left\{\hat{\theta}=\tilde{\theta}\right\}\right]}=1-\alpha,\end{aligned}$$

as we wanted to show, so (18) follows by Lemma 5 of AKM.

Since this result again holds for all $\tilde{\theta}\in\Theta$, (19) follows immediately by the same argument as in the proof of Proposition 3. □

[19] Note that when $\hat{\theta}=\tilde{\theta}$, $\mathcal{Y}^*$ is either equal to the real line or contains at least one interval with a continuously distributed endpoint. Hence, the almost-everywhere continuity established in Lemma 5 is sufficient for us to apply the continuous mapping theorem.

AKM also introduces the analogous $1-\alpha$ level hybrid confidence interval, CS_U^H, that modifies the conditioning set in its construction to condition on both $\hat{\theta}$ and the event that $\mu_Y(\hat{\theta}) \in CS_P^\beta$ for some $0 \leq \beta \leq \alpha$. Again, the feasible versions of these intervals $CS_{U,n}^H$ are defined identically to CS_U after replacing $\hat{\theta}$ by $\hat{\theta}_n$, Y by Y_n, Σ by $\hat{\Sigma}_n$, and $Z_{\tilde{\theta}}$ by $Z_{\tilde{\theta},n}$. These intervals again have asymptotically correct unconditional coverage.

Proposition 7

Under Assumptions 1-4,

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}_n} \left| Pr_P\left\{ \mu_{Y,n}\left(\hat{\theta}_n;P\right) \in CS_{U,n}^H | \hat{\theta}_n = \tilde{\theta}, \mu_{Y,n}\left(\hat{\theta}_n;P\right) \in CS_{P,n}^\beta \right\} - \frac{1-\alpha}{1-\beta} \right| = 0,$$

for all $\tilde{\theta} \in \Theta$,

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}_n} Pr_P\left\{ \mu_{Y,n}\left(\hat{\theta}_n;P\right) \in CS_{U,n}^H \right\} \geq 1-\alpha,$$

and

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_n} Pr_P\left\{ \mu_{Y,n}\left(\hat{\theta}_n;P\right) \in CS_{U,n}^H \right\} \leq \frac{1-\alpha}{1-\beta} \leq 1-\alpha+\beta.$$

Proof: Follows by the same argument as the proof of Proposition 5, using Lemma 5 rather than Lemma 3. □

To state the following lemma, let

$$\left(c_l\left(\mu, \Sigma_Y(\theta), \mathcal{Y}^K\right), c_u\left(\mu, \Sigma_Y(\theta), \mathcal{Y}^K\right) \right) \tag{20}$$

solve

$$Pr\{\zeta \in [c_l, c_u]\} = 1-\alpha$$

$$E[\zeta 1\{\zeta \in [c_l, c_u]\}] = (1-\alpha)E[\zeta]$$

for ζ as in Lemma 3.

Lemma 5

The function (20) is continuous in $\left(\mu, \Sigma_Y(\theta), \mathcal{Y}^K\right)$ *for Lebesgue almost-every* $\left\{\ell^k, u^k\right\}_{k=1}^K$ *on the set*

$$\left\{ \begin{array}{c} (\mu, \Sigma_Y(\theta)) \in \mathbb{R}^2, \ell^1 \in [-\infty, \infty), \\ \left\{\ell^k\right\}_{k=2}^K \in \mathbb{R}^{K-1}, \left\{u^k\right\}_{k=1}^{K-1} \in \mathbb{R}^{K-1}, u^K \in (-\infty, \infty] \end{array} : \Sigma_Y(\theta) > 0, \sum_k \left|u^k - \ell^k\right| > 0, u^k \geq \ell^k \geq u^{k-1} \text{ for all } k \right\}.$$

Moreover, if we fix any $(\mu, \Sigma_Y(\theta))$ *in this set, and fix all but one element of* $\left\{\ell^k, u^k\right\}_{k=1}^K$, *(20) is almost-everywhere continuous in the remaining element.*

