Supplement to the paper

Inference on Winners

Isaiah Andrews Toru Kitagawa Adam McCloskey

September 16, 2019

This supplement contains proofs and additional results for the paper "Inference on Winners." Section A collects proofs for results stated in the main text. Section B contains additional details and derivations for the EWM example introduced in Section 3 of the paper. Section C constructs procedures that dominate conventional sample splitting as discussed in Section 4.3 of the paper. Section D translates our finite-sample results for the normal model to uniform asymptotic results over large classes of data generating processes. Section E reports additional simulation results for the stylized example of Section 2 of the paper. Finally, Section F reports additional simulations results for the EWM simulations discussed in Section 6 of the paper.

A Proofs

Proof of Proposition 1 For ease of reference, let us abbreviate $(Y(\tilde{\theta}),\mu_Y(\tilde{\theta}),Z_{\tilde{\theta}})$ by $(\tilde{Y},\tilde{\mu}_Y,\tilde{Z})$. Let $Y(-\tilde{\theta})$ collect the elements of Y other than $Y(\tilde{\theta})$ and define $\mu_Y(-\theta)$ analagously. Let

$$Y^*=Y(-\tilde{\theta})-Cov\left(Y(-\tilde{\theta}),\begin{pmatrix}\tilde{Y}\\X\end{pmatrix}\right)Var\left(\begin{pmatrix}\tilde{Y}\\X\end{pmatrix}\right)^+\begin{pmatrix}\tilde{Y}\\X\end{pmatrix},$$

$$\mu_Y^*=\mu_Y(-\tilde{\theta})-Cov\left(Y(-\tilde{\theta}),\begin{pmatrix}\tilde{Y}\\X\end{pmatrix}\right)Var\left(\begin{pmatrix}\tilde{Y}\\X\end{pmatrix}\right)^+\begin{pmatrix}\tilde{\mu}_Y\\\mu_X\end{pmatrix},$$

and

$$\tilde{\mu}_Z=\mu_X-\left(\Sigma_{XY}(\cdot,\tilde{\theta})/\Sigma_Y(\tilde{\theta})\right)\mu_Y.$$

Here we use A^+ to denote the Moore-Penrose pseudoinverse of a matrix A. Note that $(\tilde{Z},\tilde{Y},Y^*)$ is a one-to-one transformation of (X,Y), and thus that observing $(\tilde{Z},\tilde{Y},Y^*)$ is equivalent to observing (X,Y). Likewise, $(\tilde{\mu}_Z,\tilde{\mu}_Y,\mu_Y^*)$ is a one-to-one linear transformation of (μ_X,μ_Y), and if the set of possible values for the latter contains an open set, that for the former does as well (relative to the appropriate linear subspace).

Note, next, that since $(\tilde{Z},\tilde{Y},Y^*)$ is a linear transformation of (X,Y), $(\tilde{Z},\tilde{Y},Y^*)$ is jointly normal (with a potentially degenerate distribution). Note next that $(\tilde{Z},\tilde{Y},Y^*)$ are mutually uncorrelated, and thus independent. That $\tilde{Z}$ and $\tilde{Y}$ are uncorrelated is straightforward to verify. To show that Y^* is likewise uncorrelated with the other elements, note that we can write $Cov\left(Y^*,(\tilde{Y},X')'\right)$ as

$$Cov\left(Y(-\tilde{\theta}),\begin{pmatrix}\tilde{Y}\\ X\end{pmatrix}\right)-Cov\left(Y(-\tilde{\theta}),\begin{pmatrix}\tilde{Y}\\ X\end{pmatrix}\right)Var\left(\begin{pmatrix}\tilde{Y}\\ X\end{pmatrix}\right)^+Var\left(\begin{pmatrix}\tilde{Y}\\ X\end{pmatrix}\right).$$

For $V\Lambda V'$ an eigendecomposition of $Var\left((\tilde{Y},X')'\right)$ (so $VV'=I$), note that we can write

$$Var\left(\begin{pmatrix}\tilde{Y}\\ X\end{pmatrix}\right)^+Var\left(\begin{pmatrix}\tilde{Y}\\ X\end{pmatrix}\right)=VDV'$$

for D a diagonal matrix with ones in the entries corresponding to the nonzero entries of Λ and zeros everywhere else. For any column v of V corresponding to a zero entry of D, $v'Var\left(\left(\tilde{Y},X'\right)'\right)v=0$, so the Cauchy-Schwarz inequality implies that

$$Cov\left(Y\left(-\tilde{\theta}\right),\begin{pmatrix}\tilde{Y}\\ X\end{pmatrix}\right)v=0.$$

Thus,

$$Cov\left(Y\left(-\tilde{\theta}\right),\begin{pmatrix}\tilde{Y}\\ X\end{pmatrix}\right)VDV'=Cov\left(Y\left(-\tilde{\theta}\right),\begin{pmatrix}\tilde{Y}\\ X\end{pmatrix}\right)VV'=Cov\left(Y\left(-\tilde{\theta}\right),\begin{pmatrix}\tilde{Y}\\ X\end{pmatrix}\right),$$

so Y^* is uncorrelated with $\left(\tilde{Y},X'\right)'$.

Using independence, the joint density of $(\tilde{Z},\tilde{Y},Y^*)$ absent truncation is given by

$$f_{N,\tilde{Z}}(\tilde{z};\tilde{\mu}_Z)f_{N,\tilde{Y}}(\tilde{y};\tilde{\mu}_Y)f_{N,Y^*}(\tilde{y}^*;\mu_Y^*)$$

for f_N normal densities with respect to potentially degenerate base measures:

$$f_{N,\tilde{Z}}(\tilde{z};\tilde{\mu}_Z)=\tilde{\det}(2\pi\Sigma_{\tilde{Z}})^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(\tilde{z}-\tilde{\mu}_Z)'\Sigma_{\tilde{Z}}^{+}(\tilde{z}-\tilde{\mu}_Z)\right)$$

$$f_{N,\tilde{Y}}(\tilde{y};\tilde{\mu}_Y)=(2\pi\Sigma_{\tilde{Y}})^{-\frac{1}{2}}\exp\left(-\frac{(\tilde{y}-\tilde{\mu}_Y)^2}{2\Sigma_{\tilde{Y}}}\right)$$

$$f_{N,Y^*}(y^*;\mu_Y^*)=\tilde{\det}(2\pi\Sigma_{Y^*})^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(y^*-\tilde{\mu}_Y^*)'\Sigma_{Y^*}^{+}(y^*-\mu_Y^*)\right),$$

where $\tilde{\det}(A)$ denotes the pseudodeterminant of a matrix A, $\Sigma_{\tilde{Z}}=Var(\tilde{Z})$, $\Sigma_{\tilde{Y}}=\Sigma_Y(\tilde{\theta})$, and $\Sigma_{Y^*}=Var(Y^*)$.

The event $\left\{X\in\mathcal{X}(\tilde{\theta},\tilde{\gamma})\right\}$ depends only on $(\tilde{Z},\tilde{Y})$ since it can be expressed as

$$\left\{\left(\tilde{Z}+\frac{\Sigma_{XY}(\cdot,\tilde{\theta})}{\Sigma_Y(\tilde{\theta})}\tilde{Y}\right)\in\mathcal{X}(\tilde{\theta},\tilde{\gamma})\right\},$$

so conditional on this event Y^* remains independent of $(\tilde{Z},\tilde{Y})$. In particular, we can write the joint density conditional on $\left\{X\in\mathcal{X}(\tilde{\theta},\tilde{\gamma})\right\}$ as

$$\frac{1\left\{\left(\tilde{z}+\Sigma_{XY}(\cdot,\tilde{\theta})\Sigma_Y(\tilde{\theta})^{-1}\tilde{y}\right)\in\mathcal{X}(\tilde{\theta},\tilde{\gamma})\right\}}{Pr_{\tilde{\mu}_Z,\tilde{\mu}_Y}\left\{X\in\mathcal{X}(\tilde{\theta},\tilde{\gamma})\right\}}f_{N,\tilde{Z}}(\tilde{z};\tilde{\mu}_Z)f_{N,\tilde{Y}}(\tilde{y};\tilde{\mu}_Y)f_{N,Y^*}(\tilde{y}^*;\mu_Y^*). \tag{20}$$

The density (20) has the same structure as (5.5.14) of Pfanzagl (1994), and satisfies properties (5.5.1)-(5.5.3) of Pfanzagl (1994) as well. Part 1 of the proposition then follows immediately from Theorem 5.5.9 of Pfanzagl (1994). Part 2 of the proposition follows by using Theorem 5.5.9 of Pfanzagl (1994) to verify the conditions of Theorem 5.5.15 of Pfanzagl (1994). □

Proof of Proposition 2 In the proof of Proposition 1, we showed that the joint density of $(\tilde{Z},\tilde{Y},Y^*)$ (defined in that proof) has the exponential family structure assumed in equation 4.10 of Lehmann and Romano (2005). Moreover, Assumption 1 implies that the parameter space for (μ_X,μ_Y) is convex and is not contained in any proper linear subspace. Thus, the parameter space for $(\tilde{\mu}_Z,\tilde{\mu}_Y,\mu_Y^*)$ inherits the same property, and satisfies the conditions of Theorem 4.4.1 of Lehmann and Romano (2005). The result follows immediately. □

Proof of Proposition 3 Let us number the elements of Θ as $\left\{\theta_1,\theta_2,...,\theta_{|\Theta|}\right\}$, where $X(\theta_1)$ is the first element of X, $X(\theta_2)$ is the second element, and so on. Let us further assume without loss of generality that $\tilde{\theta}=\theta_1$. Note that the conditioning event $\{\max_{\theta\in\Theta}X(\theta)=X(\theta_1)\}$ is equivalent to $\{MX\geq 0\}$, where

$$M\equiv\begin{pmatrix} 1 & -1 & 0 & 0 & ... & 0 \\ 1 & 0 & -1 & 0 & ... & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & ... & -1 \end{pmatrix}$$

is a $(|\Theta|-1)\times|\Theta|$ matrix and the inequality is taken element-wise. Let $A=\begin{bmatrix}-M & 0_{(|\Theta|-1)\times|\Theta|}\end{bmatrix}$, where $0_{(|\Theta|-1)\times|\Theta|}$ denotes the $(|\Theta|-1)\times|\Theta|$ matrix of zeros. Let $W=(X',Y')'$ and note that we can re-write the event of interest as $\{W:AW\leq 0\}$ and that we are interested in inference on $\eta'\mu$ for η the $2|\Theta|\times 1$ vector with one in the $(|\Theta|+1)$st entry and zeros everywhere else. Define

$$Z^*_{\tilde{\theta}}=W-cY(\tilde{\theta}),$$

for $c=Cov(W,Y(\tilde{\theta}))/\Sigma_Y(\tilde{\theta})$, noting that the definition of $Z_{\tilde{\theta}}$ in (11) corresponds to extracting the elements of $Z^*_{\tilde{\theta}}$ corresponding to X. By Lemma 5.1 of Lee et al. (2016),

$$\{W:AW\leq 0\}=\left\{W:\mathcal{L}(\tilde{\theta},Z^*_{\tilde{\theta}})\leq Y(\tilde{\theta})\leq\mathcal{U}(\tilde{\theta},Z^*_{\tilde{\theta}}),\mathcal{V}(\tilde{\theta},Z^*_{\tilde{\theta}})\geq 0\right\},$$

where for $(v)_j$ the jth element of a vector v,

$$\mathcal{L}(\tilde{\theta},z)=\max_{j:(Ac)_j<0}\frac{-(Az)_j}{(Ac)_j}$$

$$\mathcal{U}(\tilde{\theta},z)=\min_{j:(Ac)_j>0}\frac{-(Az)_j}{(Ac)_j}$$

$$\mathcal{V}(\tilde{\theta},z)=\min_{j:(Ac)_j=0}-(Az)_j.$$

Note, however, that

$$\left(AZ^*_{\tilde{\theta}}\right)_j=Z_{\tilde{\theta}}(\theta_j)-Z_{\tilde{\theta}}(\theta_1)$$

and

$$(Ac)_j = -\frac{\Sigma_{XY}(\theta_1,\theta_1) - \Sigma_{XY}(\theta_1,\theta_j)}{\Sigma_Y(\theta_1)}.$$

Hence, we can re-write

$$\frac{-(AZ^*_{\tilde{\theta}})_j}{(Ac)_j} = \frac{\Sigma_Y(\theta_1)(Z_{\tilde{\theta}}(\theta_j) - Z_{\tilde{\theta}}(\theta_1))}{\Sigma_{XY}(\theta_1,\theta_1) - \Sigma_{XY}(\theta_1,\theta_j)},$$

$$\mathcal{L}(\tilde{\theta},Z^*_{\tilde{\theta}}) = \max_{j:\Sigma_{XY}(\theta_1,\theta_1) > \Sigma_{XY}(\theta_1,\theta_j)} \frac{\Sigma_Y(\theta_1)(Z_{\tilde{\theta}}(\theta_j) - Z_{\tilde{\theta}}(\theta_1))}{\Sigma_{XY}(\theta_1,\theta_1) - \Sigma_{XY}(\theta_1,\theta_j)},$$

$$\mathcal{U}(\tilde{\theta},Z^*_{\tilde{\theta}}) = \min_{j:\Sigma_{XY}(\theta_1,\theta_1) < \Sigma_{XY}(\theta_1,\theta_j)} \frac{\Sigma_Y(\theta_1)(Z_{\tilde{\theta}}(\theta_j) - Z_{\tilde{\theta}}(\theta_1))}{\Sigma_{XY}(\theta_1,\theta_1) - \Sigma_{XY}(\theta_1,\theta_j)},$$

and

$$\mathcal{V}(\tilde{\theta},Z^*_{\tilde{\theta}}) = \min_{j:\Sigma_{XY}(\theta_1,\theta_1) = \Sigma_{XY}(\theta_1,\theta_j)} -(Z_{\tilde{\theta}}(\theta_j) - Z_{\tilde{\theta}}(\theta_1)).$$

Note, however, that these are functions of $Z_{\tilde{\theta}}$, as expected. The result follows. $\square$

Proof of Lemma 1 Recall that conditional on $Z_{\tilde{\theta}} = z_{\tilde{\theta}}$, $\hat{\theta} = \tilde{\theta}$ and $\hat{\gamma} = \tilde{\gamma}$ if and only if $Y(\tilde{\theta}) \in \mathcal{Y}(\tilde{\theta},\tilde{\gamma},z_{\tilde{\theta}})$. Hence, the assumption of the lemma implies that

$$Pr_{\mu_{Y,m}}\left\{Y(\tilde{\theta}) \in \mathcal{Y}(\tilde{\theta},\tilde{\gamma},Z_{\tilde{\theta}}) | Z_{\tilde{\theta}} = z_{\tilde{\theta},m}\right\} \to 1.$$

Note, next, that both the conventional and conditional confidence intervals are equivariant under shifts, in the sense that the conditional confidence interval for $\mu_Y(\tilde{\theta})$ based on observing $Y(\tilde{\theta})$ conditional on $Y(\tilde{\theta}) \in \mathcal{Y}(\tilde{\theta},\tilde{\gamma},Z_{\tilde{\theta}})$ is equal to the conditional confidence interval for $\mu_Y(\tilde{\theta})$ based on observing $Y(\tilde{\theta}) - \mu^*_Y(\tilde{\theta})$ conditional on $Y(\tilde{\theta}) - \mu^*_Y(\tilde{\theta}) \in \mathcal{Y}(\tilde{\theta},\tilde{\gamma},Z_{\tilde{\theta}}) - \mu^*_Y(\tilde{\theta})$ for any constant $\mu^*_Y(\tilde{\theta})$. Hence, rather than considering a sequence of values $\mu_{Y,m}$, we can fix some μ^*_Y and note that

$$Pr_{\mu^*_Y}\left\{Y(\tilde{\theta}) \in \mathcal{Y}^*_m | Z_{\tilde{\theta}} = z_{\tilde{\theta},m}\right\} \to 1,$$

where $\mathcal{Y}^*_m = \mathcal{Y}(\tilde{\theta},\tilde{\gamma},Z_{\tilde{\theta}}) - \mu_{Y,m}(\tilde{\theta}) + \mu^*_Y(\tilde{\theta})$. Confidence intervals for $\mu_{Y,m}(\tilde{\theta})$ in the original problem are equal to those for $\mu^*_Y(\tilde{\theta})$ in the new problem, shifted by $\mu_{Y,m}(\tilde{\theta}) - \mu^*_Y(\tilde{\theta})$. Hence, to prove the result it suffices to prove the equivalence of conditional and conventional confidence intervals in the problem with μ_Y fixed (and likewise for estimators).

To prove the result, we make use of the following lemma, which is proved below. First, we must introduce the following notation. Let $(c_{l,ET}(\mu_{Y,0},\mathcal{Y}),c_{u,ET}(\mu_{Y,0},\mathcal{Y}))$ denote the critical values for an equal-tailed test of $H_0: \mu_Y(\tilde{\theta}) = \mu_{Y,0}$ for $Y(\tilde{\theta}) \sim N\left(\mu_Y(\tilde{\theta}),\Sigma_Y(\tilde{\theta})\right)$ conditional on $Y(\tilde{\theta}) \in \mathcal{Y}$. That is, $(c_{l,ET}(\mu_{Y,0},\mathcal{Y}),c_{u,ET}(\mu_{Y,0},\mathcal{Y}))$ solve

$$F_{TN}(c_{l,ET}(\mu_{Y,0},\mathcal{Y});\mu_{Y,0},\mathcal{Y}) = \frac{\alpha}{2}$$

$$F_{TN}(c_{u,ET}(\mu_{Y,0},\mathcal{Y});\mu_{Y,0},\mathcal{Y}) = 1-\frac{\alpha}{2},$$

where $F_{TN}(\cdot;\mu_{Y,0},\mathcal{Y})$ is the distribution function for the normal distribution $N\left(\mu_{Y,0},\Sigma_Y(\tilde{\theta})\right)$ truncated to $\mathcal{Y}$. Similarly, let $(c_{l,U}(\mu_{Y,0},\mathcal{Y}),c_{u,U}(\mu_{Y,0},\mathcal{Y}))$ denote the critical values for the corresponding unbiased test. That is, $(c_{l,U}(\mu_{Y,0},\mathcal{Y}),c_{u,U}(\mu_{Y,0},\mathcal{Y}))$ solve

$$Pr\{\zeta \in [c_{l,U}(\mu_{Y,0},\mathcal{Y}),c_{u,U}(\mu_{Y,0},\mathcal{Y})]\} = 1-\alpha$$

$$E[\zeta 1\{\zeta \in [c_{l,U}(\mu_{Y,0},\mathcal{Y}),c_{u,U}(\mu_{Y,0},\mathcal{Y})]\}] = (1-\alpha)E[\zeta]$$

for $\zeta \sim \xi | \xi \in \mathcal{Y}$ where $\xi \sim N\left(\mu_{Y,0},\Sigma_Y(\tilde{\theta})\right)$.

Lemma 3

Suppose that we observe $Y(\tilde{\theta}) \sim N\left(\mu_Y(\tilde{\theta}),\Sigma_Y(\tilde{\theta})\right)$ conditional on $Y(\tilde{\theta})$ falling in a set $\mathcal{Y}$. If we hold $\left(\Sigma_Y(\tilde{\theta}),\mu_{Y,0}\right)$ fixed and consider a sequence of sets $\mathcal{Y}_m$ such that $Pr\left\{Y(\tilde{\theta}) \in \mathcal{Y}_m\right\} \to 1$, we have that for

$$\phi_{ET}(\mu_{Y,0}) = 1\left\{Y(\tilde{\theta}) \notin [c_{l,ET}(\mu_{Y,0},\mathcal{Y}_m),c_{u,ET}(\mu_{Y,0},\mathcal{Y}_m)]\right\} \tag{21}$$

and

$$\phi_U(\mu_{Y,0}) = 1\left\{Y(\tilde{\theta}) \notin [c_{l,U}(\mu_{Y,0},\mathcal{Y}_m),c_{u,U}(\mu_{Y,0},\mathcal{Y}_m)]\right\}, \tag{22}$$

$$(c_{l,ET}(\mu_{Y,0},\mathcal{Y}_m),c_{u,ET}(\mu_{Y,0},\mathcal{Y}_m)) \to \left(\mu_{Y,0} - c_{\frac{\alpha}{2},N}\sqrt{\Sigma_Y(\tilde{\theta})},\mu_{Y,0} + c_{\frac{\alpha}{2},N}\sqrt{\Sigma_Y(\tilde{\theta})}\right)$$

and

$$(c_{l,U}(\mu_{Y,0},\mathcal{Y}_m),c_{u,U}(\mu_{Y,0},\mathcal{Y}_m)) \to \left(\mu_{Y,0} - c_{\frac{\alpha}{2},N}\sqrt{\Sigma_Y(\tilde{\theta})},\mu_{Y,0} + c_{\frac{\alpha}{2},N}\sqrt{\Sigma_Y(\tilde{\theta})}\right).$$

To complete the proof, first note that CS_{ET} and CS_U are formed by inverting (families of) equal-tailed and unbiased tests, respectively. Let CS_m denote a generic conditional confidence interval formed by inverting a family of tests

$$\phi_m(\mu_{Y,0}) = 1\left\{Y(\tilde{\theta}) \notin [c_l(\mu_{Y,0},\mathcal{Y}_m^*), c_u(\mu_{Y,0},\mathcal{Y}_m^*)]\right\}.$$

Hence, we want to show that

$$CS_m \to_p \left[Y(\tilde{\theta}) - c_{\frac{\alpha}{2},N}, Y(\tilde{\theta}) + c_{\frac{\alpha}{2},N}\right], \tag{23}$$

as $m \to \infty$, for CS_m formed by inverting either (21) or (22).