Proof: Note that

$$Pr\{\zeta\in[c_l,c_u]\}=\frac{\sum_k 1\left\{u^k\geq c_l,c_u\geq \ell^k\right\}\left(F_N\left(\frac{u^k\wedge c_u-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-F_N\left(\frac{\ell^k\vee c_l-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)\right)}{\sum_k\left(F_N\left(\frac{u^k-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-F_N\left(\frac{\ell^k-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)\right)}$$

while

$$E[\zeta 1\{\zeta\in[c_l,c_u]\}]=E[\zeta|\zeta\in[c_l,c_u]]Pr\{\zeta\in[c_l,c_u]\}$$

where

$$E[\zeta|\zeta\in[c_l,c_u]]=\mu+\sqrt{\Sigma_Y(\theta)}\frac{\sum_k 1\left\{u^k\geq c_l,c_u\geq \ell^k\right\}\left(f_N\left(\frac{\ell^k\vee c_l-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-f_N\left(\frac{u^k\wedge c_u-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)\right)}{\sum_k 1\{u^k\geq c_l,c_u\geq \ell^k\}\left(F_N\left(\frac{u^k\wedge c_u-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-F_N\left(\frac{\ell^k\vee c_l-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)\right)}.$$

Thus,

$$\begin{aligned}E[\zeta 1\{\zeta\in[c_l,c_u]\}]=\mu&\frac{\sum_k 1\left\{u^k\geq c_l,c_u\geq \ell^k\right\}\left(F_N\left(\frac{u^k\wedge c_u-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-F_N\left(\frac{\ell^k\vee c_l-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)\right)}{\sum_k\left(F_N\left(\frac{u^k-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-F_N\left(\frac{\ell^k-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)\right)}\\ &+\sqrt{\Sigma_Y(\theta)}\frac{\sum_k 1\left\{u^k\geq c_l,c_u\geq \ell^k\right\}\left(f_N\left(\frac{\ell^k\vee c_l-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-f_N\left(\frac{u^k\wedge c_u-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)\right)}{\sum_k\left(F_N\left(\frac{u^k-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-F_N\left(\frac{\ell^k-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)\right)}\end{aligned}$$

and

$$E[\zeta]=\mu+\sqrt{\Sigma_Y(\theta)}\frac{\sum_k\left(f_N\left(\frac{\ell^k-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-f_N\left(\frac{u^k-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)\right)}{\sum_k\left(F_N\left(\frac{u^k-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-F_N\left(\frac{\ell^k-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)\right)}.$$

Using analogous reasoning to that in the proof of Lemma 10 in AKM, we can write (20) as the solution to

$$g\left(c;\mu,\sqrt{\Sigma_Y(\theta)},\mathcal{Y}^K\right)=0 \tag{21}$$

for

$$g\left(c;\mu,\sqrt{\Sigma_Y(\theta)},\mathcal{Y}^K\right)=$$

$$\begin{pmatrix} \sum_k 1\{u^k \geq c_l, c_u \geq \ell^k\}\left(F_N\left(\frac{u^k \wedge c_u - \mu}{\sqrt{\Sigma_Y(\theta)}}\right) - F_N\left(\frac{\ell^k \vee c_l - \mu}{\sqrt{\Sigma_Y(\theta)}}\right) - (1-\alpha)\left(F_N\left(\frac{u^k - \mu}{\sqrt{\Sigma_Y(\theta)}}\right) - F_N\left(\frac{\ell^k - \mu}{\sqrt{\Sigma_Y(\theta)}}\right)\right)\right) \\ \sum_k 1\{u^k \geq c_l, c_u \geq \ell^k\}\left(f_N\left(\frac{\ell^k \vee c_l - \mu}{\sqrt{\Sigma_Y(\theta)}}\right) - f_N\left(\frac{u^k \wedge c_l - \mu}{\sqrt{\Sigma_Y(\theta)}}\right) - (1-\alpha)\left(f_N\left(\frac{\ell^k - \mu}{\sqrt{\Sigma_Y(\theta)}}\right) - f_N\left(\frac{u^k - \mu}{\sqrt{\Sigma_Y(\theta)}}\right)\right)\right) \end{pmatrix}.$$