We assume that CS_m is a finite interval for all m, which holds trivially for the equal-tailed confidence interval CS_{ET}, and holds for C_U by Lemma 5.5.1 of Lehmann and Romano (2005). For each value $\mu_{Y,0}$ our Lemma 3 implies that

$$\phi_m(\mu_{Y,0}) \to_p 1\left\{Y\left(\tilde{\theta}\right) \notin \left[\mu_{Y,0} - c_{\frac{\alpha}{2},N}, \mu_{Y,0} + c_{\frac{\alpha}{2},N}\right]\right\}$$

for ϕ_m equal to either (21) or (22). This convergence in probability holds jointly for all finite collections of values $\mu_{Y,0}$, however, which implies (23). The same argument works for the median unbiased estimator $\hat{\mu}_{\frac{1}{2}}$, which can also be viewed as the upper endpoint of a one-sided 50% confidence interval. $\square$

Proof of Proposition 4 We prove this result for the unconditional case, noting that since $Pr_{\mu_m}\left\{\hat{\theta} = \tilde{\theta}, \hat{\gamma} = \tilde{\gamma}\right\} \to 1$, the result conditional on $\left\{\hat{\theta} = \tilde{\theta}, \hat{\gamma} = \tilde{\gamma}\right\}$ follows immediately.

Note that by the law of iterated expectations, $Pr_{\mu_m}\left\{\hat{\theta} = \tilde{\theta}, \hat{\gamma} = \tilde{\gamma}\right\} \to 1$ implies that $Pr_{\mu_{Y,m}}\left\{\hat{\theta} = \tilde{\theta}, \hat{\gamma} = \tilde{\gamma} | Z_{\tilde{\theta}}\right\} \to_p 1$. Hence, if we define

$$g(\mu_Y, z) = Pr_{\mu_Y}\left\{\hat{\theta} = \tilde{\theta}, \hat{\gamma} = \tilde{\gamma} | Z_{\tilde{\theta}} = z\right\},$$

we see that $g(\mu_{Y,m}, Z_{\tilde{\theta}}) \to_p 1$.

Note, next, that for d the euclidian distance between the endpoints, if we define

$$h_\varepsilon(\mu_Y, z) = Pr_{\mu_Y}\{d(CS_U, CS_N) > \varepsilon | Z_{\tilde{\theta}} = z\},$$

Lemma 1 implies that for any sequence $(\mu_{Y,m}, z_m)$ such that $g(\mu_{Y,m}, z_m) \to 1$, $h_\varepsilon(\mu_{Y,m}, z_m) \to 0$. Hence, if we define $\mathcal{G}(\delta) = \{(\mu_Y, z) : g(\mu_Y, z) > 1 - \delta\}$ and $\mathcal{H}(\varepsilon) = \{(\mu_Y, z) : h_\varepsilon(\mu_Y, z) < \varepsilon\}$, we see that for all $\varepsilon > 0$ there exists $\delta(\varepsilon) > 0$ such that $\mathcal{G}(\delta(\varepsilon)) \subseteq \mathcal{H}(\varepsilon)$.

Hence, since our argument above implies that for all $\delta > 0$,

$$Pr_{\mu_m}\{(\mu_{Y,m}, Z_{\tilde{\theta}}) \in \mathcal{G}(\delta)\} \to 1,$$

we see that for all $\varepsilon > 0$,

$$Pr_{\mu_m}\{(\mu_{Y,m}, Z_{\tilde{\theta}}) \in \mathcal{H}(\varepsilon)\} \to 1$$

as well, which suffices to prove the desired claim for confidence intervals. The same argument likewise implies the result for our median unbiased estimator. □

Proof of Proposition 5 Provided $\hat{\theta}$ is unique with probability one, we can write

$$Pr_\mu\left\{\mu(\hat{\theta}) \in CS\right\} = \sum_{\tilde{\theta}\in\Theta,\tilde{\gamma}\in\Gamma} Pr_\mu\left\{\hat{\theta} = \tilde{\theta}, \hat{\gamma} = \tilde{\gamma}\right\} Pr_\mu\left\{\mu(\tilde{\theta}) \in CS | \hat{\theta} = \tilde{\theta}, \hat{\gamma} = \tilde{\gamma}\right\}.$$

Since $\sum_{\tilde{\theta}\in\Theta,\tilde{\gamma}\in\Gamma} Pr_\mu\left\{\hat{\theta} = \tilde{\theta}, \hat{\gamma} = \tilde{\gamma}\right\} = 1$, the result of the proposition follows immediately. □

Proof of Lemma 2 The assumption of the lemma implies that $X(\tilde{\theta}) - X(\theta)$ has a non-degenerate normal distribution for all μ. Since Θ is finite, almost-sure uniqueness of $\hat{\theta}$ follows immediately.

Proof of Proposition 6 The first part of the proposition follows immediately from Proposition 2. For the second part of the proposition, note that for CS^H either of the hybrid confidence intervals,

$$Pr_\mu\left\{\mu_Y(\hat{\theta}) \in CS^H\right\} = Pr_\mu\left\{\mu_Y(\hat{\theta}) \in CS_P^\beta\right\} \times$$

$$\sum_{\tilde{\theta}\in\Theta,\tilde{\gamma}\in\Gamma} Pr_\mu\left\{\hat{\theta} = \tilde{\theta}, \hat{\gamma} = \tilde{\gamma} | \mu_Y(\hat{\theta}) \in CS_P^\beta\right\} Pr_\mu\left\{\mu_Y(\tilde{\theta}) \in CS^H | \hat{\theta} = \tilde{\theta}, \hat{\gamma} = \tilde{\gamma}, \mu_Y(\tilde{\theta}) \in CS_P^\beta\right\}$$

$$= Pr_\mu\left\{\mu_Y(\hat{\theta}) \in CS_P^\beta\right\} \frac{1-\alpha}{1-\beta} \geq (1-\beta)\frac{1-\alpha}{1-\beta} = 1-\alpha,$$

where the second equality follows from the first part of the proposition. The upper bound follows by the same argument and the fact that $Pr_\mu\left\{\mu_Y(\hat{\theta}) \in CS_P^\beta\right\} \leq 1$. □

Proof of Proposition 7 We first establish uniqueness of $\hat{\mu}_\alpha^H$. To do so, it suffices to show that $F_{TN}^H(Y(\tilde{\theta}); \mu_Y(\tilde{\theta}), \tilde{\theta}, \tilde{\gamma}, Z_{\tilde{\theta}})$ is strictly decreasing in $\mu_Y(\tilde{\theta})$. Note first that this holds for the truncated normal assuming truncation that does not depend on $\mu_Y(\tilde{\theta})$ by Lemma A.1 of Lee et al. (2016). When we instead consider $F_{TN}^{H}(Y(\tilde{\theta});\mu_Y(\tilde{\theta}),\tilde{\theta},\tilde{\gamma},Z_{\tilde{\theta}})$, we impose truncation to

$$Y(\tilde{\theta})\in\left[\mu_Y(\tilde{\theta})-c_\beta\sqrt{\Sigma_Y(\tilde{\theta})},\mu_Y(\tilde{\theta})+c_\beta\sqrt{\Sigma_Y(\tilde{\theta})}\right].$$

Since this interval shifts upwards as we increase $\mu_Y(\tilde{\theta})$, $F_{TN}^{H}(Y(\hat{\theta});\mu_Y(\tilde{\theta}),\tilde{\theta},\tilde{\gamma},Z_{\tilde{\theta}})$ is a fortiori decreasing in $\mu_Y(\tilde{\theta})$. Uniqueness of $\hat{\mu}_\alpha^H$ for $\alpha\in(0,1)$ follows. Note, next, that $F_{TN}^{H}(Y(\tilde{\theta});\mu_Y(\tilde{\theta}),\tilde{\theta},\tilde{\gamma},Z_{\tilde{\theta}})\in\{0,1\}$ for $\mu_Y(\tilde{\theta})\notin CS_P^\beta$ from which we immediately see that $\hat{\mu}_\alpha^H\in CS_P^\beta$.

Finally, note that for $\mu_Y(\tilde{\theta})$ the true value,

$$F_{TN}^{H}(Y(\hat{\theta});\mu_Y(\tilde{\theta}),\tilde{\theta},\tilde{\gamma},Z_{\tilde{\theta}})\sim U[0,1]$$

conditional on $\left\{\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma},Z_{\hat{\theta}}=z_{\tilde{\theta}},\mu_Y(\tilde{\theta})\in CS_P^\beta\right\}$. Since $F_{TN}^{H}(Y(\hat{\theta});\mu_Y(\tilde{\theta}),\tilde{\theta},\tilde{\gamma},Z_{\tilde{\theta}})$ is decreasing in $\mu_Y(\tilde{\theta})$,

$$Pr_\mu\left\{\hat{\mu}_\alpha^H\geq\mu_Y(\tilde{\theta})|\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma},Z_{\hat{\theta}}=z_{\tilde{\theta}},\mu_Y(\tilde{\theta})\in CS_P^\beta\right\}$$

$$=Pr_\mu\left\{F_{TN}^{H}(Y(\hat{\theta});\mu_Y(\tilde{\theta}),\tilde{\theta},\tilde{\gamma},Z_{\tilde{\theta}})\geq 1-\alpha|\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma},Z_{\hat{\theta}}=z_{\tilde{\theta}},\mu_Y(\tilde{\theta})\in CS_P^\beta\right\}=\alpha,$$

and thus $\hat{\mu}_\alpha^H$ is α-quantile-unbiased conditional on $\left\{\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma},Z_{\hat{\theta}}=z_{\tilde{\theta}},\mu_Y(\tilde{\theta})\in CS_P^\beta\right\}$. We can drop the conditioning on $Z_{\tilde{\theta}}$ by the law of iterated expectations, and α-quantile-unbiasedness conditional on $\mu_Y(\tilde{\theta})\in CS_P^\beta$ follows by the same argument as in the proof of Proposition 5.

Proof of Lemma 3 Note that we can assume without loss of generality that $\mu_{Y,0}=0$ and $\Sigma_Y(\tilde{\theta})=1$ since we can define $Y^*(\tilde{\theta})=\left(Y(\tilde{\theta})-\mu_{Y,0}\right)/\sqrt{\Sigma_Y(\tilde{\theta})}$ and consider the problem of testing that the mean of $Y^*(\tilde{\theta})$ is zero (transforming the set $\mathcal{Y}_m$ accordingly). After deriving critical values (c_l^*,c_u^*) in this transformed problem, we can recover critical values for our original problem as $(c_l,c_u)=\sqrt{\Sigma_Y(\tilde{\theta})}(c_l^*,c_u^*)+\mu_{Y,0}$. Hence, for the remainder of the proof we assume that $\mu_{Y,0}=0$ and $\Sigma_Y(\tilde{\theta})=1$.

Equal-Tailed Test We consider first the equal-tailed test. Note that this test rejects if and only if

$$Y(\tilde{\theta})\notin[c_{l,ET}(\mathcal{Y}),c_{u,ET}(\mathcal{Y})],$$

where we suppress the dependence of the critical values on $\mu_{Y,0}=0$ for simplicity, and $(c_{l,ET}(\mathcal{Y}),c_{u,ET}(\mathcal{Y}))$ solve

$$F_{TN}(c_{l,ET}(\mathcal{Y}),\mathcal{Y})=\frac{\alpha}{2}$$

$$F_{TN}(c_{u,ET}(\mathcal{Y}),\mathcal{Y})=1-\frac{\alpha}{2}.$$

for $F_{TN}(\cdot,\mathcal{Y})$ the distribution function of a standard normal random variable truncated to $\mathcal{Y}$. Recall that we can write the density corresponding to $F_{TN}(y,\mathcal{Y})$ as $\frac{1\{y\in\mathcal{Y}\}}{Pr\{\xi\in\mathcal{Y}\}}f_N(y)$ where f_N is the standard normal density and $Pr\{\xi\in\mathcal{Y}\}$ is the probability that $\xi\in\mathcal{Y}$ for $\xi\sim N(0,1)$. Hence, we can write

$$F_{TN}(y,\mathcal{Y})=\frac{\int_{-\infty}^{y}1\{\tilde{y}\in\mathcal{Y}\}f_N(\tilde{y})d\tilde{y}}{Pr\{\xi\in\mathcal{Y}\}}.$$

Note that that for all y we can write

$$F_{TN}(y,\mathcal{Y}_m)=a_m(y)+F_N(y),$$

where F_N is the standard normal distribution function and

$$a_m(y)=\frac{\int_{-\infty}^{y}1\{\tilde{y}\in\mathcal{Y}_m\}f_N(\tilde{y})d\tilde{y}}{Pr\{\xi\in\mathcal{Y}_m\}}-F_N(y).$$

Recall, however, that $Pr\{\xi\in\mathcal{Y}_m\}\to 1$ and

$$\left|\int_{-\infty}^{y}1\{\tilde{y}\in\mathcal{Y}_m\}f_N(\tilde{y})d\tilde{y}-F_N(y)\right|=\left|\int_{-\infty}^{y}[1\{\tilde{y}\in\mathcal{Y}_m\}-1]f_N(\tilde{y})d\tilde{y}\right|$$
$$=\int_{-\infty}^{y}1\{\tilde{y}\notin\mathcal{Y}_m\}f_N(\tilde{y})d\tilde{y}\leq Pr\{\xi\notin\mathcal{Y}_m\}\to 0$$

for all y, so $a_m(y)\to 0$ for all y. Theorem 2.11 in Van der Vaart (1998) then implies that $a_m(y)\to 0$ uniformly in y as well.

Note next that

$$F_{TN}(c_{l,ET}(\mathcal{Y}_m),\mathcal{Y}_m)=a_m(c_{l,ET}(\mathcal{Y}_m))+F_N(c_{l,ET}(\mathcal{Y}_m))=\frac{\alpha}{2}$$

implies

$$c_{l,ET}(\mathcal{Y}_m)=F_N^{-1}\left(\frac{\alpha}{2}-a_m(c_{l,ET}(\mathcal{Y}_m))\right),$$

and thus that $c_{l,ET}(\mathcal{Y}_m) \to F_N^{-1}\left(\frac{\alpha}{2}\right)$. Using the same argument, we can show that $c_{u,ET}(\mathcal{Y}_m) \to F_N^{-1}\left(1-\frac{\alpha}{2}\right)$, as desired.

Unbiased Test We next consider the unbiased test. Recall that critical values $c_{l,U}(\mathcal{Y})$, $c_{u,U}(\mathcal{Y})$ for the unbiased test solve

$$Pr\{\zeta \in [c_{l,U}(\mathcal{Y}), c_{u,U}(\mathcal{Y})]\} = 1-\alpha$$

$$E[\zeta 1\{\zeta \in [c_{l,U}(\mathcal{Y}), c_{u,U}(\mathcal{Y})]\}] = (1-\alpha)E[\zeta]$$

for $\zeta \sim \xi | \xi \in \mathcal{Y}$ where $\xi \sim N(0,1)$.

Note that for ζ_m the truncated normal random variable corresponding to $\mathcal{Y}_m$, we can write

$$Pr\{\zeta_m \in [c_l, c_u]\} = a_m(c_l, c_u) + (F_N(c_u) - F_N(c_l))$$

with

$$a_m(c_l, c_u) = (F_N(c_l) - Pr\{\zeta_m \le c_l\}) - (F_N(c_u) - Pr\{\zeta_m \le c_u\}).$$

As in the argument for equal-tailed tests above, we see that both $F_N(c_u) - Pr\{\zeta_m \le c_u\}$ and $F_N(c_l) - Pr\{\zeta_m \le c_l\}$ converge to zero pointwise, and thus uniformly in c_u and c_l by Theorem 2.11 in Van der Vaart (1998). Hence, $a_m(c_l, c_u) \to 0$ uniformly in (c_l, c_u).

Note, next, that we can write

$$E[\zeta_m 1\{\zeta_m \in [c_l, c_u]\}] = [\xi 1\{\xi \in [c_l, c_u]\}] + b_m(c_l, c_u)$$

for

$$\begin{aligned} b_m(c_l, c_u) &= E[\zeta_m 1\{\zeta_m \in [c_l, c_u]\}] - [\xi 1\{\xi \in [c_l, c_u]\}] \\ &= \int_{c_l}^{c_u} \left(\frac{1\{y \in \mathcal{Y}_m\}}{Pr\{\xi \in \mathcal{Y}_m\}} - 1 \right) y f_N(y) dy. \end{aligned}$$

Note, however, that

$$\int_{c_l}^{c_u} (1\{y \in \mathcal{Y}_m\} - 1) y f_N(y) dy \le E[|\xi| 1\{\xi \notin \mathcal{Y}_m\}].$$

Hence, since

$$\left| \int_{c_l}^{c_u} \left(\frac{1\{y \in \mathcal{Y}_m\}}{Pr\{\xi \in \mathcal{Y}_m\}} - 1\{y \in \mathcal{Y}_m\} \right) y f_N(y) dy \right|$$

$$\leq \left| \left(\frac{1}{Pr\{\xi \in \mathcal{Y}_m\}} - 1 \right) \right| E[|\xi| 1\{\xi \notin \mathcal{Y}_m\}] \leq \left| \left(\frac{1}{Pr\{\xi \in \mathcal{Y}_m\}} - 1 \right) \right| \sqrt{P(\xi \notin \mathcal{Y}_m)}$$

by the Cauchy-Schwartz Inequality, where the right hand side tends to zero and doesn't depend on (c_l, c_u), $b_m(c_l, c_u)$ converges to zero uniformly in (c_l, c_u).

Next, let us define $(c_{l,m}, c_{u,m})$ as the solutions to

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

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

From our results above, we can re-write the problem solved by $(c_{l,m}, c_{u,m})$ as

$$F_N(c_u) - F_N(c_l) = 1 - \alpha - a_m(c_l, c_u)$$

$$E[\xi 1\{\xi \in [c_l, c_u]\}] = (1 - \alpha) E[\zeta_m] - b_m(c_l, c_u).$$

Letting

$$\bar{a}_m = \sup_{c_l, c_u} |a_m(c_l, c_u)|,$$

$$\bar{b}_m = \sup_{c_l, c_u} |b_m(c_l, c_u)|$$

we thus see that $(c_{l,m}, c_{u,m})$ solves

$$F_N(c_u) - F_N(c_l) = 1 - \alpha - a_m^*$$

$$E[\xi 1\{\xi \in [c_l, c_u]\}] = (1 - \alpha) E[\zeta_m] - b_m^*$$

for some $a_m^* \in [-\bar{a}_m, \bar{a}_m]$, $b_m^* \in \left[-\bar{b}_m, \bar{b}_m\right]$. We will next show that for any sequence of values (a_m^*, b_m^*) such that $a_m^* \in [-\bar{a}_m, \bar{a}_m]$ and $b_m^* \in \left[-\bar{b}_m, \bar{b}_m\right]$ for all m, the implied solutions $c_{l,m}(a_m^*, b_m^*)$, $c_{u,m}(a_m^*, b_m^*)$ converge to $F_N^{-1}\left(\frac{\alpha}{2}\right)$ and $F_N^{-1}\left(1 - \frac{\alpha}{2}\right)$. This follows from the next lemma, which is proved below.