Note that by construction

$$g\left(c;\mu,\sqrt{\Sigma_Y(\theta)},\mathcal{Y}^K\right) = g\left(c-\mu;0,\sqrt{\Sigma_Y(\theta)},\mathcal{Y}^K - \mu\right),$$

which implies that

$$\left(c_l\left(\mu,\Sigma_Y(\theta),\mathcal{Y}^K\right),c_u\left(\mu,\Sigma_Y(\theta),\mathcal{Y}^K\right)\right) = \left(\mu + c_l\left(0,\Sigma_Y(\theta),\mathcal{Y}^K - \mu\right),\mu + c_u\left(0,\Sigma_Y(\theta),\mathcal{Y}^K - \mu\right)\right)$$

so to prove continuity it suffices to consider the case with $\mu = 0$.

Next, note that $g\left(c;0,\sqrt{\Sigma_Y(\theta)},\mathcal{Y}^K\right)$ is almost everywhere differentiable with respect to (c_l,c_u), with derivative

$$\begin{pmatrix} \sum_k 1\left\{u^k > c_l > \ell^k\right\} \frac{-1}{\sqrt{\Sigma_Y(\theta)}} f_N\left(\frac{c_l}{\sqrt{\Sigma_Y(\theta)}}\right) & \sum_k 1\left\{u^k > c_u > \ell^k\right\} \frac{1}{\sqrt{\Sigma_Y(\theta)}} f_N\left(\frac{c_u}{\sqrt{\Sigma_Y(\theta)}}\right) \\ \sum_k 1\left\{u^k > c_l > \ell^k\right\} \frac{-c_l}{\Sigma_Y(\theta)} f_N\left(\frac{c_l}{\sqrt{\Sigma_Y(\theta)}}\right) & \sum_k 1\left\{u^k > c_u > \ell^k\right\} \frac{c_u}{\Sigma_Y(\theta)} f_N\left(\frac{c_u}{\sqrt{\Sigma_Y(\theta)}}\right) \end{pmatrix},$$

though it is non-differentiable if $c_u \in \left\{u^k,\ell^k\right\}$ or $c_l \in \left\{u^k,\ell^k\right\}$ for some k.

Note, however, that if we fix all but one element of $\left\{\ell^k,u^k\right\}_{k=1}^K$ and change the remaining element, there exists a solution c to (21) with $c_u \in (\ell^j,u^j)$ and $c_l \in \left(\ell^k,u^k\right)$ for some j,k Lebesgue almost-everywhere by arguments along the same lines as in the proof of Lemma 10 of AKM. Likewise, the set of values such that there exists a solution c to (21) with $c_l = c_u$ has Lebesgue measure zero. The implicit function theorem thus implies that (20) is almost-everywhere continuously differentiable in the element we have selected. Since we can repeat this argument for each element of $\left\{\ell^k,u^k\right\}_{k=1}^K$, we obtain that (20) is elementwise continuously differentiable in $\left\{\ell^k,u^k\right\}_{k=1}^K$ Lebesgue almost everywhere. Moreover, as in the proof of Lemma 10 of AKM, the form of (20) implies that the same remains true if we take $\ell^1 \to -\infty$ or $u^K \to \infty$. $\square$

C Additional Results for Tipping Point Simulations

We begin by presenting results analogous to those presented in Tables 1 and 2 of the main text for the conditional and hybrid confidence intervals based upon the uniformly most accurate unbiased approach. More specifically, Table 5 reports the unconditional coverage probability for the confidence intervals CS_U and CS_U^H while Table 6 compares the lengths of CS_U and CS_U^H to CS_P^α. The values in these two tables can be seen to be quite similar to the values corresponding to the confidence intervals CS and CS^H in Tables 1 and 2 of the main text.