Lemma 4

Suppose that $c_{l,m}$ and $c_{u,m}$ solve

$$Pr\{\xi \in [c_l, c_u]\} = 1 - \alpha + a_m,$$

$$E[\xi 1\{\xi \in [c_l, c_u]\}] = d_m$$

for a_m, $d_m \to 0$. *Then* $(c_{l,m}, c_{u,m}) \to \left(-c_{\frac{\alpha}{2},N}, c_{\frac{\alpha}{2},N}\right)$.

Using this lemma, since $E[\zeta_m] \to 0$ as $m \to \infty$ we see that for any sequence of values $(a_m^*, b_m^*) \to 0$,

$$(c_{l,m}(a_m^*, b_m^*), c_{u,m}(a_m^*, b_m^*)) \to \left(-c_{\frac{\alpha}{2},N}, c_{\frac{\alpha}{2},N}\right).$$

However, since $\bar{a}_m, \bar{b}_m \to 0$ we know that the values a_m^* and b_m^* corresponding to the true $c_{l,m}$, $c_{u,m}$ must converge to zero. Hence $(c_{l,m}, c_{u,m}) \to \left(-c_{\frac{\alpha}{2},N}, c_{\frac{\alpha}{2},N}\right)$ as we wanted to show. $\square$

Proof of Lemma 4 Note that the critical values solve

$$f(a_m, d_m, c) = \left(\begin{array}{c} F_N(c_u) - F_N(c_l) - (1-\alpha) - a_m \\ \int_{c_l}^{c_u} y f_N(y) dy - d_m \end{array} \right) = 0.$$

We can simplify this expression, since $\frac{\partial}{\partial y} f_N(y) = -y f_N(y)$, so

$$\int_{c_l}^{c_u} y f_N(y) dy = f_N(c_l) - f_N(c_u).$$

We thus must solve the system of equations

$$F_N(c_u) - F_N(c_l) = (1-\alpha) - a_m$$

$$f_N(c_l) - f_N(c_u) = d_m$$

or more compactly $g(c) - v_m = 0$, for

$$g(c) = \left(\begin{array}{c} F_N(c_u) - F_N(c_l) \\ f_N(c_l) - f_N(c_u) \end{array} \right), \quad v_m = \left(\begin{array}{c} a_m + (1-\alpha) \\ d_m \end{array} \right).$$

Note that for $v_m = (1-\alpha, 0)'$ this system is solved by $c = \left(-c_{\frac{\alpha}{2},N}, c_{\frac{\alpha}{2},N}\right)$. Further,

$$\frac{\partial}{\partial c} g(c) = \left(\begin{array}{cc} -f_N(c_l) & f_N(c_u) \\ -c_l f_N(c_l) & c_u f_N(c_u) \end{array} \right),$$

which evaluated at $c = \left(-c_{\frac{\alpha}{2},N}, c_{\frac{\alpha}{2},N}\right)$ is equal to

$$\left(\begin{array}{cc} -f_N\left(c_{\frac{\alpha}{2},N}\right) & f_N\left(c_{\frac{\alpha}{2},N}\right) \\ c_{\frac{\alpha}{2},N} f_N\left(c_{\frac{\alpha}{2},N}\right) & c_{\frac{\alpha}{2},N} f_N\left(c_{\frac{\alpha}{2},N}\right) \end{array} \right)$$

and has full rank for all $\alpha \in (0,1)$. Thus, by the implicit function theorem there exists an open neighborhood V of $v_\infty = (1-\alpha,0)$ such that $g(c) - v = 0$ has a unique solution $c(v)$ for $v \in V$ and $c(v)$ is continuously differentiable. Hence, if we consider any sequence of values $v_m \to (1-\alpha,0)$, we see that

$$c(v_m) \to \begin{pmatrix} -c_{\frac{\alpha}{2},N} \\ c_{\frac{\alpha}{2},N} \end{pmatrix},$$

again as we wanted to show. $\square$

B Additional Results: Details for Empirical Welfare Maximization Example

Here, we derive the form of the conditioning event $\mathcal{Y}_\gamma(1,Z_{\tilde{\theta}})$ discussed in Section 4.2, including for cases when $\Sigma_{XY}(\tilde{\theta}) - \Sigma_{XY}(\tilde{\theta},0) \leq 0$. Note that we can write

$$\left\{X(\tilde{\theta}) - X(0) \geq c\right\} = \left\{Z_{\tilde{\theta}}(\tilde{\theta}) - Z_{\tilde{\theta}}(0) + \frac{\Sigma_{XY}(\tilde{\theta}) - \Sigma_{XY}(\tilde{\theta},0)}{\Sigma_Y(\tilde{\theta})} Y(\tilde{\theta}) \geq c\right\}.$$

Rearranging, we see that

$$\mathcal{Y}_\gamma(1,Z_{\tilde{\theta}}) = \begin{cases} \left\{y : y \geq \frac{\Sigma_Y(\tilde{\theta})\left(c - Z_{\tilde{\theta}}(\tilde{\theta}) + Z_{\tilde{\theta}}(0)\right)}{\Sigma_{XY}(\tilde{\theta}) - \Sigma_{XY}(\tilde{\theta},0)}\right\} & \text{if } \Sigma_{XY}(\tilde{\theta}) - \Sigma_{XY}(\tilde{\theta},0) > 0 \\ \left\{y : y \leq \frac{\Sigma_Y(\tilde{\theta})\left(c - Z_{\tilde{\theta}}(\tilde{\theta}) + Z_{\tilde{\theta}}(0)\right)}{\Sigma_{XY}(\tilde{\theta}) - \Sigma_{XY}(\tilde{\theta},0)}\right\} & \text{if } \Sigma_{XY}(\tilde{\theta}) - \Sigma_{XY}(\tilde{\theta},0) < 0 \\ \mathbb{R} & \begin{array}{c} \text{if } \Sigma_{XY}(\tilde{\theta}) - \Sigma_{XY}(\tilde{\theta},0) = 0 \\ \text{and } Z_{\tilde{\theta}}(\tilde{\theta}) - Z_{\tilde{\theta}}(0) \geq c \end{array} \\ \emptyset & \begin{array}{c} \text{if } \Sigma_{XY}(\tilde{\theta}) - \Sigma_{XY}(\tilde{\theta},0) = 0 \\ \text{and } Z_{\tilde{\theta}}(\tilde{\theta}) - Z_{\tilde{\theta}}(0) < c. \end{array} \end{cases}$$

C Alternatives to Conventional Sample Splitting

In Section 4.3 of the main text, we discuss the relationship of our conditional approach to conventional sample splitting methods and note that the results of Fithian et al. (2017) imply that traditional sample splitting methods are dominated in our setting. Here, we derive optimal split-sample confidence intervals and estimators as well as easy-to-implement confidence intervals and estimators that dominate their conventional split-sample counterparts in the asymptotic version of the split-sample problem.

The Split-Sample Limit Experiment Let τ denote the fraction of the full sample used to compute the estimated maximum and (X_n^1,Y_n^1) and (X_n^2,Y_n^2) denote rescaled data corresponding to the first and second portions of the data such that

$$(X_n^1,Y_n^1)=\tau^{-1/2}(X_{[\tau\cdot n]},Y_{[\tau\cdot n]}),$$

$$(X_n^2,Y_n^2)=(1-\tau)^{-1}\left((X_n,Y_n)-\sqrt{\tau}(X_{[\tau\cdot n]+1},Y_{[\tau\cdot n]+1})\right)$$

with $[a]$ denoting the nearest integer to $a\in\mathbb{R}$. Finally, let $\hat{\theta}_n^1=\text{argmax}_{\theta\in\Theta}X_n^1(\theta)$ or $\hat{\theta}_n^1=\text{argmax}_{\theta\in\Theta}\|X_n^1(\theta)\|$, as in Andrews et al. (2019), denote the estimated maximum from the first part of the sample. In large samples, (X_n^1,Y_n^1), (X_n^2,Y_n^2) and $\hat{\theta}_n^1$ behave according to[23]

$$\begin{pmatrix} X^1 \\ Y^1 \end{pmatrix}\sim N(\mu,\Sigma),$$

$$\begin{pmatrix} X^2 \\ Y^2 \end{pmatrix}\sim N\left(\mu,c^{-1}\Sigma\right)$$

and

$$\hat{\theta}^1=\text{argmax}_{\theta\in\Theta}X^1(\theta)$$

or

$$\hat{\theta}^1=\text{argmax}_{\theta\in\Theta}\left\|X^1(\theta)\right\|,$$

where $c=(1-\tau)/\tau$ and (X^1,Y^1) is independent of (X^2,Y^2). This is the generalization of the asymptotic problem discussed in Section 4.3 of the main text to arbitrary sample splits.[24]

Traditional sample splitting methods base inference on $Y^2(\hat{\theta}^1)$. Since Y^2 is independent of X^1, and thus of $\hat{\theta}^1$, this ensures the (conditional) median-unbiasedness of conventional split-sample estimates $Y^2(\hat{\theta}^1)$ and the (conditional) validity of conventional split-sample confidence intervals

$$CS_{SS}=\left[Y^2(\hat{\theta}^1)-\sqrt{c^{-1}\Sigma_Y(\hat{\theta}^1)}c_{\alpha/2,N},Y^2(\hat{\theta}^1)+\sqrt{c^{-1}\Sigma_Y(\hat{\theta}^1)}c_{\alpha/2,N}\right]$$

[23]The quantity Σ in the exposition of this section corresponds to the quantity Σ in the main text, multiplied by τ^{-1}.

[24]For simplicity of exposition, in this section we suppress the possibility of using additional conditioning variables $\hat{\gamma}_n=\gamma\left(X_n^1\right)$ with asymptotic counterpart $\hat{\gamma}=\gamma\left(X^1\right)$.

but does not make full use of the information in the data. To derive optimal procedures in the sample splitting framework, we first derive a sufficient statistic for the unknown parameter μ conditional on $\left\{\hat{\theta}^1=\tilde{\theta}\right\}$ and then apply classical exponential family results as in Section 4 of the main text.

Optimal Estimators and Confidence Sets The joint (unconditional) density of (X^1,Y^1,X^2,Y^2) is proportional to

$$\exp\left(-\frac{1}{2}\left(\begin{pmatrix}X^1\\Y^1\end{pmatrix}-\mu\right)'\Sigma^{-1}\left(\begin{pmatrix}X^1\\Y^1\end{pmatrix}-\mu\right)\right)\exp\left(-\frac{c}{2}\left(\begin{pmatrix}X^2\\Y^2\end{pmatrix}-\mu\right)'\Sigma^{-1}\left(\begin{pmatrix}X^2\\Y^2\end{pmatrix}-\mu\right)\right).$$

The conditional density given $\left\{\hat{\theta}^1=\tilde{\theta}\right\}$ is thus propotional to

$$\frac{1\left\{X^1\in\mathcal{X}^1\left(\tilde{\theta}\right)\right\}}{Pr_\mu\left\{X^1\in\mathcal{X}^1\left(\tilde{\theta}\right)\right\}}\exp\left(-\frac{1}{2}\left(\begin{pmatrix}X^1\\Y^1\end{pmatrix}-\mu\right)'\Sigma^{-1}\left(\begin{pmatrix}X^1\\Y^1\end{pmatrix}-\mu\right)\right)\times$$
$$\exp\left(-\frac{c}{2}\left(\begin{pmatrix}X^2\\Y^2\end{pmatrix}-\mu\right)'\Sigma^{-1}\left(\begin{pmatrix}X^2\\Y^2\end{pmatrix}-\mu\right)\right)$$

with $\mathcal{X}^1(\tilde{\theta})=\{X^1:\hat{\theta}=\tilde{\theta}\}$, which we can re-write as

$$g_1\left(X^1,Y^1\right)g_2\left(X^2,Y^2\right)h(\mu)\exp\left(\left(\begin{pmatrix}X^1\\Y^1\end{pmatrix}+c\begin{pmatrix}X^2\\Y^2\end{pmatrix}\right)'\Sigma^{-1}\mu\right)$$

for

$$g_1\left(X^1,Y^1\right)=1\left\{X^1\in\mathcal{X}^1\left(\tilde{\theta}\right)\right\}\exp\left(-\frac{1}{2}\begin{pmatrix}X^1\\Y^1\end{pmatrix}'\Sigma^{-1}\begin{pmatrix}X^1\\Y^1\end{pmatrix}\right),$$
$$g_2\left(X^2,Y^2\right)=\exp\left(-\frac{c}{2}\begin{pmatrix}X^2\\Y^2\end{pmatrix}'\Sigma^{-1}\begin{pmatrix}X^2\\Y^2\end{pmatrix}\right),$$

and

$$h(\mu)=\frac{1}{Pr_\mu\left\{X^1\in\mathcal{X}^1\left(\tilde{\theta}\right)\right\}}\exp\left(-\frac{1+c}{2}\mu'\Sigma^{-1}\mu\right).$$

This exponential family structure shows that $\left(\begin{array}{c} X^* \\ Y^* \end{array}\right)=\left(\left(\begin{array}{c} X^1 \\ Y^1 \end{array}\right)+c\left(\begin{array}{c} X^2 \\ Y^2 \end{array}\right)\right)$ is sufficient for μ. Hence, for any function of (X^1,Y^1,X^2,Y^2), there exists a (potentially randomized) function of (X^*,Y^*) with the same distribution for all μ. Thus, to study questions of optimality it is without loss to limit attention to confidence intervals and estimators that depend only on (X^*,Y^*).

Now that we have derived a sufficient statistic (X^*,Y^*) for μ, we turn to the question of how to construct optimal estimators and confidence intervals for $\mu_Y(\tilde{\theta})$ conditional on $\left\{\hat{\theta}=\tilde{\theta}\right\}$. Note that the unconditional density of (X^*,Y^*) is proportional to

$$\exp\left(-\frac{1}{2+2c}\left(\left(\begin{array}{c} X^* \\ Y^* \end{array}\right)-(1+c)\mu\right)'\Sigma^{-1}\left(\left(\begin{array}{c} X^* \\ Y^* \end{array}\right)-(1+c)\mu\right)\right).$$

The density of (X^*,Y^*) given $\left\{\hat{\theta}^1=\tilde{\theta}\right\}$ is thus proportional to

$$\frac{Pr\left\{X^1\in\mathcal{X}^1\left(\tilde{\theta}\right)|X^*,Y^*\right\}}{Pr_\mu\left\{X^1\in\mathcal{X}^1\left(\tilde{\theta}\right)\right\}}\exp\left(-\frac{1}{2+2c}\left(\left(\begin{array}{c} X^* \\ Y^* \end{array}\right)-(1+c)\mu\right)'\Sigma^{-1}\left(\left(\begin{array}{c} X^* \\ Y^* \end{array}\right)-(1+c)\mu\right)\right),$$

where we have used sufficiency to drop dependence of the numerator on μ.

This joint distribution has the same exponential family structure used to derive the optimal estimators and confidence intervals in the main text (see the proofs of Propositions 1 and 2). Hence, the same arguments deliver optimal procedures for the split-sample setting. Specifically, for

$$Z^*_{\tilde{\theta}}=\left(\begin{array}{c} X^* \\ Y^* \end{array}\right)-\left(Cov\left(\left(\begin{array}{c} X^* \\ Y^* \end{array}\right),Y^*\left(\tilde{\theta}\right)\right)/\Sigma_{Y^*}\left(\tilde{\theta}\right)\right)Y^*\left(\tilde{\theta}\right),$$

where Σ_{Y^*} denotes the variance of Y^*, we can re-write

$$\exp\left(\left(\left(\begin{array}{c} X^1 \\ Y^1 \end{array}\right)+c\left(\begin{array}{c} X^2 \\ Y^2 \end{array}\right)\right)\Sigma^{-1}\mu\right)=\exp\left(Y^*\left(\tilde{\theta}\right)\mu_{Y^*}\left(\tilde{\theta}\right)/\Sigma_{Y^*}\left(\tilde{\theta}\right)+Z^*_{\tilde{\theta}}\Sigma^+_{Z^*}\mu_{Z^*}\right)$$

for Σ_{Z^*} the variance of Z^*, A^+ the Moore-Penrose pseudoinverse of a matrix A, and

$$\mu_{Z^*}=(1+c)\mu-\left(Cov\left(\left(\begin{array}{c} X^* \\ Y^* \end{array}\right),Y^*\left(\tilde{\theta}\right)\right)/Var\left(Y^*\left(\tilde{\theta}\right)\right)\right)\mu_{Y^*}\left(\tilde{\theta}\right).$$

This expression shows that when we are interested in inference on $\mu_Y(\tilde{\theta})$ conditional on $\left\{\hat{\theta}^1 = \tilde{\theta}\right\}$, μ_{Z^*} is the nuisance parameter, and $Z^*_{\tilde{\theta}}$ is minimal sufficient for this parameter relative to observing (X^1, Y^1, X^2, Y^2).

If we let $F^*_{SS}(Y^*(\tilde{\theta}); \mu_{Y^*}(\tilde{\theta}), \tilde{\theta}, z^*)$ denote the conditional distribution function of $Y^*|Z^* = z^*, \hat{\theta}^1 = \tilde{\theta}$, then the same arguments used to prove Proposition 1 show that the optimal α quantile-unbiased estimator $\hat{\mu}^*_{SS,\alpha}$ in the sample splitting problem solves

$$F^*_{SS}(Y^*(\hat{\theta}^1); (1+c)\hat{\mu}^*_{SS,\alpha}, \tilde{\theta}, Z^*_{\tilde{\theta}}) = 1-\alpha.$$

Likewise, the same arguments used to prove Proposition 2 show that the optimal two-sided unbiased test rejects $H_0 : \mu_Y(\tilde{\theta}) = \mu_{Y,0}$ when

$$Y^*(\tilde{\theta}) \notin \left[c_l\left(Z^*_{\tilde{\theta}}\right), c_u\left(Z^*_{\tilde{\theta}}\right)\right],$$

where $c_l(z)$, $c_u(z)$ solve

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

with ζ distributed according to $F^*_{SS}(\cdot; (1+c)\mu_{Y,0}, \tilde{\theta}, z)$. These optimal procedures condition on $Z^*_{\tilde{\theta}}$ rather than (X^1, Y^1) and so, unlike conventional sample splitting, continue to treat (X^1, Y^1) as random for inference.

Feasible Dominating Estimators and Confidence Sets To implement the optimal split-sample proecdures, we need to evaluate (or at least be able to draw from) the conditional distribution $F^*_{SS}(\cdot; (1+c)\mu_{Y,0}, \tilde{\theta}, z)$. Unfortunately, however, it is not computationally straightforward to do so since $Y^*|Z^* = z^*, \hat{\theta}^1 = \tilde{\theta}$ is distributed as a normal random variable truncated to a dependent random set. We thus introduce side constraints to derive procedures that, although they are not fully optimal in the unconstrained problem, are computationally straightforward to implement and dominate conventional sample splitting procedures. These computationally feasible procedures are optimal within the class of split-sample procedures that condition on $\{\hat{\theta}^1 = \tilde{\theta}\}$ and the realizations of

$$Z^i_{\tilde{\theta}} = X^i - \left(\Sigma_{XY}\left(\cdot, \tilde{\theta}\right) / \Sigma_Y\left(\tilde{\theta}\right)\right) Y^i\left(\tilde{\theta}\right)$$

for $i = 1, 2$, where $(Z^1_{\tilde{\theta}}, Z^2_{\tilde{\theta}})$ is a sufficient statistic for the nuisance parameter μ_X. Since $Y^2(\hat{\theta}^1)|\{\hat{\theta}^1 = \tilde{\theta}, (Z^1_{\tilde{\theta}}, Z^2_{\tilde{\theta}}) = (z^1, z^1)\} \sim Y^2(\tilde{\theta})$, the conventional split-sample estimator $Y^2(\hat{\theta}^1)$

and confidence interval CS_{SS} fall within the class of split-sample conditional procedures that condition on $\{\hat{\theta}^1 = \tilde{\theta}\}$ and $(Z^1_{\tilde{\theta}}, Z^2_{\tilde{\theta}})$. These conventional procedures are therefore dominated by the optimal procedures within this class, which we now describe.