Table 5: Unconditional Coverage Probability

DGP	CS_U	CS_U^H
Chicago Data Calibration		
(i)	0.95	0.949
(ii)	0.95	0.955
(iii)	0.946	0.951
Los Angeles Data Calibration		
(i)	0.948	0.948
(ii)	0.952	0.956
(iii)	0.951	0.954

This table reports the unconditional coverage probability of $\mu_Y(\hat{\theta})$ for the conditionally valid uniformly most accurate unbiased confidence interval (CS_U) and the hybrid confidence interval based upon the uniformly most accurate unbiased conditional confidence interval (CS_U^H), both evaluated at the nominal coverage level of 95%. In the Chicago (Los Angeles) data calibrations, the covariance matrix Σ is set equal to a consistent estimate from the Chicago (Los Angeles) Card et al. (2008) data. The column "DGP" refers to the specification of the nuisance function $\Sigma_{Cg}(\cdot)$, which along with other parameters, determines the value of the mean vector μ (see Appendix A.1 of the main text for details). The function $\Sigma_{Cg}(\cdot)$ is set equal to the value it takes when there is no coefficient change in DGP (i), the value it takes when there is a single large coefficient change in DGP (ii) and its data-calibrated value in DGP (iii). For DGP (ii) the true threshold location is set to equal the estimate from the Card et al. (2008) data. All other parameters that determine μ are set equal to consistent estimates from the Card et al. (2008) data.

Tables 7 and 8 provide the ratios of the 5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} quantiles of the lengths of CS, CS_U, CS^H and CS_U^H relative to the corresponding length quantiles of CS_P^α for the tipping point data-calibrated designs described in Section 6 of the main text. Looking at the upper quantiles in Table 7, we can see that the conditional confidence intervals CS and CS_U can become very wide in the absence of a clear break. Conversely, as seen in Table 8, the hybrid intervals CS^H and CS_U^H dominate CS_P^α across all quantiles and simulation designs we examined.

Table 9 reports the same quantiles of the studentized absolute errors of $\hat{\mu}_{\frac{1}{2}}$, $\hat{\mu}_{\frac{1}{2}}^H$ and $Y(\hat{\theta})$. The main features of this table are similar to those of Table 7: the unconditional estimator $\hat{\mu}_{\frac{1}{2}}$ can exhibit very large absolute errors while the hybrid estimator $\hat{\mu}_{\frac{1}{2}}^H$ does not exhibit such extreme values. In addition, note that the hybrid estimator $\hat{\mu}_{\frac{1}{2}}^H$ not only exhibits minimal bias, in contrast to the standard estimator $Y(\hat{\theta})$, but also exhibits lower studentized absolute errors across most quantiles and designs considered.

Table 6: Length of Confidence Sets Relative to CS_P^α in Tipping Point Simulations

	Median Length Relative to CS_P^α		Probability Longer than CS_P^α	
DGP	CS_U	CS_U^H	CS_U	CS_U^H
		Chicago Data Calibration		
(i)	1.38	0.94	0.89	0
(ii)	0.72	0.74	0	0
(iii)	0.93	0.87	0.44	0
		Los Angeles Data Calibration		
(i)	1.29	0.85	0.62	0
(ii)	0.68	0.69	0	0
(iii)	0.70	0.72	0.19	0

This table reports the median length of the conditionally valid uniformly most accurate unbiased confidence interval (CS_U) and the hybrid confidence interval based upon the uniformly most accurate unbiased conditional confidence interval (CS_U^H), divided by the median length of the projection confidence interval (CS_P^α), as well as the frequency with which CS_U and CS_U^H is longer than CS_P^α. In the Chicago (Los Angeles) data calibrations, the covariance matrix Σ is set equal to a consistent estimate from the Chicago (Los Angeles) Card et al. (2008) data. The column "DGP" refers to the specification of the nuisance function $\Sigma_{Cg}(\cdot)$, which along with other parameters, determines the value of the mean vector μ (see Appendix A.1 of the main text for details). The function $\Sigma_{Cg}(\cdot)$ is set equal to the value it takes when there is no coefficient change in DGP (i), the value it takes when there is a single large coefficient change in DGP (ii) and its data-calibrated value in DGP (iii). For DGP (ii) the true threshold location is set to equal the estimate from the Card et al. (2008) data. All other parameters that determine μ are set equal to consistent estimates from the Card et al. (2008) data.