Standard exponential family arguments show that $(Z^1_{\tilde{\theta}},Z^2_{\tilde{\theta}})$ is sufficient for the nuisance parameter μ_X and, conditional on $\{\hat{\theta}^1=\tilde{\theta}\}$ and $(Z^1_{\tilde{\theta}},Z^2_{\tilde{\theta}})$, optimal estimation and inference is based upon the conditional distribution of $Y^*(\tilde{\theta})$. Note that since $Y^2(\tilde{\theta})$ is independent of $(Z^1_{\tilde{\theta}},Z^2_{\tilde{\theta}})$ and both $\hat{\theta}^1$ and $Y^2(\tilde{\theta})$ are independent of $Z^2_{\tilde{\theta}}$,

$$Y^*(\tilde{\theta})|\{\hat{\theta}^1=\tilde{\theta},(Z^1_{\tilde{\theta}},Z^2_{\tilde{\theta}})=(z^1,z^2)\}\sim Y^1(\tilde{\theta})|\{\hat{\theta}^1=\tilde{\theta},Z^1_{\tilde{\theta}}=z^1\}+cY^2(\tilde{\theta}).$$

Thus, the feasible dominating split-sample procedures rely upon the computation of the distribution function of $Y^1(\tilde{\theta})|\{\hat{\theta}^1=\tilde{\theta},Z^1_{\tilde{\theta}}=z^1\}+cY^2(\tilde{\theta})$. We now describe a fast method for computing this object.

In analogy with full sample inference, let

$$\mathcal{Y}^1(\tilde{\theta},z^1)=\left\{y^1:z^1+\left(\Sigma_{XY}\left(\cdot,\tilde{\theta}\right)/\Sigma_Y\left(\tilde{\theta}\right)\right)y^1\in\mathcal{X}^1(\tilde{\theta})\right\}$$

so that conditional on $\{\hat{\theta}^1 = \tilde{\theta}\}$ and $Z^1_{\tilde{\theta}} = z^1$, $Y^1(\tilde{\theta})$ follows a one-dimensional truncated normal distribution with truncation set $\mathcal{Y}^1(\tilde{\theta}, z^1)$. Note that in both the level and norm maximization contexts, $\mathcal{Y}^1(\tilde{\theta},z^1)$ can be expressed as a finite union of disjoint intervals: $\mathcal{Y}^1(\tilde{\theta},z^1)=\bigcup_{k=1}^K[\ell_k(z^1),u_k(z^1)]$, where the dependence of $\ell_k(z^1)$ and $u_k(z^1)$ for $k=1,\ldots,K$ on $\tilde{\theta}$ is suppressed for notational simplicity. Note that $Y^1(\tilde{\theta})|\{\hat{\theta}^1=\tilde{\theta},Z^1_{\tilde{\theta}}=z^1\}$ is distributed as $\xi^1|\xi^1\in\mathcal{Y}^1(\tilde{\theta},z^1)$, where $\xi^1\sim N(\mu_Y(\tilde{\theta}),\Sigma_Y(\tilde{\theta}))$. The density function of $Y^1(\tilde{\theta})|\{\hat{\theta}^1=\tilde{\theta},Z^1_{\tilde{\theta}}=z^1\}$ is thus

$$f^1(y^1)=\frac{\sum_{k=1}^K f_N\left((y^1-\mu_Y(\tilde{\theta}))/\sqrt{\Sigma_Y(\tilde{\theta})}\right)1(\ell_k(z^1)\leq y^1\leq u_k(z^1))}{\sqrt{\Sigma_Y(\tilde{\theta})}\sum_{k=1}^K\left(F_N\left((u_k(z^1)-\mu_Y(\tilde{\theta}))/\sqrt{\Sigma_Y(\tilde{\theta})}\right)-F_N\left((\ell_k(z^1)-\mu_Y(\tilde{\theta}))/\sqrt{\Sigma_Y(\tilde{\theta})}\right)\right)}$$

and $cY^2(\tilde{\theta})$ has density function $f^2(y^2)=c^{-1/2}\Sigma_Y(\tilde{\theta})^{-1/2}f_N\left((y^2-c\mu)/\sqrt{c\Sigma_Y(\tilde{\theta})}\right)$. Therefore, since $Y^1(\tilde{\theta})|\{\hat{\theta}^1=\tilde{\theta},Z^1_{\tilde{\theta}}=z^1\}$ and $cY^2(\tilde{\theta})$ are independent, the density function of

$Y^*(\tilde{\theta})|\{\hat{\theta}^1=\tilde{\theta},Z^1_{\tilde{\theta}}=z^1\}$ is equal to

$$\frac{\sum_{k=1}^K\int_{\ell_k(z^1)}^{u_k(z^1)} f_N\left((t-\mu_Y(\tilde{\theta}))/\sqrt{\Sigma_Y(\tilde{\theta})}\right) f_N\left((y^*-t-c\mu_Y(\tilde{\theta}))/\sqrt{c\Sigma_Y(\tilde{\theta})}\right)dt}{\sqrt{c}\Sigma_Y(\tilde{\theta})\sum_{k=1}^K\left(F_N\left((u_k(z^1)-\mu_Y(\tilde{\theta}))/\sqrt{\Sigma_Y(\tilde{\theta})}\right)-F_N\left((\ell_k(z^1)-\mu_Y(\tilde{\theta}))/\sqrt{\Sigma_Y(\tilde{\theta})}\right)\right)}$$

with corresponding distribution function

$$F^A_{SS}(y^*;\mu_Y(\tilde{\theta}),\tilde{\theta},z^1)$$

$$=\frac{\sum_{k=1}^K\int_{\ell_k(z^1)}^{u_k(z^1)} f_N\left((t-\mu_Y(\tilde{\theta}))/\sqrt{\Sigma_Y(\tilde{\theta})}\right) F_N\left((y^*-t-c\mu_Y(\tilde{\theta}))/\sqrt{c\Sigma_Y(\tilde{\theta})}\right)dt}{\sqrt{\Sigma_Y(\tilde{\theta})}\sum_{k=1}^K\left(F_N\left((u_k(z^1)-\mu_Y(\tilde{\theta}))/\sqrt{\Sigma_Y(\tilde{\theta})}\right)-F_N\left((\ell_k(z^1)-\mu_Y(\tilde{\theta}))/\sqrt{\Sigma_Y(\tilde{\theta})}\right)\right)}$$

$$=\frac{E\left[F_N\left((y^*-\xi^1-c\mu_Y(\tilde{\theta}))/\sqrt{c\Sigma_Y(\tilde{\theta})}\right)1\left(\xi^1\in\bigcup_{k=1}^K[\ell_k(z^1),u_k(z^1)]\right)\right]}{\sum_{k=1}^K\left(F_N\left((u_k(z^1)-\mu_Y(\tilde{\theta}))/\sqrt{\Sigma_Y(\tilde{\theta})}\right)-F_N\left((\ell_k(z^1)-\mu_Y(\tilde{\theta}))/\sqrt{\Sigma_Y(\tilde{\theta})}\right)\right)},$$

where the expectation is taken with respect to $\xi^1\sim N(\mu_Y(\tilde{\theta}),\Sigma_Y(\tilde{\theta}))$. This latter expression for $F^A_{SS}(y^*;\mu_Y(\tilde{\theta}),\tilde{\theta},z^1)$ is very easy to compute by generating normal random variables in standard software packages. This makes the computation of optimal estimators, tests and confidence intervals within the class discussed here computationally straightforward.

Similarly to the optimal case above, the same arguments used to prove Proposition 1 show that the optimal α quantile-unbiased estimator $\hat{\mu}^A_{SS,\alpha}$ in the sample splitting problem that conditions on $\{\hat{\theta}^1=\tilde{\theta}\}$ and the realizations of $Z^1_{\tilde{\theta}}$ and $Z^2_{\tilde{\theta}}$ solves

$$F^A_{SS}(Y^*(\hat{\theta}^1);\hat{\mu}^A_{SS,\alpha},\tilde{\theta},Z^1_{\tilde{\theta}})=1-\alpha.$$

Therefore, our (equal-tailed) alternative split-sample confidence interval is $C^A_{SS}=[\hat{\mu}^A_{SS,\alpha/2},\hat{\mu}^A_{SS,1-\alpha/2}]$. Likewise, the same arguments used to prove Proposition 2 show that the optimal two-sided unbiased test rejects $H_0:\mu_Y(\tilde{\theta})=\mu_{Y,0}$ when

$$Y^*(\tilde{\theta})\notin\left[c_l\left(Z^1_{\tilde{\theta}}\right),c_u\left(Z^1_{\tilde{\theta}}\right)\right],$$

where $c_l(z)$, $c_u(z)$ solve

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

with ζ distributed according to $F_{SS}^A(\cdot;\mu_{Y,0},\tilde{\theta},z)$. These dominating procedures condition on $Z_{\tilde{\theta}}^1$ rather than (X^1,Y^1), and so unlike conventional sample splitting continue to treat (X^1,Y^1) as random for inference.

D Uniformity Results

In this section, we show that the results derived in the main text for the finite-sample normal model translate to uniform asymptotic results over a large class of data generating processes for level-maximization problems. To state and prove these results, it will be important to distinguish between finite-sample and asymptotic objects. To keep this distinction clear, we will subscript finite-sample objects by the sample size, writing X_n, Y_n, $\widehat{\Sigma}_n$, and so on. Moreover, the estimators and confidence intervals $\hat{\mu}_{\alpha,n}$, $\hat{\mu}_{\alpha,n}^H$, $CS_{ET,n}$, $CS_{ET,n}^H$, $CS_{U,n}$, $CS_{U,n}^H$ and $CS_{P,n}$ are equal to their asymptotic counterparts $\hat{\mu}_\alpha$, $\hat{\mu}_\alpha^H$, CS_{ET}, CS_{ET}^H, CS_U, CS_U^H and CS_P after replacing X, Y, Σ with X_n, Y_n, $\widehat{\Sigma}_n$.

With this notation, we aim to prove, for example, that for $\hat{\mu}_{\alpha,n}$ our α-quantile unbiased estimator calculated using $\left(X_n,Y_n,\widehat{\Sigma}_n\right)$, $\mu_{Y,n}(\theta;P)$ the analog of $\mu_Y(\theta)$ in the sample of size n, and data generating process P,

$$\lim_{n\to\infty}\sup_{P\in\mathcal{P}_n}\left|Pr_P\left\{\hat{\mu}_{\alpha,n}\geq\mu_{Y,n}\left(\hat{\theta}_n;P\right)\right\}-\alpha\right|=0,$$

so $\hat{\mu}_{\alpha,n}$ is (unconditionally) asymptotically α-quantile unbiased uniformly over the (possibly sample-size dependent) class of data generating processes $\mathcal{P}_n$. Moreover, we will show that for all $\tilde{\theta}\in\Theta$

$$\lim_{n\to\infty}\sup_{P\in\mathcal{P}_n}\left|Pr_P\left\{\hat{\mu}_{\alpha,n}\geq\mu_{Y,n}\left(\hat{\theta}_n;P\right)|\hat{\theta}_n=\tilde{\theta}\right\}-\alpha\right|Pr_P\left\{\hat{\theta}_n=\tilde{\theta}\right\}=0,$$

so asymptotic quantile unbiasedness also holds conditional on the event $\left\{\hat{\theta}_n=\tilde{\theta}\right\}$ provided this event occurs with non-trivial asymptotic probability. One could use arguments along the same lines as those below to derive results for additional conditioning variables $\hat{\gamma}_n$, but since such arguments would be case-specific, and we do not pursue such an extension here.

Asymptotic uniformity results for conditional inference procedures that, like our corrections, rely on truncated normal distributions were previously established by Tibshirani et al. (2018). Their results cover a class of models that nests our level maximization problem but impose an assumption that implies bounded asymptotic means. Since we do not impose this assumption in our analysis of level-maximization, our results on conditional confidence intervals are not nested by theirs. Moreover, these authors do not cover hybrid inference procedures, which are new to the literature, and also do not provide results for quantile-unbiased estimation. Our proofs are based on subsequencing arguments as in Andrews et al. (2018), though due to the differences in our setting (our interest in conditional inference, and the fact that our target is random from an unconditional perspective) we cannot directly apply their results. In the subsequent analysis, F_N and f_N denote the cdf and pdf of the standard normal distribution.

D.1 Asymptotic Validity for Level Maximization

Section D.1.1 collects the assumptions we use to prove uniform asymptotic validity. Section D.1.2 then states our uniformity results. Section D.1.3 collects a series of technical lemmas which we use to prove our uniformity results. Finally, Sections D.1.4 and D.1.5 collect proofs for the lemmas and the uniformity results, respectively.

D.1.1 Assumptions

To derive our asymptotic uniformity results, we use the fact that all our estimates and confidence intervals are functions of $\left(X_n,Y_n,\widehat{\Sigma}_n\right)$. Hence, to derive our results it suffices to state assumptions in terms of the behavior of these objects.

Assumption 2
Our estimator $\widehat{\Sigma}_n$ is uniformly consistent for some function $\Sigma(P)$,

$$\lim_{n\to\infty}\sup_{P\in\mathcal{P}_n} Pr_P\left\{\left\|\widehat{\Sigma}_n-\Sigma(P)\right\|>\varepsilon\right\}=0$$

for all $\varepsilon>0$.

This assumption requires that our variance estimator $\widehat{\Sigma}_n$ be consistent for some $\Sigma(P)$, which our later assumptions will take to be the asymptotic variance matrix of $(X_n',Y_n')'$ under P, uniformly over $\mathcal{P}_n$.

Assumption 3
There exists a finite $\bar{\lambda}>0$ such that for $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ the minimum and maximum

eigenvalues of a matrix A,

$$1/\bar{\lambda} \leq \lambda_{\min}(\Sigma_X(P)) \leq \lambda_{\max}(\Sigma_X(P)) \leq \bar{\lambda} \textit{ for all } P \in \mathcal{P}_n$$

and

$$1/\bar{\lambda} \leq \Sigma_Y(\theta;P) \leq \bar{\lambda} \textit{ for all } \theta \in \Theta \textit{ and all } P \in \mathcal{P}_n.$$

This assumption bounds the variance matrix $\Sigma_X(P)$ above and away from singularity, and likewise bounds the diagonal elements of $\Sigma_Y(P)$ above and away from zero. This ensures that the set of covariance matrices consistent with $P \in \mathcal{P}_n$ is a subset of a compact set, and that $X_n(\theta)$ has a unique maximum with probability tending to one.

Assumption 4

For BL_1 the class of Lipschitz functions that are bounded in absolute value by one and have Lipschitz constant bounded by one, and $\xi_P \sim N(0,\Sigma(P))$,

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}_n} \sup_{f\in BL_1} \left| E_P\left[f\left(\begin{array}{c} X_n - \mu_{X,n}(P) \\ Y_n - \mu_{Y,n}(P) \end{array} \right) \right] - E[f(\xi_P)] \right| = 0$$

for some sequence of functions $\mu_{X,n}(P)$ and $\mu_{Y,n}(P)$.

Bounded Lipschitz distance metrizes convergence in distribution, so uniform convergence in bounded Lipschitz, as we assume here, is one formalization for uniform convergence in distribution. Hence, this assumption requires that

$$\left(X_n' - \mu_{X,n}(P)', Y_n' - \mu_{Y,n}(P)'\right)'$$

be asymptotically $N(0,\Sigma(P))$ distributed, uniformly over $P \in \mathcal{P}_n$.

D.1.2 Level Maximization Uniformity Results

For $\hat{\theta}_n = \text{argmax}_\theta X_n(\theta)$ we obtain the following results.

Proposition 8

Under Assumptions 2-4, for $\hat{\theta}_n = \text{argmax}_\theta X_n(\theta)$ and $\hat{\mu}_{\alpha,n}$ the α-quantile unbiased estimator,

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}_n} \left| Pr_P\left\{ \hat{\mu}_{\alpha,n} \geq \mu_{Y,n}\left(\hat{\theta}_n;P\right) | \hat{\theta}_n = \tilde{\theta} \right\} - \alpha \right| Pr_P\left\{ \hat{\theta}_n = \tilde{\theta} \right\} = 0, \tag{24}$$

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

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}_n} \left| Pr_P\left\{ \hat{\mu}_{\alpha,n} \geq \mu_{Y,n}\left(\hat{\theta}_n;P\right)\right\} - \alpha \right| = 0. \tag{25}$$

Corollary 1

Under Assumptions 2-4, for $\hat{\theta}_n = \mathrm{argmax}_\theta X_n(\theta)$ and $CS_{ET,n}$ the level $1-\alpha$ equal-tailed confidence interval,

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

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

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

Proposition 9

Under Assumptions 2-4, for $\hat{\theta}_n = \arg\max_\theta X_n(\theta)$ and $CS_{U,n}$ the level $1-\alpha$ unbiased confidence interval,

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}_n} \left| Pr_P\left\{ \mu_{Y,n}\left(\hat{\theta}_n;P\right) \in CS_{U,n} | \hat{\theta}_n = \tilde{\theta}\right\} - (1-\alpha)\right| Pr_P\left\{\hat{\theta}_n = \tilde{\theta}\right\} = 0, \tag{26}$$

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

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}_n} \left| Pr_P\left\{ \mu_{Y,n}\left(\hat{\theta}_n;P\right) \in CS_{U,n}\right\} - (1-\alpha)\right| = 0. \tag{27}$$

Proposition 10

Under Assumptions 2-4, for $\hat{\theta}_n = \arg\max_\theta X_n(\theta)$ and $CS_{P,n}$ the level $1-\alpha$ projection confidence interval,

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}_n} Pr_P\left\{ \mu_{Y,n}\left(\hat{\theta}_n;P\right) \in CS_{P,n}\right\} \geq 1-\alpha. \tag{28}$$

Proposition 11

Under Assumptions 2-4, for $\hat{\theta}_n = \arg\max_\theta X_n(\theta)$, $\hat{\mu}^H_{\alpha,n}$ the α-quantile unbiased hybrid estimator based on initial confidence interval $CS^\beta_{P,n}$, and

$$C^H_n\left(\tilde{\theta};P\right) = 1\left\{\hat{\theta}_n = \tilde{\theta}, \mu_{Y,n}\left(\hat{\theta}_n;P\right) \in CS^\beta_{P,n}\right\},$$

we have

$$\lim_{n\to\infty}\sup_{P\in\mathcal{P}_n}\left|Pr_P\left\{\hat{\mu}^H_{\alpha,n}\geq\mu_{Y,n}\left(\hat{\theta}_n;P\right)|C^H_n\left(\tilde{\theta};P\right)=1\right\}-\alpha\right|E_P\left\{C^H_n\left(\tilde{\theta};P\right)\right\}=0, \tag{29}$$

for all $\tilde{\theta}\in\Theta$. *Moreover*

$$\limsup_{n\to\infty}\sup_{P\in\mathcal{P}_n}\left|Pr_P\left\{\hat{\mu}^H_{\alpha,n}\geq\mu_{Y,n}\left(\hat{\theta}_n;P\right)\right\}-\alpha\right|\leq\max\{\alpha,1-\alpha\}\beta. \tag{30}$$

Corollary 2
Under Assumptions 2-4, for $\hat{\theta}_n=\text{argmax}_\theta X_n(\theta)$ *and* $CS^H_{ET,n}$ *the level* $1-\alpha$ *equal-tailed hybrid confidence set based on initial confidence interval* $CS^\beta_{P,n}$,

$$\lim_{n\to\infty}\sup_{P\in\mathcal{P}_n}\left|Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS^H_{ET,n}|C^H_n\left(\tilde{\theta};P\right)=1\right\}-\frac{1-\alpha}{1-\beta}\right|E_P\left\{C^H_n\left(\tilde{\theta};P\right)\right\}=0, \tag{31}$$

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

$$\liminf_{n\to\infty}\inf_{P\in\mathcal{P}_n}Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS^H_{ET,n}\right\}\geq1-\alpha, \tag{32}$$

and

$$\limsup_{n\to\infty}\sup_{P\in\mathcal{P}_n}Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS^H_{ET,n}\right\}\leq\frac{1-\alpha}{1-\beta}\leq1-\alpha+\beta. \tag{33}$$

Proposition 12
Under Assumptions 2-4, for $\hat{\theta}_n=\text{argmax}_\theta X_n(\theta)$ *and* $CS^H_{U,n}$ *the level* $1-\alpha$ *unbiased hybrid confidence interval based on initial confidence interval* $CS^\beta_{P,n}$,

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

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

$$\liminf_{n\to\infty}\inf_{P\in\mathcal{P}_n}Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS^H_{U,n}\right\}\geq1-\alpha,$$

and

$$\limsup_{n\to\infty}\sup_{P\in\mathcal{P}_n}Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS^H_{U,n}\right\}\leq\frac{1-\alpha}{1-\beta}\leq1-\alpha+\beta.$$

D.1.3 Auxiliary Lemmas

This section collects lemmas that we will use to prove our uniformity results.

Lemma 5
Under Assumption 3, for any sequence of confidence intervals 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. *$\Sigma(P_{n_s})\to\Sigma^*\in\mathcal{S}$ for*

$$\mathcal{S}=\left\{\Sigma:1/\bar{\lambda}\le\lambda_{\min}(\Sigma_X)\le\lambda_{\max}(\Sigma_X)\le\bar{\lambda},1/\bar{\lambda}\le\Sigma_Y(\theta)\le\bar{\lambda}\right\},\tag{34}$$

2. *$Pr_{P_{n_s}}\{C_{n_s}(P_{n_s})=1\}\to p^*\in(0,1]$, and*

3. *$\mu_{X,n_s}(P_{n_s})-\max_\theta\mu_{X,n_s}(\theta;P_{n_s})\to\mu_X^*\in\mathcal{M}_X^*$ for*

$$\mathcal{M}_X^*=\left\{\mu_X\in[-\infty,0]^{|\Theta|}:\max_\theta\mu_X(\theta)=0\right\},$$

we have

$$\lim_{s\to\infty}Pr_{P_{n_s}}\left\{\mu_{Y,n_s}\left(\hat{\theta}_{n_s};P_{n_s}\right)\in CS_{n_s}|C_{n_s}(P_{n_s})=1\right\}=\alpha.\tag{35}$$

Lemma 6

For a collection of sequences of sets $\mathcal{C}_{n,1}(P),\ldots,\mathcal{C}_{n,J}(P)$ and

$$C_{n,j}(P)=1\left\{\left(X_n,Y_n,\widehat{\Sigma}_n\right)\in\mathcal{C}_{n,j}(P)\right\},$$

if

$$\lim_{n\to\infty}\sup_{P\in\mathcal{P}_n}Pr_P\{C_{n,j}(P)=1,C_{n,j'}(P)=1\}=0 \text{ for all } j\neq j'$$

and

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

for all j, then

$$\liminf_{n\to\infty}\inf_{P\in\mathcal{P}_n}Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS_n\right\}\ge(1-\alpha)\cdot\liminf_{n\to\infty}\inf_{P\in\mathcal{P}_n}\sum_j Pr_P\{C_{n,j}(P)=1\}$$

and

$$\limsup_{n\to\infty}\sup_{P\in\mathcal{P}_n} Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS_n\right\}\leq 1-\alpha\liminf_{n\to\infty}\inf_{P\in\mathcal{P}_n}\sum_j Pr_P\{C_{n,j}(P)=1\}.$$

To state the next lemma, define

$$\mathcal{L}\left(\tilde{\theta},Z,\Sigma\right)=\max_{\theta\in\Theta:\Sigma_{XY}\left(\tilde{\theta}\right)>\Sigma_{XY}\left(\tilde{\theta},\theta\right)}\frac{\Sigma_Y\left(\tilde{\theta}\right)\left(Z(\theta)-Z\left(\tilde{\theta}\right)\right)}{\Sigma_{XY}\left(\tilde{\theta}\right)-\Sigma_{XY}\left(\tilde{\theta},\theta\right)} \tag{36}$$

$$\mathcal{U}\left(\tilde{\theta},Z,\Sigma\right)=\min_{\theta\in\Theta:\Sigma_{XY}\left(\tilde{\theta}\right)<\Sigma_{XY}\left(\tilde{\theta},\theta\right)}\frac{\Sigma_Y\left(\tilde{\theta}\right)\left(Z(\theta)-Z\left(\tilde{\theta}\right)\right)}{\Sigma_{XY}\left(\tilde{\theta}\right)-\Sigma_{XY}\left(\tilde{\theta},\theta\right)}, \tag{37}$$

where we define a maximum over the empty set as $-\infty$ and a minimum over the empty set as $+\infty$. For

$$\begin{pmatrix} X_n^* \\ Y_n^* \end{pmatrix}=\begin{pmatrix} X_n-\max_\theta\mu_{X,n}(\theta;P) \\ Y_n-\mu_{Y,n}(P) \end{pmatrix},$$

we next show that using $\left(X_n^*,Y_n^*,\widehat{\Sigma}_n\right)$ in our calculations yields the same bounds $\mathcal{L}$ and $\mathcal{U}$ as using $\left(X_n,Y_n,\widehat{\Sigma}_n\right)$, up to additive shifts

Lemma 7
For $\mathcal{L}\left(\tilde{\theta},Z,\Sigma\right)$ *and* $\mathcal{U}\left(\tilde{\theta},Z,\Sigma\right)$ *as defined in (36) and (37), and*

$$Z_{\tilde{\theta},n}=X_n(\theta)-\frac{\widehat{\Sigma}_{XY,n}\left(\theta,\tilde{\theta}\right)}{\widehat{\Sigma}_{Y,n}\left(\tilde{\theta}\right)}Y_n\left(\tilde{\theta}\right),\quad Z^*_{\tilde{\theta},n}=X_n^*(\theta)-\frac{\widehat{\Sigma}_{XY,n}\left(\theta,\tilde{\theta}\right)}{\widehat{\Sigma}_{Y,n}\left(\tilde{\theta}\right)}Y_n^*\left(\tilde{\theta}\right),$$

we have

$$\mathcal{L}\left(\tilde{\theta},Z^*_{\tilde{\theta},n},\widehat{\Sigma}_n\right)=\mathcal{L}\left(\tilde{\theta},Z_{\tilde{\theta},n},\widehat{\Sigma}_n\right)-\mu_{Y,n}\left(\tilde{\theta};P\right)$$
$$\mathcal{U}\left(\tilde{\theta},Z^*_{\tilde{\theta},n},\widehat{\Sigma}_n\right)=\mathcal{U}\left(\tilde{\theta},Z_{\tilde{\theta},n},\widehat{\Sigma}_n\right)-\mu_{Y,n}\left(\tilde{\theta};P\right).$$

For brevity, going forward we use the shorthand notation

$$\left(\mathcal{L}\left(\tilde{\theta},Z_{\tilde{\theta},n},\widehat{\Sigma}_n\right),\mathcal{U}\left(\tilde{\theta},Z_{\tilde{\theta},n},\widehat{\Sigma}_n\right),\mathcal{L}\left(\tilde{\theta},Z^*_{\tilde{\theta},n},\widehat{\Sigma}_n\right),\mathcal{U}\left(\tilde{\theta},Z^*_{\tilde{\theta},n},\widehat{\Sigma}_n\right)\right)=(\mathcal{L}_n,\mathcal{U}_n,\mathcal{L}_n^*,\mathcal{U}_n^*).$$

Lemma 8

Under Assumptions 2 and 4, for any $\{n_s\}$ and $\{P_{n_s}\}$ satisfying conditions (1)-(3) of Lemma 5 and any $\tilde{\theta}$ with $\mu_X^*\left(\tilde{\theta}\right) > -\infty$,

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

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

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

with $\mu^* = (\mu_X^{*\prime}, 0')'$.

Lemma 9

For F_N again the standard normal distribution function, the function

$$F_{TN}(Y(\theta); \mu, \Sigma_Y(\theta), \mathcal{L}, \mathcal{U}) = \frac{F_N\left(\frac{Y(\theta)\wedge\mathcal{U}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right) - F_N\left(\frac{\mathcal{L}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)}{F_N\left(\frac{\mathcal{U}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right) - F_N\left(\frac{\mathcal{L}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)} 1(Y(\theta) \geq \mathcal{L}) \tag{38}$$

is continuous in $(Y(\theta), \mu, \Sigma_Y(\theta), \mathcal{L}, \mathcal{U})$ on the set

$$\left\{(Y(\theta), \mu, \Sigma_Y(\theta)) \in \mathbb{R}^3, \mathcal{L} \in \mathbb{R} \cup \{-\infty\}, \mathcal{U} \in \mathbb{R} \cup \{\infty\} : \Sigma_Y(\theta) > 0, \mathcal{L} < Y(\theta) < \mathcal{U}\right\}.$$

To state the next lemma, let $(c_l(\mu, \Sigma_Y(\theta), \mathcal{L}, \mathcal{U}), c_u(\mu, \Sigma_Y(\theta), \mathcal{L}, \mathcal{U}))$ solve

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

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

for

$$\zeta \sim \xi | \xi \in [\mathcal{L}, \mathcal{U}], \xi \sim N(\mu, \Sigma_Y(\theta)).$$

Lemma 10

The function $(c_l(\mu, \Sigma_Y(\theta), \mathcal{L}, \mathcal{U}), c_u(\mu, \Sigma_Y(\theta), \mathcal{L}, \mathcal{U}))$ satisfies

$$\begin{gathered}(c_l(\mu, \Sigma_Y(\theta), \mathcal{L}, \mathcal{U}), c_u(\mu, \Sigma_Y(\theta), \mathcal{L}, \mathcal{U})) \\ = (\mu + c_l(0, \Sigma_Y(\theta), \mathcal{L}-\mu, \mathcal{U}-\mu), \mu + c_u(0, \Sigma_Y(\theta), \mathcal{L}-\mu, \mathcal{U}-\mu))\end{gathered}$$

and is continuous in $(\mu,\Sigma_Y(\theta),\mathcal{L},\mathcal{U})$ *on the set*

$$\left\{(\mu,\Sigma_Y(\theta))\in\mathbb{R}^2,\mathcal{L}\in\mathbb{R}\cup\{-\infty\},\mathcal{U}\in\mathbb{R}\cup\{\infty\}:\Sigma_Y(\theta)>0,\mathcal{L}<\mathcal{U}\right\}.$$

D.1.4 Proofs for Auxiliary Lemmas

Proof of Lemma 5 To prove 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

$$\liminf_{n\to\infty}\inf_{P\in\mathcal{P}_n}\left(Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS_n|C_n(P)=1\right\}-\alpha\right)Pr_P\{C_n(P)=1\}\geq 0 \tag{39}$$

and

$$\limsup_{n\to\infty}\sup_{P\in\mathcal{P}_n}\left(Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS_n|C_n(P)=1\right\}-\alpha\right)Pr_P\{C_n(P)=1\}\leq 0. \tag{40}$$

We prove that to show (39), it suffices to show that for all $\{n_s\}$, $\{P_{n_s}\}$ satisfying conditions (1)-(3) of the lemma,

$$\liminf_{s\to\infty}Pr_{P_{n_s}}\left\{\mu_{Y,n_s}\left(\hat{\theta}_{n_s};P_{n_s}\right)\in CS_{n_s}|C_{n_s}(P_{n_s})=1\right\}\geq\alpha. \tag{41}$$

An argument along the same lines implies that to prove (40) it suffices to show that

$$\limsup_{s\to\infty}Pr_{P_{n_s}}\left\{\mu_{Y,n_s}\left(\hat{\theta}_{n_s};P_{n_s}\right)\in CS_{n_s}|C_{n_s}(P_{n_s})=1\right\}\leq\alpha. \tag{42}$$

Note, however, that (41) and (42) together are equivalent to (35).

Towards contradiction, suppose that (39) fails, so

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

for some $\varepsilon>0$ but that (41) holds for all sequences satisfying conditions (1)-(3) of the lemma. Then there exists an increasing sequence of sample sizes n_q and some sequence $\left\{P_{n_q}\right\}$ with $P_{n_q}\in\mathcal{P}_{n_q}$ for all q such that

$$\limsup_{q\to\infty}\left(Pr_{P_{n_q}}\left\{\mu_{Y,n_q}\left(\hat{\theta}_{n_q};P_{n_q}\right)\in CS_{n_q}|C_{n_q}(P_{n_q})=1\right\}-\alpha\right)Pr_{P_{n_q}}\left\{C_{n_q}(P_{n_q})=1\right\}<-\varepsilon. \tag{43}$$

We want to show that there exists a further subsequence $\{n_s\}\subseteq\{n_q\}$ satisfying (1)-(3) in the statement of the lemma, and so establish a contradiction.

Note that since the set $\mathcal{S}$ defined in (34) is compact (e.g. in the Frobenius norm), and Assumption 3 implies that $\Sigma\left(P_{n_q}\right)\in\mathcal{S}$ for all q, there exists a further subsequence $\{n_r\}\subseteq\{n_q\}$ such that

$$\lim_{r\to\infty}\Sigma(P_{n_r})\to\Sigma^*$$

for some $\Sigma^*\in\mathcal{S}$.

Note, next, that $Pr_{P_{n_r}}\{C_{n_r}(P_{n_r})=1\}\in[0,1]$ for all r, and so converges along a subsequence $\{n_t\}\subseteq\{n_r\}$. However, (43) implies that $Pr_{P_{n_r}}\{C_{n_r}(P_{n_r})=1\}\geq\frac{\varepsilon}{\alpha}$ for all r, and thus that

$$Pr_{P_{n_t}}\{C_{n_t}(P_{n_t})=1\}\to p^*\in\left[\frac{\varepsilon}{\alpha},1\right].$$

Finally, let us define

$$\mu^*_{X,n}(P)=\mu_{X,n}(P)-\max_{\theta}\mu_{X,n}(\theta;P),$$

and note that $\mu^*_{X,n}(P)\leq 0$ by construction. Since $\mu^*_{X,n}(P)$ is finite-dimensional and $\max_\theta\mu^*_{X,n}(P;\theta)=0$, there exists some $\theta\in\Theta$ such that $\mu^*_{X,n}(P;\theta)$ is equal to zero infinitely often. Let $\{n_u\}\subseteq\{n_t\}$ extract the corresponding sequence of sample sizes. The set $[-\infty,0]^{|\Theta|}$ is compact under the metric $d(\mu_X,\tilde{\mu}_X)=\|F_N(\mu_X)-F_N(\tilde{\mu}_X)\|$ for $F_N(\cdot)$ the standard normal cdf applied elementwise, and $\|\cdot\|$ the Euclidean norm. Hence, there exists a further subsequence $\{n_s\}\subseteq\{n_u\}$ along which $\mu^*_{X,n_s}(P_{n_s})$ converges to a limit in this metric. Note, however, that this means that $\mu^*_{X,n_s}(P_{n_s})$ converges to a limit $\mu^*\in\mathcal{M}^*$ in the usual metric.

Hence, we have shown that there exists a subsequence $\{n_s\}\subseteq\{n_q\}$ that satisfies (1)-(3). By supposition, (41) must hold along this subsequence. Thus,

$$\liminf_{n\to\infty}\left(Pr_{P_{n_s}}\left\{\mu_{Y,n_s}\left(\hat{\theta}_{n_s};P_{n_s}\right)\in CS_{n_s}|C_{n_s}(P_{n_s})=1\right\}-\alpha\right)Pr_P\{C_{n_s}(P_{n_s})=1\}\geq 0,$$

which contradicts (43). Hence, we have established a contradiction and so proved that (41) for all subsequences satisfying conditions (1)-(3) of the lemma implies (39). An argument along the same lines shows that (42) along all subsequences satisfying conditions (1)-(3) of lemma implies (40). □

Proof of Lemma 6 Let us define

$$C_{n,J+1}(P)=1\{C_{n,j}(P)=0\text{ for all }j\in\{1,...,J\}\}.$$

Note that

$$Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS_n\right\}$$
$$=\sum_{j=1}^{J+1} Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS_n|C_{n,j}(P)=1\right\}Pr_P\{C_{n,j}(P)=1\}+o(1)$$

where the $o(1)$ term is negligible uniformly over $P\in\mathcal{P}_n$ as $n\to\infty$. Hence,

$$Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS_n\right\}-(1-\alpha)$$
$$=\sum_{j=1}^{J+1}\left(Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS_n|C_{n,j}(P)=1\right\}-(1-\alpha)\right)Pr_P\{C_{n,j}(P)=1\}+o(1)$$

and

$$\liminf_{n\to\infty}\inf_{P\in\mathcal{P}} Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS_n\right\}-(1-\alpha)$$
$$=\liminf_{n\to\infty}\inf_{P\in\mathcal{P}_n}\sum_{j=1}^{J+1}\left(Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS_n|C_{n,j}(P)=1\right\}-(1-\alpha)\right)Pr_P\{C_{n,j}(P)=1\}$$
$$=\liminf_{n\to\infty}\inf_{P\in\mathcal{P}_n}\left(Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS_n|C_{n,J+1}(P)=1\right\}-(1-\alpha)\right)Pr_P\{C_{n,J+1}(P)=1\}$$
$$\geq-(1-\alpha)\limsup_{n\to\infty}\sup_{P\in\mathcal{P}_n} Pr_P\{C_{n,J+1}(P)=1\}$$
$$=-(1-\alpha)\left(1-\liminf_{n\to\infty}\inf_{P\in\mathcal{P}_n}\sum_{j=1}^{J} Pr_P\{C_{n,j}(P)=1\}\right)$$

which immediately implies that

$$\liminf_{n\to\infty}\inf_{P\in\mathcal{P}_n} Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS_n\right\}\geq(1-\alpha)\liminf_{n\to\infty}\inf_{P\in\mathcal{P}_n}\sum_{j=1}^{J} Pr_P\{C_{n,j}(P)=1\}.$$

Likewise,

$$\limsup_{n\to\infty}\sup_{P\in\mathcal{P}_n} Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS_n\right\}-(1-\alpha)$$
$$=\limsup_{n\to\infty}\sup_{P\in\mathcal{P}_n}\sum_{j=1}^{J+1}\left(Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS_n|C_{n,j}(P)=1\right\}-(1-\alpha)\right)Pr_P\{C_{n,j}(P)=1\}$$
$$=\limsup_{n\to\infty}\sup_{P\in\mathcal{P}_n}\left(Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS_n|C_{n,J+1}(P)=1\right\}-(1-\alpha)\right)Pr_P\{C_{n,J+1}(P)=1\}$$

$$\leq \alpha \cdot \limsup_{n\to\infty} \sup_{P\in\mathcal{P}_n} Pr_P\{C_{n,J+1}(P)=1\} = \alpha\left(1-\liminf_{n\to\infty} \inf_{P\in\mathcal{P}_n} \sum_{j=1}^{J} Pr_P\{C_{n,j}(P)=1\}\right).$$

This immediately implies that

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_n} Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS_n\right\} \leq 1-\alpha\cdot \liminf_{n\to\infty} \inf_{P\in\mathcal{P}_n} \sum_{j=1}^{J} Pr_P\{C_{n,j}(P)=1\},$$

as we wanted to show. □

Proof of Lemma 7 Note that

$$Z^*_{\tilde{\theta},n} = Z_{\tilde{\theta},n} - \max_{\theta} \mu_{X,n}(\theta;P) + \widehat{\Sigma}_{XY,n}\left(\cdot,\tilde{\theta}\right)\frac{\mu_{Y,n}\left(\tilde{\theta};P\right)}{\widehat{\Sigma}_{Y,n}\left(\tilde{\theta}\right)},$$

so

$$Z^*_{\tilde{\theta},n}(\theta) - Z^*_{\tilde{\theta},n}\left(\tilde{\theta}\right) = Z_{\tilde{\theta},n}(\theta) - Z_{\tilde{\theta},n}\left(\tilde{\theta}\right) + \left(\widehat{\Sigma}_{XY,n}\left(\theta,\tilde{\theta}\right) - \widehat{\Sigma}_{XY,n}\left(\tilde{\theta}\right)\right)\frac{\mu_{Y,n}\left(\tilde{\theta};P\right)}{\widehat{\Sigma}_{Y,n}\left(\tilde{\theta}\right)}.$$

The result follows immediately. □

Proof of Lemma 8 By Assumption 4

$$\begin{pmatrix} X_{n_s} - \mu_{X,n_s}(P_{n_s}) \\ Y_{n_s} - \mu_{Y,n_s}(P_{n_s}) \end{pmatrix} \to_d N(0,\Sigma^*).$$

Hence, by Slutsky's lemma

$$\begin{pmatrix} X^*_{n_s} \\ Y^*_{n_s} \end{pmatrix} = \begin{pmatrix} X_{n_s} - \max_\theta \mu_{X,n_s}(\theta;P_{n_s}) \\ Y_{n_s} - \mu_{Y,n_s}(P_{n_s}) \end{pmatrix} \to_d \begin{pmatrix} X^* \\ Y^* \end{pmatrix} \sim N(\mu^*,\Sigma^*).$$