Table 7: Ratios of Length Quantiles Relative to CS_P^α

	CS Quantile					CS_U Quantile				
DGP	5^{th}	25^{th}	50^{th}	75^{th}	95^{th}	5^{th}	25^{th}	50^{th}	75^{th}	95^{th}
	Chicago Data Calibration									
(i)	0.88	1.13	1.33	1.54	1.87	0.92	1.20	1.38	1.58	1.89
(ii)	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.74
(iii)	0.74	0.74	0.82	1.22	3.30	0.74	0.76	0.93	1.45	3.65
	Los Angeles Data Calibration									
(i)	0.92	1.27	1.26	0.99	0.76	0.94	1.31	1.29	1.00	0.77
(ii)	0.68	0.68	0.68	0.68	0.68	0.67	0.68	0.68	0.68	0.69
(iii)	0.68	0.68	0.68	0.79	2.12	0.68	0.68	0.70	0.89	2.32

This table reports the 5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} quantiles of the length of the conditionally valid equal-tailed confidence interval (CS) and conditionally valid uniformly most accurate unbiased confidence interval (CS_U), divided by the corresponding length quantiles of the projection confidence interval (CS_P^α). In the Chicago (Los Angeles) data calibrations, the covariance matrix Σ is set equal to a consistent estimate from the Chicago (Los Angeles) Card et al. (2008) data. The column "DGP" refers to the specification of the nuisance function $\Sigma_{Cg}(\cdot)$, which along with other parameters, determines the value of the mean vector μ (see Appendix A.1 of the main text for details). The function $\Sigma_{Cg}(\cdot)$ is set equal to the value it takes when there is no coefficient change in DGP (i), the value it takes when there is a single large coefficient change in DGP (ii) and its data-calibrated value in DGP (iii). For DGP (ii) the true threshold location is set to equal the estimate from the Card et al. (2008) data. All other parameters that determine μ are set equal to consistent estimates from the Card et al. (2008) data.

Table 8: Ratios of Length Quantiles Relative to CS_P^α

	CS^H Quantile					CS_U^H Quantile				
DGP	5^{th}	25^{th}	50^{th}	75^{th}	95^{th}	5^{th}	25^{th}	50^{th}	75^{th}	95^{th}
				Chicago Data Calibration						
(i)	0.69	0.91	0.94	0.93	0.96	0.60	0.90	0.94	0.93	0.96
(ii)	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.75
(iii)	0.75	0.75	0.82	0.93	0.97	0.76	0.78	0.87	0.94	0.97
				Los Angeles Data Calibration						
(i)	0.73	0.91	0.86	0.82	0.76	0.65	0.91	0.85	0.82	0.76
(ii)	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.70
(iii)	0.69	0.69	0.70	0.79	0.91	0.68	0.69	0.72	0.84	0.92

This table reports the 5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} quantiles of the length of the hybrid confidence interval based upon the equal-tailed conditional confidence interval (CS^H) and the hybrid confidence interval based upon the uniformly most accurate unbiased conditional confidence interval (CS_U^H), divided by the corresponding length quantiles of the projection confidence interval (CS_P^α). In the Chicago (Los Angeles) data calibrations, the covariance matrix Σ is set equal to a consistent estimate from the Chicago (Los Angeles) Card et al. (2008) data. The column "DGP" refers to the specification of the nuisance function $\Sigma_{Cg}(\cdot)$, which along with other parameters, determines the value of the mean vector μ (see Appendix A.1 of the main text for details). The function $\Sigma_{Cg}(\cdot)$ is set equal to the value it takes when there is no coefficient change in DGP (i), the value it takes when there is a single large coefficient change in DGP (ii) and its data-calibrated value in DGP (iii). For DGP (ii) the true threshold location is set to equal the estimate from the Card et al. (2008) data. All other parameters that determine μ are set equal to consistent estimates from the Card et al. (2008) data.

Table 9: Quantiles of $\left|\hat{\mu} - \mu_Y(\hat{\theta})\right| / \sqrt{\Sigma_Y(\hat{\theta})}$

	$\hat{\mu}_{\frac{1}{2}}$ Quantile					$\hat{\mu}^H_{\frac{1}{2}}$ Quantile					$Y(\hat{\theta})$ Quantile				
DGP	5^{th}	25^{th}	50^{th}	75^{th}	95^{th}	5^{th}	25^{th}	50^{th}	75^{th}	95^{th}	5^{th}	25^{th}	50^{th}	75^{th}	95^{th}
Chicago Data Calibration															
(i)	0.15	0.74	1.51	2.65	6.38	0.15	0.71	1.38	2.02	2.63	0.81	1.16	1.52	1.95	2.70
(ii)	0.06	0.32	0.66	1.14	1.95	0.06	0.32	0.66	1.14	1.95	0.06	0.32	0.66	1.14	1.95
(iii)	0.08	0.38	0.83	1.50	4.81	0.08	0.38	0.83	1.48	2.94	0.07	0.34	0.71	1.19	2.05
Los Angeles Data Calibration															
(i)	0.13	0.67	1.38	2.32	5.25	0.13	0.64	1.29	1.93	2.60	1.07	1.45	1.80	2.20	2.89
(ii)	0.07	0.32	0.67	1.14	1.93	0.07	0.32	0.67	1.14	1.93	0.07	0.32	0.67	1.14	1.93
(iii)	0.07	0.35	0.74	1.31	2.56	0.07	0.35	0.74	1.30	2.46	0.06	0.33	0.68	1.17	2.00

This table reports the the 5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} quantiles of the studentized absolute estimation error for the conditionally median-unbiased estimator ($\hat{\mu}_{\frac{1}{2}}$), the hybrid estimator ($\hat{\mu}^H_{\frac{1}{2}}$) and the conventional estimator ($Y(\hat{\theta})$). In the Chicago (Los Angeles) data calibrations, the covariance matrix Σ is set equal to a consistent estimate from the Chicago (Los Angeles) Card et al. (2008) data. The column "DGP" refers to the specification of the nuisance function $\Sigma_{Cg}(\cdot)$, which along with other parameters, determines the value of the mean vector μ (see Appendix A.1 of the main text for details). The function $\Sigma_{Cg}(\cdot)$ is set equal to the value it takes when there is no coefficient change in DGP (i), the value it takes when there is a single large coefficient change in DGP (ii) and its data-calibrated value in DGP (iii). For DGP (ii) the true threshold location is set to equal the estimate from the Card et al. (2008) data. All other parameters that determine μ are set equal to consistent estimates from the Card et al. (2008) data.

C.1 Additional Results for Split-Sample Approaches

Table 10 provides the ratios of the 5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} quantiles of the length of our newly proposed equal-tailed split-sample confidence interval CS_{SS}^{A} relative to the corresponding length quantiles of the conventional split-sample confidence interval CS_{SS} for each of the tipping point data-calibrated designs described in Section 6 of the main text. Since every entry in this table is less than one, we can see that the dominance result illustrated in Table 4 of the main text is further reinforced: the length quantiles of CS_{SS}^{A} are shorter than those of CS_{SS} across all quantiles and simulation designs considered. Table 11 reports the same quantiles of the studentized absolute errors of our newly proposed split-sample estimator $\hat{\mu}_{SS,\frac{1}{2}}^{A}$ and those of the conventional split-sample estimator $Y^2(\hat{\theta}^1)$. Though both of these estimators are median-unbiased for $\mu_Y(\hat{\theta}^1)$, $\hat{\mu}_{SS,\frac{1}{2}}^{A}$ dominates $Y^2(\hat{\theta}^1)$ in terms of studentized absolute errors across all quantiles and simulation designs considered.