We begin by considering one $\theta\in\Theta\setminus\left\{\tilde{\theta}\right\}$ at a time. Since $\widehat{\Sigma}_{n_s} \to_p \Sigma^*$ by Assumption 2, if $\Sigma^*_{XY}\left(\tilde{\theta}\right) - \Sigma^*_{XY}\left(\tilde{\theta},\theta\right) \neq 0$ then

$$\frac{\widehat{\Sigma}_{Y,n_s}\left(\tilde{\theta}\right)\left(Z^*_{\tilde{\theta},n_s}(\theta) - Z^*_{\tilde{\theta},n_s}\left(\tilde{\theta}\right)\right)}{\widehat{\Sigma}_{XY,n_s}\left(\tilde{\theta}\right) - \widehat{\Sigma}_{XY,n_s}\left(\tilde{\theta},\theta\right)} \to_d \frac{\Sigma^*_Y\left(\tilde{\theta}\right)\left(Z^*_{\tilde{\theta}}(\theta) - Z^*_{\tilde{\theta}}\left(\tilde{\theta}\right)\right)}{\Sigma^*_{XY}\left(\tilde{\theta}\right) - \Sigma^*_{XY}\left(\tilde{\theta},\theta\right)},$$

where the terms on the right hand side are based on (X^*,Y^*,Σ^*). The limit is finite if $\mu_X^*(\theta)>-\infty$, while otherwise $\mu_X^*(\theta)=-\infty$ and

$$\frac{\Sigma_Y^*\left(\tilde{\theta}\right)\left(Z_{\tilde{\theta}}^*(\theta)-Z_{\tilde{\theta}}^*\left(\tilde{\theta}\right)\right)}{\Sigma_{XY}^*\left(\tilde{\theta}\right)-\Sigma_{XY}^*\left(\tilde{\theta},\theta\right)}=\begin{cases}-\infty & \text{if } \Sigma_{XY}^*\left(\tilde{\theta}\right)-\Sigma_{XY}^*\left(\tilde{\theta},\theta\right)>0\\ +\infty & \text{if } \Sigma_{XY}^*\left(\tilde{\theta}\right)-\Sigma_{XY}^*\left(\tilde{\theta},\theta\right)<0\end{cases}.$$

If instead $\Sigma_{XY}^*\left(\tilde{\theta}\right)-\Sigma_{XY}^*\left(\tilde{\theta},\theta\right)=0$, then since Σ_X^* has full rank,

$$Z_{\tilde{\theta}}^*(\theta)-Z_{\tilde{\theta}}^*\left(\tilde{\theta}\right)=X^*(\theta)-X^*\left(\tilde{\theta}\right)$$

is normally distributed with non-zero variance. Hence, in this case

$$\left|\frac{\widehat{\Sigma}_{Y,n_s}\left(\tilde{\theta}\right)\left(Z_{n_s,\tilde{\theta}}^*(\theta)-Z_{n_s,\tilde{\theta}}^*\left(\tilde{\theta}\right)\right)}{\widehat{\Sigma}_{XY,n_s}\left(\tilde{\theta}\right)-\widehat{\Sigma}_{XY,n_s}\left(\tilde{\theta},\theta\right)}\right|\to\infty. \tag{44}$$

Let us define

$$\Theta^*\left(\tilde{\theta}\right)=\left\{\theta\in\Theta\backslash\tilde{\theta}:\Sigma_{XY}^*\left(\tilde{\theta}\right)-\Sigma_{XY}^*\left(\tilde{\theta},\theta\right)\neq0\right\}.$$

The argument above implies that

$$\max_{\theta\in\Theta^*\left(\tilde{\theta}\right):\widehat{\Sigma}_{XY,n_s}\left(\tilde{\theta}\right)>\widehat{\Sigma}_{XY,n_s}\left(\tilde{\theta},\theta\right)}\frac{\widehat{\Sigma}_{Y,n_s}\left(\tilde{\theta}\right)\left(Z_{\tilde{\theta},n_s}^*(\theta)-Z_{\tilde{\theta},n_s}^*\left(\tilde{\theta}\right)\right)}{\widehat{\Sigma}_{XY,n_s}\left(\tilde{\theta}\right)-\widehat{\Sigma}_{XY,n_s}\left(\tilde{\theta},\theta\right)}$$

$$\to_d\mathcal{L}^*=\max_{\theta\in\Theta:\Sigma_{XY}^*\left(\tilde{\theta}\right)>\Sigma_{XY}^*\left(\tilde{\theta},\theta\right)}\frac{\Sigma_Y^*\left(\tilde{\theta}\right)\left(Z_{\tilde{\theta}}^*(\theta)-Z_{\tilde{\theta}}^*\left(\tilde{\theta}\right)\right)}{\Sigma_{XY}^*\left(\tilde{\theta}\right)-\Sigma_{XY}^*\left(\tilde{\theta},\theta\right)},$$

and

$$\min_{\theta\in\Theta^*\left(\tilde{\theta}\right):\widehat{\Sigma}_{XY,n_s}\left(\tilde{\theta}\right)<\widehat{\Sigma}_{XY,n_s}\left(\tilde{\theta},\theta\right)}\frac{\widehat{\Sigma}_{Y,n_s}\left(\tilde{\theta}\right)\left(Z_{\tilde{\theta},n_s}^*(\theta)-Z_{\tilde{\theta},n_s}^*\left(\tilde{\theta}\right)\right)}{\widehat{\Sigma}_{XY,n_s}\left(\tilde{\theta}\right)-\widehat{\Sigma}_{XY,n_s}\left(\tilde{\theta},\theta\right)}$$

$$\to_d\mathcal{U}^*=\min_{\theta\in\Theta:\Sigma_{XY}^*\left(\tilde{\theta}\right)<\Sigma_{XY}^*\left(\tilde{\theta},\theta\right)}\frac{\Sigma_Y^*\left(\tilde{\theta}\right)\left(Z_{\tilde{\theta}}^*(\theta)-Z_{\tilde{\theta}}^*\left(\tilde{\theta}\right)\right)}{\Sigma_{XY}^*\left(\tilde{\theta}\right)-\Sigma_{XY}^*\left(\tilde{\theta},\theta\right)}.$$

By (44), the same convergence holds when we minimize and maximize over Θ rather than

$\Theta^*(\tilde{\theta})$. Hence,

$$\left(\mathcal{L}_{n_s}^*, \mathcal{U}_{n_s}^*\right) \to_d \left(\mathcal{L}^*, \mathcal{U}^*\right).$$

Moreover, $\hat{\theta}_{n_s}$ is almost everywhere continuous in $X_{n_s}^*$, so

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

by the continuous mapping theorem, and this convergence holds jointly with that for $\left(\mathcal{L}_{n_s}^*, \mathcal{U}_{n_s}^*\right)$. Hence, we have established the desired convergence. $\square$

Proof of Lemma 9 Continuity for $\Sigma_Y(\theta) > 0, \mathcal{L} < Y(\theta) < \mathcal{U}$ with all elements finite is immediate from the functional form. Moreover, for fixed $(Y(\theta), \mu, \Sigma_Y(\theta)) \in \mathbb{R}^3$ with $\Sigma_Y(\theta) > 0$ and $\mathcal{L} < Y(\theta) < \mathcal{U}$,

$$\lim_{\mathcal{U}\to\infty} \frac{F_N\left(\frac{Y(\theta)\wedge\mathcal{U}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right) - F_N\left(\frac{\mathcal{L}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)}{F_N\left(\frac{\mathcal{U}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right) - F_N\left(\frac{\mathcal{L}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)} 1(Y(\theta) \geq \mathcal{L}) = \frac{F_N\left(\frac{Y(\theta)-\mu}{\sqrt{\Sigma_Y(\theta)}}\right) - F_N\left(\frac{\mathcal{L}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)}{F_N\left(\frac{\infty}{\sqrt{\Sigma_Y(\theta)}}\right) - F_N\left(\frac{\mathcal{L}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)}$$

$$\lim_{\mathcal{L}\to-\infty} \frac{F_N\left(\frac{Y(\theta)\wedge\mathcal{U}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right) - F_N\left(\frac{\mathcal{L}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)}{F_N\left(\frac{\mathcal{U}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right) - F_N\left(\frac{\mathcal{L}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)} 1(Y(\theta) \geq \mathcal{L}) = \frac{F_N\left(\frac{Y(\theta)-\mu}{\sqrt{\Sigma_Y(\theta)}}\right) - F_N\left(\frac{-\infty}{\sqrt{\Sigma_Y(\theta)}}\right)}{F_N\left(\frac{\mathcal{U}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right) - F_N\left(\frac{-\infty}{\sqrt{\Sigma_Y(\theta)}}\right)}$$

and

$$\lim_{(\mathcal{L},\mathcal{U})\to(-\infty,\infty)} \frac{F_N\left(\frac{Y(\theta)\wedge\mathcal{U}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right) - F_N\left(\frac{\mathcal{L}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)}{F_N\left(\frac{\mathcal{U}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right) - F_N\left(\frac{\mathcal{L}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)} 1(Y(\theta) \geq \mathcal{L}) = \frac{F_N\left(\frac{Y(\theta)-\mu}{\sqrt{\Sigma_Y(\theta)}}\right) - F_N\left(\frac{-\infty}{\sqrt{\Sigma_Y(\theta)}}\right)}{F_N\left(\frac{\infty}{\sqrt{\Sigma_Y(\theta)}}\right) - F_N\left(\frac{-\infty}{\sqrt{\Sigma_Y(\theta)}}\right)}.$$

Hence, we obtain the desired result. $\square$

Proof of Lemma 10 Note that for f_N again the standard normal density,

$$Pr\{\zeta \in [c_l, c_u]\} = \frac{F_N\left(\frac{\mathcal{U}\wedge c_u-\mu}{\sqrt{\Sigma_Y(\theta)}}\right) - F_N\left(\frac{\mathcal{L}\vee c_l-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)}{F_N\left(\frac{\mathcal{U}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right) - F_N\left(\frac{\mathcal{L}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)} 1(\mathcal{U} \geq c_l, c_u \geq \mathcal{L}),$$

$$E[\zeta 1\{\zeta\in[c_l,c_u]\}]=Pr\{\zeta\in[c_l,c_u]\}\left[\mu+\frac{\sqrt{\Sigma_Y(\theta)}\left(f_N\left(\frac{\mathcal{L}\vee c_l-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-f_N\left(\frac{\mathcal{U}\wedge c_u-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)\right)}{F_N\left(\frac{\mathcal{U}\wedge c_u-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-F\left(\frac{\mathcal{L}\vee c_l-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)}\right]$$

$$=\frac{\mu\left(F_N\left(\frac{\mathcal{U}\wedge c_u-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-F_N\left(\frac{\mathcal{L}\vee c_l-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)\right)+\sqrt{\Sigma_Y(\theta)}\left(f_N\left(\frac{\mathcal{L}\vee c_l-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-f_N\left(\frac{\mathcal{U}\wedge c_u-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)\right)}{F_N\left(\frac{\mathcal{U}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-F_N\left(\frac{\mathcal{L}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)}$$

and

$$E[\zeta]=\mu+\frac{\sqrt{\Sigma_Y(\theta)}\left(f_N\left(\frac{\mathcal{L}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-f_N\left(\frac{\mathcal{U}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)\right)}{F_N\left(\frac{\mathcal{U}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-F_N\left(\frac{\mathcal{L}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)}.$$

Thus, we can write $(c_l(\mu,\Sigma_Y(\theta),\mathcal{L},\mathcal{U}),c_u(\mu,\Sigma_Y(\theta),\mathcal{L},\mathcal{U}))$ as the solution to the following system of equations:

$$F_N\left(\frac{\mathcal{U}\wedge c_u-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-F_N\left(\frac{\mathcal{L}\vee c_l-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-(1-\alpha)\left(F_N\left(\frac{\mathcal{U}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-F_N\left(\frac{\mathcal{L}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)\right)=0 \quad (45)$$

and

$$\mu\left(F_N\left(\frac{\mathcal{U}\wedge c_u-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-F_N\left(\frac{\mathcal{L}\vee c_l-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)\right)+\sqrt{\Sigma_Y(\theta)}\left(f_N\left(\frac{\mathcal{L}\vee c_l-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-f_N\left(\frac{\mathcal{U}\wedge c_u-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)\right)$$

$$-(1-\alpha)\mu\left(F_N\left(\frac{\mathcal{U}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-F_N\left(\frac{\mathcal{L}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)\right)$$

$$-(1-\alpha)\sqrt{\Sigma_Y(\theta)}\left(f_N\left(\frac{\mathcal{L}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-f_N\left(\frac{\mathcal{U}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)\right)=0$$

such that $c_l\leq\mathcal{U}$ and $c_u\geq\mathcal{L}$. Note, however, that since any $c=(c_l,c_u)$ that solves this system must satisfy (45), we can also write

$$(c_l(\mu,\Sigma_Y(\theta),\mathcal{L},\mathcal{U}),c_u(\mu,\Sigma_Y(\theta),\mathcal{L},\mathcal{U}))$$

as the solution to

$$g\left(c;\mu,\sqrt{\Sigma_Y(\theta)},\mathcal{L},\mathcal{U}\right)=0$$

such that $c_l \leq \mathcal{U}$ and $c_u \geq \mathcal{L}$, for

$$g\left(c;\mu,\sqrt{\Sigma_Y(\theta)},\mathcal{L},\mathcal{U}\right)$$
$$=\left(\begin{array}{c} F_N\left(\frac{\mathcal{U}\wedge c_u-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-F_N\left(\frac{\mathcal{L}\vee c_l-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-(1-\alpha)\left(F_N\left(\frac{\mathcal{U}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-F_N\left(\frac{\mathcal{L}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)\right) \\ f_N\left(\frac{\mathcal{L}\vee c_l-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-f_N\left(\frac{\mathcal{U}\wedge c_u-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-(1-\alpha)\left(f_N\left(\frac{\mathcal{L}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-f_N\left(\frac{\mathcal{U}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)\right)\end{array}\right).$$

This implies that

$$g\left(c;\mu,\sqrt{\Sigma_Y(\theta)},\mathcal{L},\mathcal{U}\right)=g\left(c-(\mu,\mu)';0,\sqrt{\Sigma_Y(\theta)},\mathcal{L}-\mu,\mathcal{U}-\mu\right),$$

from which the first result of the lemma follows immediately.

To prove the second part of the lemma, note that by the first part of the lemma it suffices to prove continuity of

$$(c_l(0,\Sigma_Y(\theta),\mathcal{L},\mathcal{U}),c_u(0,\Sigma_Y(\theta),\mathcal{L},\mathcal{U})). \tag{46}$$

Recall that (46) solves

$$Pr\{\zeta\in[c_l,c_u]\}=(1-\alpha) \tag{47}$$

and

$$E[\zeta 1\{\zeta\in[c_l,c_u]\}]=(1-\alpha)E[\zeta] \tag{48}$$

for $\zeta \sim \xi|\xi \in [\mathcal{L},\mathcal{U}]$ where $\xi \sim N(0,\Sigma_Y(\theta))$. Note, however, that since $\mathcal{L} < \mathcal{U}$, (47) implies that any solution has $c_l < c_u$, and that we cannot have both $c_l \leq \mathcal{L}$ and $c_u \geq \mathcal{U}$. Note, next, that if $c_l = \mathcal{L}$, then since $c_u < \mathcal{U}$, $E[\zeta|\zeta \in [c_l, c_u]] < E[\zeta]$, and thus that $E[\zeta 1\{\zeta\in[c_l,c_u]\}]<(1-\alpha)E[\zeta]$. Since the same argument applies when $c_u=\mathcal{U}$, we see that for any solution (46), $\mathcal{L}<c_l<c_u<\mathcal{U}$.

Note, next, that $g\left(c;0,\sqrt{\Sigma_Y(\theta)},\mathcal{L},\mathcal{U}\right)$ is almost everywhere differentiable with respect to c with derivative

$$\frac{\partial}{\partial c'}g\left(c;0,\sqrt{\Sigma_Y(\theta)},\mathcal{L},\mathcal{U}\right)=$$
$$\left(\begin{array}{cc} -1(c_l>\mathcal{L})f_N\left(c_l/\sqrt{\Sigma_Y(\theta)}\right)/\sqrt{\Sigma_Y(\theta)} & 1(c_u<\mathcal{U})f_N\left(c_u/\sqrt{\Sigma_Y(\theta)}\right)/\sqrt{\Sigma_Y(\theta)} \\ -1(c_l>\mathcal{L})c_l f_N\left(c_l/\sqrt{\Sigma_Y(\theta)}\right)/\Sigma_Y(\theta) & 1(c_u<\mathcal{U})c_u f_N\left(c_u/\sqrt{\Sigma_Y(\theta)}\right)/\Sigma_Y(\theta)\end{array}\right).$$

The first row is zero if and only if $c_l<\mathcal{L}$ and $c_u>\mathcal{U}$, which as argued above cannot be a solution to $g\left(c;0,\sqrt{\Sigma_Y(\theta)},\mathcal{L},\mathcal{U}\right)=0$ for $\mathcal{L}<\mathcal{U}$ finite. The second row is zero if and only if either (i) $c_l < \mathcal{L}$ and $c_u > \mathcal{U}$ or (ii) $c_l = c_u = 0$, which again cannot be a solution. Finally, apart from the cases just mentioned, the rows are proportional if and only if either (i) $c_l < \mathcal{L}$, (ii) $c_u > \mathcal{U}$ or (iii) $c_l = c_u$, none of which can be a solution. Hence, the implicit function theorem implies continuity on

$$\{\Sigma_Y(\theta) \in \mathbb{R}, \mathcal{L} \in \mathbb{R}, \mathcal{U} \in \mathbb{R} : \Sigma_Y(\theta) > 0, \mathcal{L} < \mathcal{U}\}.$$

To complete the proof, we need to establish continuity at infinity. Note, however, that we can write

$$g\left(c;0,\sqrt{\Sigma_Y(\theta)},\mathcal{L},\mathcal{U}\right) = \tilde{g}(c;0,\Sigma_Y(\theta),F_N(\mathcal{L}),F_N(\mathcal{U}))$$

where $\tilde{g}$ is continuous in all arguments and $F_N(\cdot)$ is continuous at infinity. Hence, another application of implicit function theorem implies that

$$(c_l(0,\Sigma_Y(\theta),\mathcal{L},\mathcal{U}), c_u(0,\Sigma_Y(\theta),\mathcal{L},\mathcal{U}))$$

are continuous on

$$\left\{\Sigma_Y(\theta) > 0, \mathcal{L} < \mathcal{U} : (\Sigma_Y(\theta), Y(\theta)) \in \mathbb{R}^2, \mathcal{L} \in \mathbb{R} \cup \{-\infty\}, \mathcal{U} \in \mathbb{R} \cup \{\infty\}\right\},$$

as we wanted to show. $\square$

D.1.5 Proofs for Uniformity Results

Proof of Proposition 8 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 Lemma 5, to prove that (24) holds it suffices to show that for all $\{n_s\}$ and $\{P_{n_s}\}$ such that conditions (1)-(3) of the lemma hold with $C_n(P) = 1\left\{\hat{\theta}_n = \tilde{\theta}\right\}$, we have

$$\lim_{s\to\infty} Pr_{P_{n_s}}\left\{\hat{\mu}_{Y,n_s}\left(\hat{\theta}_{n_s};P_{n_s}\right) \in CS_{U,-,n_s} | \hat{\theta}_{n_s} = \tilde{\theta}\right\} = \alpha. \quad (49)$$

To this end, recall that for $F_{TN}(Y(\theta);\mu,\Sigma_Y(\theta),\mathcal{L},\mathcal{U})$ as defined in (38), the estimator $\hat{\mu}_{\alpha,n}$ solves

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

where $(\mathcal{L}_n,\mathcal{U}_n)$ are defined following Lemma 7. This cdf is strictly decreasing in μ as argued in the proof of Proposition 7, 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

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

Note, next, that by Lemma 7 and the form of the function F_{TN},

$$F_{TN}\left(Y_n\left(\hat{\theta}_n\right);\mu_{Y,n}\left(\hat{\theta}_n;P\right),\widehat{\Sigma}_{Y,n}\left(\hat{\theta}_n\right),\mathcal{L}_n,\mathcal{U}_n\right)=F_{TN}\left(Y_n^*\left(\hat{\theta}_n\right);0,\widehat{\Sigma}_{Y,n}\left(\hat{\theta}_n\right),\mathcal{L}_n^*,\mathcal{U}_n^*\right),$$

so $\hat{\mu}_{\alpha,n}\geq\mu_{Y,n}\left(\hat{\theta}_n;P\right)$ if and only if

$$F_{TN}\left(Y_n^*\left(\hat{\theta}_n\right);0,\widehat{\Sigma}_{Y,n}\left(\hat{\theta}_n\right),\mathcal{L}_n^*,\mathcal{U}_n^*\right)\geq 1-\alpha.$$

Lemma 8 shows that $\left(Y_n^*\left(\hat{\theta}_{n_s}\right),\widehat{\Sigma}_{Y,n_s}\left(\hat{\theta}_{n_s}\right),\mathcal{L}_{n_s}^*,\mathcal{U}_{n_s}^*,\hat{\theta}_{n_s}\right)$ converges in distribution as $s\to\infty$, so since F_{TN} is continuous by Lemma 9 while $\text{argmax}_\theta X^*(\theta)$ is almost surely unique and continuous for X^* as in Lemma 8, the continuous mapping theorem implies that

$$\begin{aligned}&\left(F_{TN}\left(Y_{n_s}^*\left(\hat{\theta}_{n_s}\right);0,\widehat{\Sigma}_{Y,n_s}\left(\hat{\theta}_{n_s}\right),\mathcal{L}_{n_s}^*,\mathcal{U}_{n_s}^*\right),1\left\{\hat{\theta}_{n_s}=\tilde{\theta}\right\}\right)\\ &\to_d\left(F_{TN}\left(Y^*\left(\hat{\theta}\right);0,\Sigma_Y^*\left(\hat{\theta}\right),\mathcal{L}^*,\mathcal{U}^*\right),1\left\{\hat{\theta}=\tilde{\theta}\right\}\right).\end{aligned}$$

Since we can write

$$Pr_{P_{n_s}}\left\{F_{TN}\left(Y_{n_s}^*\left(\hat{\theta}_{n_s}\right);0,\widehat{\Sigma}_{Y,n_s}\left(\hat{\theta}_{n_s}\right),\mathcal{L}_{n_s}^*,\mathcal{U}_{n_s}^*\right)\geq 1-\alpha|\hat{\theta}_{n_s}=\tilde{\theta}\right\}$$

$$=\frac{E_{P_{n_s}}\left[1\left\{F_{TN}\left(Y_{n_s}^*\left(\hat{\theta}_{n_s}\right);0,\widehat{\Sigma}_{Y,n_s}\left(\hat{\theta}_{n_s}\right),\mathcal{L}_{n_s}^*,\mathcal{U}_{n_s}^*\right)\geq 1-\alpha\right\}1\left\{\hat{\theta}_{n_s}=\tilde{\theta}\right\}\right]}{E_{P_{n_s}}\left[1\left\{\hat{\theta}_{n_s}=\tilde{\theta}\right\}\right]},$$

and by construction (see also Proposition 1 in the main text),

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

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

$$Pr_{P_{n_s}}\left\{F_{TN}\left(Y_{n_s}^*\left(\hat{\theta}_{n_s}\right);0,\widehat{\Sigma}_{Y,n_s}\left(\hat{\theta}_{n_s}\right),\mathcal{L}_{n_s}^*,\mathcal{U}_{n_s}^*\right)\geq 1-\alpha|\hat{\theta}_{n_s}=\tilde{\theta}\right\}$$

$$\to Pr\left\{F_{TN}\left(Y^*\left(\hat{\theta}\right);0,\Sigma_Y^*\left(\hat{\theta}\right),\mathcal{L}^*,\mathcal{U}^*\right)\geq 1-\alpha|\hat{\theta}=\tilde{\theta}\right\}=\alpha,$$

which verifies (49).

Since this argument holds for all $\tilde{\theta}\in\Theta$, and Assumptions 3 and 4 imply that for all $\theta,\tilde{\theta}\in\Theta$ with $\theta\neq\tilde{\theta}$,

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

Lemma 6 implies (25). $\square$

Proof of Corollary 1 By construction, $CS_{ET,n}=\left[\hat{\mu}_{\alpha/2,n},\hat{\mu}_{1-\alpha/2,n}\right]$, and $\hat{\mu}_{1-\alpha/2,n}>\hat{\mu}_{\alpha/2,n}$ for all $\alpha<1$. Hence,

$$Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS_{ET,n}|\hat{\theta}_n=\tilde{\theta}\right\}$$
$$=Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\leq\hat{\mu}_{1-\alpha/2,n}|\hat{\theta}_n=\tilde{\theta}\right\}-Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\leq\hat{\mu}_{\alpha/2,n}|\hat{\theta}_n=\tilde{\theta}\right\},$$

so the result is immediate from Proposition 8 and Lemma 6. $\square$

Proof of Proposition 9 Note that by the definition of $CS_{U,n}$

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

where

$$(c_l(\mu,\Sigma_Y(\theta),\mathcal{L},\mathcal{U}),c_u(\mu,\Sigma_Y(\theta),\mathcal{L},\mathcal{U}))$$

are defined immediately before Lemma 10. Hence, by Lemmas 7 and 10,

$$\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS_{U,n}$$
$$\Longleftrightarrow Y_n^*\left(\hat{\theta}_n\right)\in\left[c_l\left(0,\widehat{\Sigma}_{Y,n}\left(\hat{\theta}_n\right),\mathcal{L}_n^*,\mathcal{U}_n^*\right),c_u\left(0,\widehat{\Sigma}_{Y,n}\left(\hat{\theta}_n\right),\mathcal{L}_n^*,\mathcal{U}_n^*\right)\right].$$

By Lemma 5, to prove that (26) holds it suffices to show that for all $\{n_s\}$ and $\{P_{n_s}\}$ satisfying conditions (1)-(3) of Lemma 5,

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

Thus, it suffices to show that

$$\lim_{s\to\infty} Pr_{P_{n_s}}\left\{ Y^*_{n_s}\left(\hat{\theta}_{n_s}\right) \in \begin{matrix} \left[c_l\left(0,\widehat{\Sigma}_{Y,n_s}\left(\hat{\theta}_{n_s}\right),\mathcal{L}^*_{n_s},\mathcal{U}^*_{n_s}\right),\right. \\ \left. c_u\left(0,\widehat{\Sigma}_{Y,n_s}\left(\hat{\theta}_{n_s}\right),\mathcal{L}^*_{n_s},\mathcal{U}^*_{n_s}\right)\right] \end{matrix} \middle| \hat{\theta}_{n_s}=\tilde{\theta} \right\} = 1-\alpha.$$

To this end, note that by Lemma 8,

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

and thus, by Lemma 10 and the continuous mapping theorem, that

$$\begin{aligned}&\left(Y^*_{n_s}\left(\tilde{\theta}\right),c_l\left(0,\widehat{\Sigma}_{Y,n_s}\left(\tilde{\theta}\right),\mathcal{L}^*_{n_s},\mathcal{U}^*_{n_s}\right),c_u\left(0,\widehat{\Sigma}_{Y,n_s}\left(\tilde{\theta}\right),\mathcal{L}^*_{n_s},\mathcal{U}^*_{n_s}\right),1\left\{\hat{\theta}_{n_s}=\tilde{\theta}\right\}\right)\\ &\to_d \left(Y^*\left(\tilde{\theta}\right),c_l\left(0,\Sigma^*_Y\left(\tilde{\theta}\right),\mathcal{L}^*,\mathcal{U}^*\right),c_u\left(0,\Sigma^*_Y\left(\tilde{\theta}\right),\mathcal{L}^*,\mathcal{U}^*\right),1\left\{\hat{\theta}=\tilde{\theta}\right\}\right).\end{aligned}$$

By construction (see also Proposition 2 in the main text),

$$Pr\left\{Y^*\left(\tilde{\theta}\right)\in\left[c_l\left(0,\mathcal{L}^*,\mathcal{U}^*,\Sigma^*_Y\left(\tilde{\theta}\right)\right),c_u\left(0,\mathcal{L}^*,\mathcal{U}^*,\Sigma^*_Y\left(\tilde{\theta}\right)\right)\right]|\hat{\theta}=\tilde{\theta}\right\}=1-\alpha,$$

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

$$Pr\left\{Y^*\left(\tilde{\theta}\right)=c_l\left(0,\Sigma^*_Y\left(\tilde{\theta}\right),\mathcal{L}^*,\mathcal{U}^*\right)\right\}=Pr\left\{Y^*\left(\tilde{\theta}\right)=c_u\left(0,\Sigma^*_Y\left(\tilde{\theta}\right),\mathcal{L}^*,\mathcal{U}^*\right)\right\}=0.$$

Hence,

$$\begin{aligned}&Pr_{P_{n_s}}\left\{Y^*_{n_s}\left(\hat{\theta}_{n_s}\right)\in\left[c_l\left(0,\widehat{\Sigma}_{Y,n_s}\left(\hat{\theta}_{n_s}\right),\mathcal{L}^*_{n_s},\mathcal{U}^*_{n_s}\right),c_u\left(0,\widehat{\Sigma}_{Y,n_s}\left(\hat{\theta}_{n_s}\right),\mathcal{L}^*_{n_s},\mathcal{U}^*_{n_s}\right)\right]|\hat{\theta}_{n_s}=\tilde{\theta}\right\}\\ &=\frac{E_{P_{n_s}}\left[1\left\{Y^*_{n_s}\left(\hat{\theta}_{n_s}\right)\in\left[c_l\left(0,\widehat{\Sigma}_{Y,n_s}\left(\hat{\theta}_{n_s}\right),\mathcal{L}^*_{n_s},\mathcal{U}^*_{n_s}\right),c_u\left(0,\widehat{\Sigma}_{Y,n_s}\left(\hat{\theta}_{n_s}\right),\mathcal{L}^*_{n_s},\mathcal{U}^*_{n_s}\right)\right]\right\}1\left\{\hat{\theta}_{n_s}=\tilde{\theta}\right\}\right]}{E_{P_{n_s}}\left[1\left\{\hat{\theta}_{n_s}=\tilde{\theta}\right\}\right]}\\ &\to\frac{E\left[1\left\{Y^*\left(\hat{\theta}\right)\in\left[c_l\left(0,\Sigma^*_Y\left(\hat{\theta}\right),\mathcal{L}^*,\mathcal{U}^*\right),c_u\left(0,\Sigma^*_Y\left(\hat{\theta}\right),\mathcal{L}^*,\mathcal{U}^*\right)\right]\right\}1\left\{\hat{\theta}=\tilde{\theta}\right\}\right]}{E\left[1\left\{\hat{\theta}=\tilde{\theta}\right\}\right]}=1-\alpha,\end{aligned}$$

as we wanted to show, so (26) follows by Lemma 5.

Since this result again holds for all $\tilde{\theta}\in\Theta$, (27) follows immediately by the same argument as in the proof of Proposition 8. □

Proof of Proposition 10 By the same argument as in the proof of Lemma 5, to show that (28) holds it suffices to show that for all $\{n_s\}$, $\{P_{n_s}\}$ satisfying conditions (1)-(3) of

Lemma 5,

$$\liminf_{n\to\infty} Pr_{P_{n_s}}\left\{\mu_{Y,n_s}\left(\hat{\theta}_{n_s};P_{n_s}\right)\in CS_{P,n_s}\right\}\geq 1-\alpha.$$

To this end, note that

$$\mu_{Y,n_s}\left(\hat{\theta}_{n_s};P_{n_s}\right)\in CS_{P,n_s}\ \ ifandonlyif\ \ Y^*_{n_s}\left(\hat{\theta}_{n_s}\right)\in\left[-c_\alpha\left(\widehat{\Sigma}_{Y,n_s}\right)\sqrt{\widehat{\Sigma}_Y\left(\hat{\theta}_{n_s}\right)},c_\alpha\left(\widehat{\Sigma}_{Y,n_s}\right)\sqrt{\widehat{\Sigma}_Y\left(\hat{\theta}_{n_s}\right)}\right]$$

for $c_\alpha(\Sigma_Y)$ the $1-\alpha$ quantile of $\max_\theta|\xi(\theta)|/\sqrt{\Sigma_Y(\theta)}$ where $\xi\sim N(0,\Sigma_Y)$. Next, note that $c_\alpha(\Sigma_Y)$ is continuous in Σ on $\mathcal{S}$ as defined in (34). Hence, for all θ, $c_\alpha(\Sigma_Y)\sqrt{\Sigma_Y(\theta)}$ is continuous as well. Assumptions 2 and 4 imply that

$$\left(Y^*_{n_s},\widehat{\Sigma}_{n_s},\hat{\theta}_{n_s}\right)\to_d\left(Y^*,\Sigma^*,\hat{\theta}\right),$$

which by the continuous mapping theorem implies

$$\left(Y^*_{n_s}\left(\hat{\theta}_{n_s}\right),c_\alpha\left(\widehat{\Sigma}_{Y,n_s}\right)\sqrt{\widehat{\Sigma}_Y\left(\hat{\theta}_{n_s}\right)}\right)\to_d\left(Y^*\left(\hat{\theta}\right),c_\alpha(\Sigma^*_Y)\sqrt{\Sigma^*_Y\left(\hat{\theta}\right)}\right).$$

Hence, since $Pr\left\{\left|Y^*\left(\hat{\theta}\right)\right|-c_\alpha(\Sigma^*_Y)\sqrt{\Sigma^*_Y\left(\hat{\theta}\right)}=0\right\}=0$,

$$Pr_{P_{n_s}}\left\{\mu_{Y,n_s}\left(\hat{\theta}_{n_s};P_{n_s}\right)\in CS_{P,n_s}\right\}\to Pr\left\{Y^*\left(\hat{\theta}\right)\in\left[-c_\alpha(\Sigma^*_Y)\sqrt{\Sigma^*_Y\left(\hat{\theta}\right)},c_\alpha(\Sigma^*_Y)\sqrt{\Sigma^*_Y\left(\hat{\theta}\right)}\right]\right\}\tag{50}$$

where the right hand side is at least $1-\alpha$ by construction. □

Proof of Proposition 11 Note that

$$\hat{\mu}^H_{\alpha,n}\geq\mu_{Y,n}\left(\hat{\theta}_n;P\right)$$

if and only if

$$\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS^H_{U,-,n}$$

for $CS^H_{U,-,n}=(-\infty,\hat{\mu}^H_{\alpha,n}]$. Hence, by Lemma 5, to prove that (29) holds it suffices to show that for all $\{n_s\}$ and $\{P_{n_s}\}$ such that conditions (1)-(3) of the lemma hold with $C_n(P)=1\left\{\hat{\theta}_n=\tilde{\theta},\mu_{Y,n}\left(\hat{\theta}_n;P_n\right)\in CS^\beta_{P,n}\right\}$, we have

$$\lim_{s\to\infty}Pr_{P_{n_s}}\left\{\hat{\mu}_{Y,n_s}\left(\hat{\theta}_{n_s};P_{n_s}\right)\in CS^H_{U,-,n}|\hat{\theta}_{n_s}=\tilde{\theta},\mu_{Y,n_s}\left(\hat{\theta}_{n_s};P_{n_s}\right)\in CS^\beta_{P,n_s}\right\}=\alpha.$$

Recall that for $F_{TN}(Y(\theta);\mu,\Sigma_Y(\theta),\mathcal{L},\mathcal{U})$ defined as in (38), $\hat{\mu}^H_{\alpha,n}$ solves

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

for

$$\mathcal{L}^H_n(\mu)=\max\left\{\mathcal{L}_n,\mu-c_\alpha\left(\widehat{\Sigma}_{Y,n}\right)\sqrt{\widehat{\Sigma}_Y\left(\hat{\theta}_n\right)}\right\}$$

$$\mathcal{U}^H_n(\mu)=\min\left\{\mathcal{U}_n,\mu+c_\alpha\left(\widehat{\Sigma}_{Y,n}\right)\sqrt{\widehat{\Sigma}_Y\left(\hat{\theta}_n\right)}\right\}.$$

The proof of Proposition 7 shows that $F_{TN}\left(Y_n\left(\hat{\theta}_n\right);\mu,\widehat{\Sigma}_{Y,n}\left(\hat{\theta}_n\right),\mathcal{L}^H_n(\mu),\mathcal{U}^H_n(\mu)\right)$ is strictly decreasing in μ, so for a given value $\mu_{Y,0}$,

$$\hat{\mu}^H_{\alpha,n}\geq\mu_{Y,0}\Longleftrightarrow F_{TN}\left(Y_n\left(\hat{\theta}_n\right);\mu_{Y,0},\widehat{\Sigma}_{Y,n}\left(\hat{\theta}_n\right),\mathcal{L}^H_n(\mu_{Y,0}),\mathcal{U}^H_n(\mu_{Y,0})\right)\geq1-\alpha.$$

As in the proof of Proposition 8

$$\begin{aligned}F_{TN}&\left(Y_n\left(\hat{\theta}_n\right);\mu_{Y,n}\left(\hat{\theta}_n;P_n\right),\widehat{\Sigma}_{Y,n}\left(\hat{\theta}_n\right),\mathcal{L}^H_n\left(\mu_{Y,n}\left(\hat{\theta}_n;P_n\right)\right),\mathcal{U}^H_n\left(\mu_{Y,n}\left(\hat{\theta}_n;P_n\right)\right)\right)\\&=F_{TN}\left(Y^*_n\left(\hat{\theta}_n\right);0,\widehat{\Sigma}_{Y,n}\left(\hat{\theta}_n\right),\mathcal{L}^{H*}_n,\mathcal{U}^{H*}_n\right),\end{aligned}$$

where $\mathcal{L}^{H*}_n=\max\left\{\mathcal{L}^*_n,-c_\alpha\left(\widehat{\Sigma}_{Y,n}\right)\sqrt{\widehat{\Sigma}_Y\left(\hat{\theta}_n\right)}\right\}$ and $\mathcal{U}^{H*}_n=\min\left\{\mathcal{U}^*_n,c_\alpha\left(\widehat{\Sigma}_{Y,n}\right)\sqrt{\widehat{\Sigma}_Y\left(\hat{\theta}_n\right)}\right\}$ so $\hat{\mu}^H_{\alpha,n}\geq\mu_{Y,n}\left(\hat{\theta}_n;P\right)$ if and only if

$$F_{TN}\left(Y^*_n\left(\hat{\theta}_n\right);0,\widehat{\Sigma}_{Y,n}\left(\hat{\theta}_n\right),\mathcal{L}^{H*}_n,\mathcal{U}^{H*}_n\right)\geq1-\alpha.$$

Lemma 8 implies that

$$\left(Y^*_{n_s},\widehat{\Sigma}_{Y,n_s},\mathcal{L}^{H*}_{n_s},\mathcal{U}^{H*}_{n_s},\hat{\theta}_{n_s}\right)\rightarrow_d\left(Y^*,\Sigma^*_Y,\mathcal{L}^{H*},\mathcal{U}^{H*},\hat{\theta}\right),$$

where $\mathcal{L}^{H*}$ and $\mathcal{U}^{H*}$ are equal to $\mathcal{L}^{H*}_n$ and $\mathcal{U}^{H*}_n$ after replacing $(X_n,Y_n,\widehat{\Sigma}_n)$ with (X,Y,Σ^*). Then by the continuous mapping theorem and (50),

$$\begin{aligned}&\left(F_{TN}\left(Y^*_{n_s}\left(\hat{\theta}_{n_s}\right);0,\widehat{\Sigma}_{Y,n_s}\left(\hat{\theta}_{n_S}\right),\mathcal{L}^{H*}_{n_s},\mathcal{U}^{H*}_{n_s}\right),1\left\{\hat{\theta}_{n_s}=\tilde{\theta},\mu_{Y,n_s}\left(\hat{\theta}_{n_s};P_{n_s}\right)\in CS^\beta_{P,n_s}\right\}\right)\\\rightarrow_d&\left(F_{TN}\left(Y^*\left(\hat{\theta}\right);0,\Sigma^*_Y\left(\hat{\theta}\right),\mathcal{L}^{H*},\mathcal{U}^{H*}\right),1\left\{\hat{\theta}=\tilde{\theta},Y^*\left(\hat{\theta}\right)\in\left[-c_\alpha(\Sigma^*_Y)\sqrt{\Sigma^*_Y\left(\hat{\theta}\right)},c_\alpha(\Sigma^*_Y)\sqrt{\Sigma^*_Y\left(\hat{\theta}\right)}\right]\right\}\right).\end{aligned}$$

Hence, by the same argument as in the proof of Proposition 8,

$$\lim_{s\to\infty} Pr_{P_{n_s}}\left\{\mu_{Y,n_s}\left(\hat{\theta}_{n_s};P_{n_s}\right)\in CS^H_{U,-,n_s}|\hat{\theta}_{n_s}=\tilde{\theta},\mu_{Y,n_s}\left(\hat{\theta}_{n_s};P_{n_s}\right)\in CS^\beta_{P,n_s}\right\}=\alpha,$$

as we aimed to show.