Table 10: Ratios of Length Quantiles of CS_{SS}^{A} Relative to CS_{SS}

	Quantile				
DGP	5^{th}	25^{th}	50^{th}	75^{th}	95^{th}
Chicago Data Calibration					
(i)	0.69	0.79	0.83	0.84	0.87
(ii)	0.57	0.58	0.58	0.58	0.58
(iii)	0.59	0.59	0.64	0.73	0.86
Los Angeles Data Calibration					
(i)	0.74	0.85	0.78	0.68	0.57
(ii)	0.57	0.58	0.58	0.58	0.58
(iii)	0.57	0.58	0.59	0.66	0.81

This table reports the the 5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} quantiles of the length of the alternative split-sample confidence interval (CS_{SS}^{A}), divided by the corresponding length quantiles of the conventional split-sample confidence interval (CS_{SS}). In the Chicago (Los Angeles) data calibrations, the covariance matrix Σ is set equal to a consistent estimate from the Chicago (Los Angeles) Card et al. (2008) data. The column "DGP" refers to the specification of the nuisance function $\Sigma_{Cg}(\cdot)$, which along with other parameters, determines the value of the mean vector μ (see Appendix A.1 of the main text for details). The function $\Sigma_{Cg}(\cdot)$ is set equal to the value it takes when there is no coefficient change in DGP (i), the value it takes when there is a single large coefficient change in DGP (ii) and its data-calibrated value in DGP (iii). For DGP (ii) the true threshold location is set to equal the estimate from the Card et al. (2008) data. All other parameters that determine μ are set equal to consistent estimates from the Card et al. (2008) data.

Table 11: Quantiles of $\left|\hat{\mu} - \mu_Y(\hat{\theta}^1)\right| / \sqrt{\Sigma_Y(\hat{\theta})^1}$

	$\hat{\mu}^A_{SS,\frac{1}{2}}$ Quantile					$Y^2(\hat{\theta}^1)$ Quantile				
DGP	5^{th}	25^{th}	50^{th}	75^{th}	95^{th}	5^{th}	25^{th}	50^{th}	75^{th}	95^{th}
Chicago Data Calibration										
(i)	0.05	0.27	0.57	0.95	1.61	0.06	0.31	0.67	1.15	1.97
(ii)	0.04	0.18	0.38	0.65	1.13	0.06	0.31	0.66	1.14	1.96
(iii)	0.04	0.21	0.44	0.77	1.38	0.07	0.32	0.67	1.15	2.00
Los Angeles Data Calibration										
(i)	0.05	0.25	0.55	0.93	1.56	0.07	0.32	0.69	1.16	1.96
(ii)	0.04	0.18	0.39	0.66	1.13	0.06	0.31	0.67	1.15	1.96
(iii)	0.04	0.20	0.42	0.71	1.25	0.06	0.32	0.68	1.16	1.98

This table reports the the 5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} quantiles of the studentized absolute estimation error of the median-unbiased alternative split-sample estimator ($\hat{\mu}^A_{\frac{1}{2},SS}$) and of the conventional split-sample estimator ($Y^2(\hat{\theta}^1)$). In the Chicago (Los Angeles) data calibrations, the covariance matrix Σ is set equal to a consistent estimate from the Chicago (Los Angeles) Card et al. (2008) data. The column "DGP" refers to the specification of the nuisance function $\Sigma_{Cg}(\cdot)$, which along with other parameters, determines the value of the mean vector μ (see Appendix A.1 of the main text for details). The function $\Sigma_{Cg}(\cdot)$ is set equal to the value it takes when there is no coefficient change in DGP (i), the value it takes when there is a single large coefficient change in DGP (ii) and its data-calibrated value in DGP (iii). For DGP (ii) the true threshold location is set to equal the estimate from the Card et al. (2008) data. All other parameters that determine μ are set equal to consistent estimates from the Card et al. (2008) data.