To prove (30), note that for $\widetilde{CS}^H_{U,+,n}=(\hat{\mu}^H_{\alpha,n},\infty)$,

$$\hat{\mu}^H_{\alpha,n}\geq\mu_{Y,n}\left(\hat{\theta}_n;P\right)\Longleftrightarrow\mu_{Y,n}\left(\hat{\theta}_n;P\right)\notin\widetilde{CS}^H_{U,+,n}$$

and thus that the argument above proves that

$$\lim_{n\to\infty}\sup_{P\in\mathcal{P}_n}\left|Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in\widetilde{CS}^H_{U,+,n}|C^H_n\left(\tilde{\theta};P\right)\right\}-(1-\alpha)\right|Pr_P\left\{C^H_n\left(\tilde{\theta};P\right)\right\}=0$$

for $C^H_n\left(\tilde{\theta};P\right)$ as in the statement of the proposition. Since

$$\sum_{\tilde{\theta}\in\Theta}Pr_P\left\{\hat{\theta}_{n_s}=\tilde{\theta},\mu_{Y,n_s}\left(\hat{\theta}_{n_s};P_{n_s}\right)\in CS^\beta_{P,n_s}\right\}=Pr_P\left\{\mu_{Y,n_s}\left(\hat{\theta}_{n_s};P_{n_s}\right)\in CS^\beta_{P,n_s}\right\}+o(1), \tag{51}$$

and Proposition 10 shows that

$$\liminf_{s\to\infty}\inf_{P\in\mathcal{P}_{n_s}}Pr_P\left\{\mu_{Y,n_s}\left(\hat{\theta}_{n_s};P_{n_s}\right)\in CS^\beta_{P,n_s}\right\}\geq1-\beta,$$

Lemma 6 together with (29) implies that

$$\liminf_{n\to\infty}\inf_{P\in\mathcal{P}_n}Pr_P\left\{\hat{\mu}^H_{\alpha,n}<\mu_{Y,n}\left(\hat{\theta}_n;P\right)\right\}\geq(1-\alpha)(1-\beta)=(1-\alpha)-\beta(1-\alpha)$$

and

$$\limsup_{n\to\infty}\sup_{P\in\mathcal{P}_n}Pr_P\left\{\hat{\mu}^H_{\alpha,n}<\mu_{Y,n}\left(\hat{\theta}_n;P\right)\right\}\leq1-\alpha(1-\beta)=(1-\alpha)+\beta\alpha$$

from which the second result of the proposition follows immediately. $\square$

Proof of Corollary 2 Note that by construction

$$CS^H_{ET,n}=\left[\hat{\mu}^H_{\frac{\alpha-\beta}{2(1-\beta)},n},\hat{\mu}^H_{1-\frac{\alpha-\beta}{2(1-\beta)},n}\right],$$

where $\hat{\mu}^H_{\frac{\alpha-\beta}{2(1-\beta)},n} < \hat{\mu}^H_{1-\frac{\alpha-\beta}{2(1-\beta)},n}$ provided $\frac{\alpha-\beta}{1-\beta} < 1$. Hence,

$$Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right) \in CS^H_{ET,n} | C^H_n\left(\tilde{\theta},P\right)\right\}$$

$$= Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right) \leq \hat{\mu}^H_{1-\frac{\alpha-\beta}{2(1-\beta)},n} | C^H_n\left(\tilde{\theta},P\right)\right\} - Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right) < \hat{\mu}^H_{\frac{\alpha-\beta}{2(1-\beta)},n} | C^H_n\left(\tilde{\theta},P\right)\right\},$$

so Proposition 11 immediately implies (31).

Equation (51) in the proof of Proposition 11 together with Lemma 6 implies that

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}_n} Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right) \in CS^H_{ET,n}\right\} \geq \frac{1-\alpha}{1-\beta}(1-\beta) = 1-\alpha$$

so (32) holds. We could likewise get an upper bound on coverage using Lemma 6, but obtain a sharper bound by proving the result directly. Specifically, note that

$$\mu_{Y,n}\left(\hat{\theta}_n;P_n\right) \in CS^H_{ET,n} \Rightarrow \mu_{Y,n}\left(\hat{\theta}_n;P_n\right) \in CS^\beta_{P,n}.$$

Hence,

$$Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right) \in CS^H_{ET,n}\right\}$$
$$= Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right) \in CS^H_{ET,n} | \hat{\mu}_{Y,n}\left(\hat{\theta}_n;P_n\right) \in CS^\beta_{P,n}\right\} Pr\left\{\mu_{Y,n}\left(\hat{\theta}_n;P_n\right) \in CS^\beta_{P,n}\right\}.$$

By the first part of the proposition, this implies that

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_n} Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right) \in CS^H_{ET,n}\right\} \leq \frac{1-\alpha}{1-\beta} \limsup_{n\to\infty} \sup_{P\in\mathcal{P}_n} Pr\left\{\mu_{Y,n}\left(\hat{\theta}_n;P_n\right) \in CS^\beta_{P,n}\right\}$$

$$\leq \frac{1-\alpha}{1-\beta},$$

so (33) holds as well. □

Proof of Proposition 12 The first part of the result follows by the same argument as in the proof of Proposition 9, where as in the proof of Proposition 11 we use the conditioning event $\left\{\hat{\theta}_n = \tilde{\theta}, \mu_{Y,n}\left(\hat{\theta}_n;P_n\right) \in CS^\beta_{P,n}\right\}$ and replace $(\mathcal{L}_n,\mathcal{U}_n)$ by $\left(\mathcal{L}^H_n,\mathcal{U}^H_n\right)$. The second part of the result follows by the same argument as in the proof of Corollary 2. □

E Additional Simulation Results for Stylized Example

In the stylized example discussed in Section 2 of the main text, we focus on the median length of confidence intervals and the median absolute error of estimators. In this section, we report results for other quantiles, in particular that τ-th quantiles for $\tau \in \{0.05, 0.25, 0.5, 0.75, 0.95\}$.

Figures 6 and 7 show the unconditional quantiles of the length of the 95% confidence intervals CS_U and CS_{ET}, for cases with $|\Theta|$=2, 10, and 50 policies. In each case and for each $\tau \in \{0.05, 0.25, 0.5, 0.75, 0.95\}$, the τ-th quantile is monotonically decreasing in $\mu(\theta_1) - \mu(\theta_{-1})$. Noting the different scales of the y-axes, we see that the upper quantiles grow as the number of policies increase, particularly for small $\mu(\theta_1) - \mu(\theta_{-1})$.

Figures 8 and 9 show the unconditional quantiles of the length of 95% hybrid confidence intervals CS_U^H and CS_{ET}^H with $\beta = 0.005$. Compared with Figures 6 and 7, the upper quantiles are much smaller, especially for small $\mu(\theta_1) - \mu(\theta_{-1})$. This substantial reduction in length directly comes from the construction of the hybrid confidence intervals, which ensures that CS_U^H and CS_{ET}^H are contained in CS_P^β. For the case of $|\Theta| = 50$, even the 95% quantiles of the length of CS_U^H and CS_{ET}^H are shorter than the length of CS_P uniformly over the range of $\mu(\theta_1) - \mu(\theta_{-1})$ values we consider.

Figures 10, 11, and 12 examine the performance of point estimators for $\mu(\hat{\theta})$. They plot the unconditional quantiles of the absolute error of the conventional estimator, the median unbiased estimator, and the hybrid estimator, respectively. In spite of the severe median bias shown in Figure 1 in the main text, the distribution of the conventional estimator is relatively concentrated compared to that of the median unbiased estimator. In particular, the upper quantiles of the absolute errors of $\hat{\mu}_{1/2}$ are very large for small $\mu(\theta_1) - \mu(\theta_{-1})$ (similar to the quantile plots of the length of CS_U and CS_{ET} shown in Figures 6 and 7).

At the cost of a small median bias, the hybrid estimator substantially reduces the absolute errors (Figure 12). The 95% quantile of the absolute errors of the hybrid estimator is overall similar to the 95% quantile of the absolute errors of the conventional estimator with a notable exception of the case of 2 policies. In contrast, for $|\Theta| = 10$ and 50, and for quantiles other than 95%, the hybrid estimator outperforms the conventional estimator over a wide range of values for $\mu(\theta_1) - \mu(\theta_{-1})$. These numerical results show that the hybrid estimator successfully reduces bias without greatly inflating the variability of the estimator.

F Additional Results for EWM Simulations

Tables 4 and 5 provide the ratios of the 5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} quantiles of the lengths of CS_{ET}, CS_U, CS_{ET}^H and CS_U^H relative to the corresponding length quantiles of CS_P for the

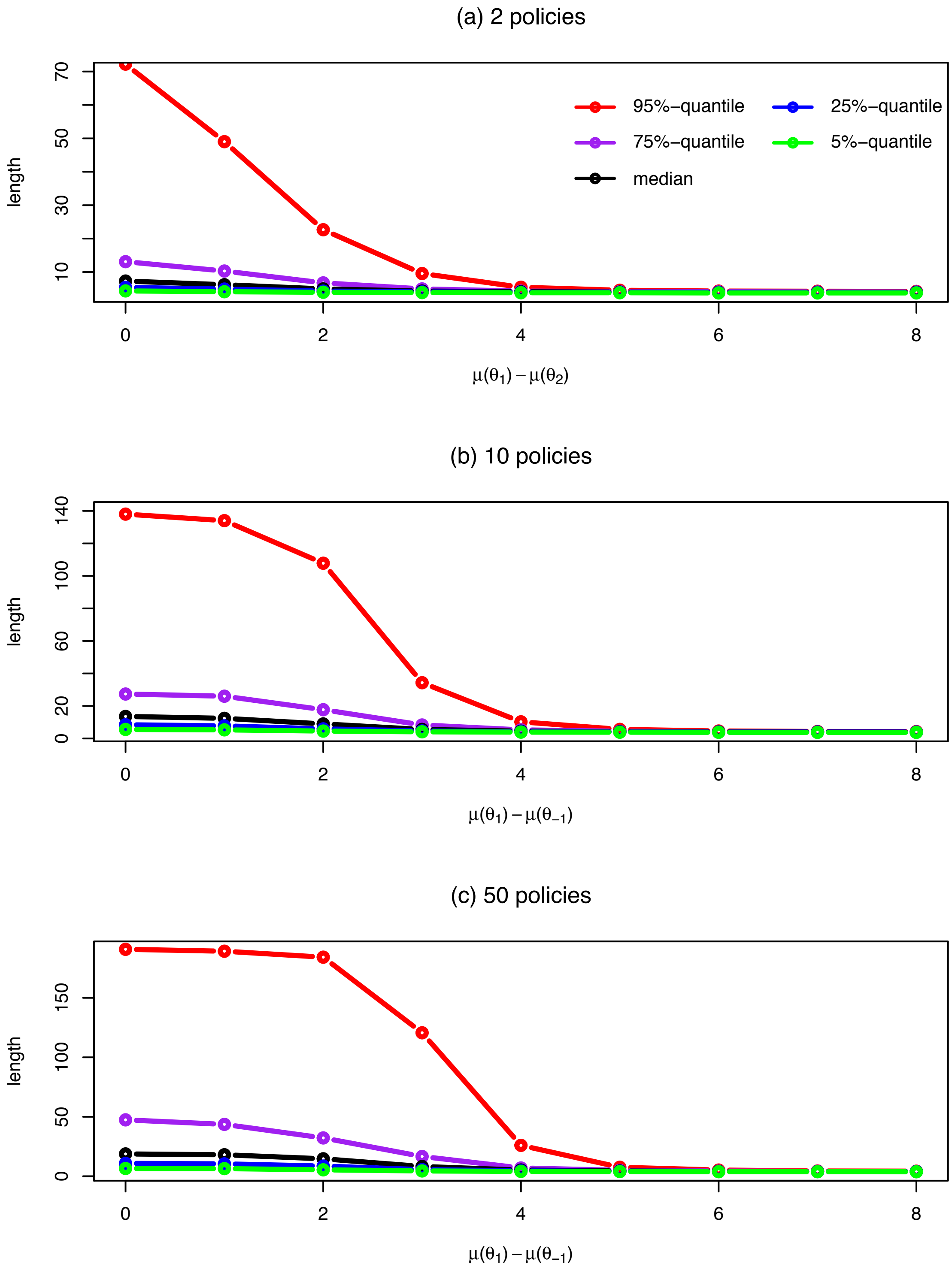

Figure 6: Quantiles of the length of 95% conditionally UMAU confidences sets CS_U.

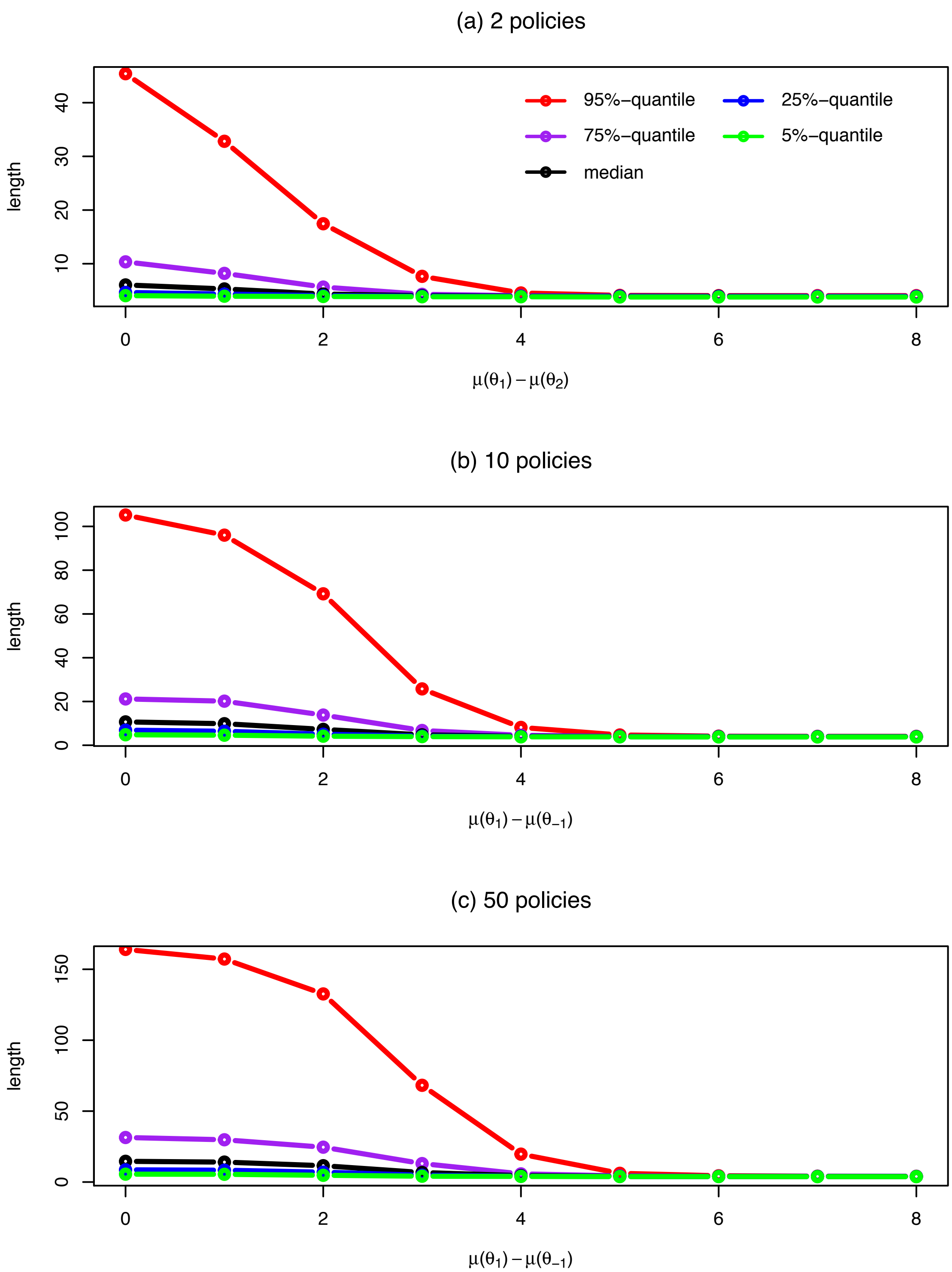

Figure 7: Quantiles of the length of 95% conditionally equal-tailed confidences sets CS_{ET}.

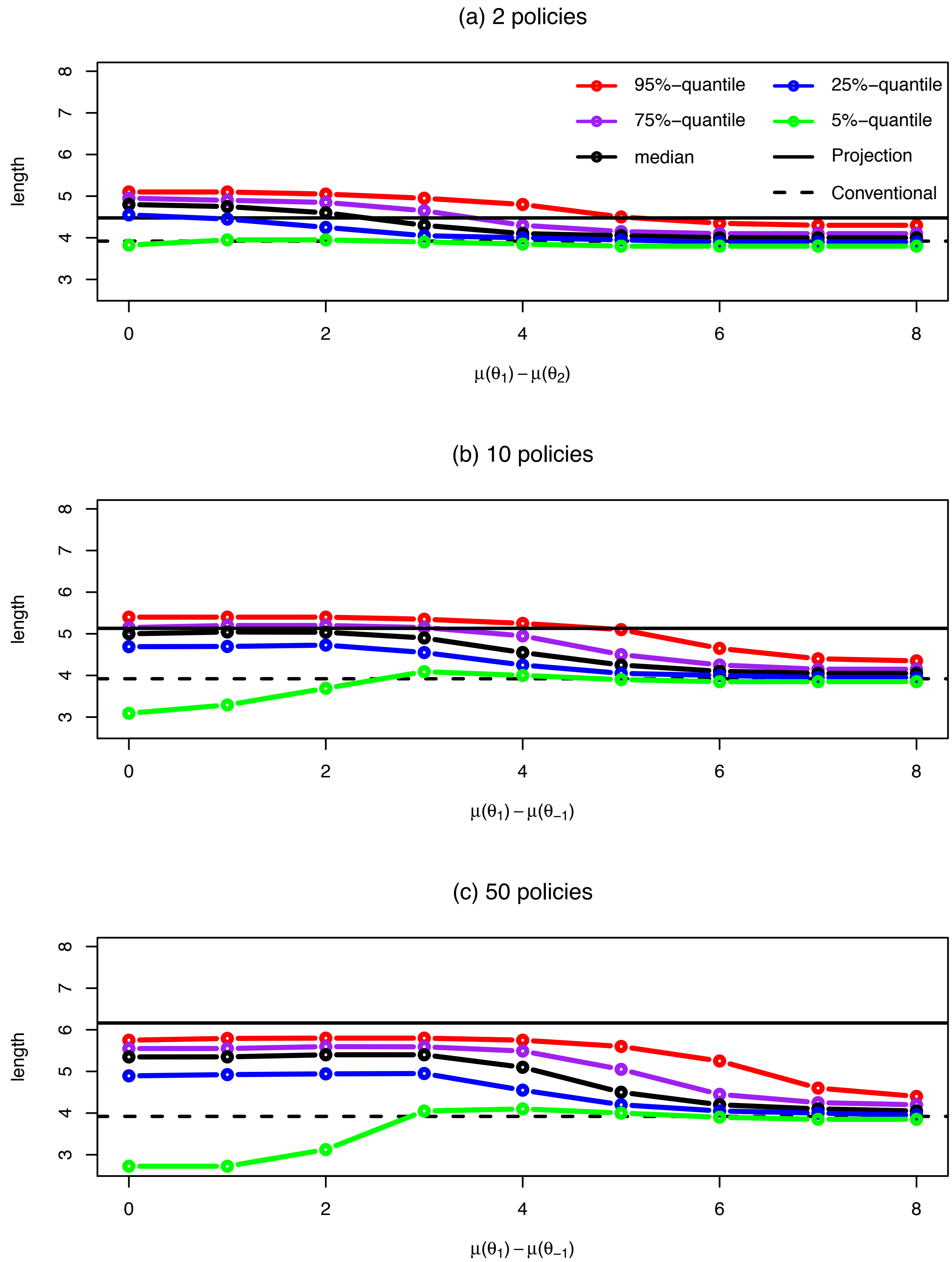

Figure 8: Quantiles of the length of 95% hybrid confidence intervals CS_U^H, with $\beta = 0.005$.

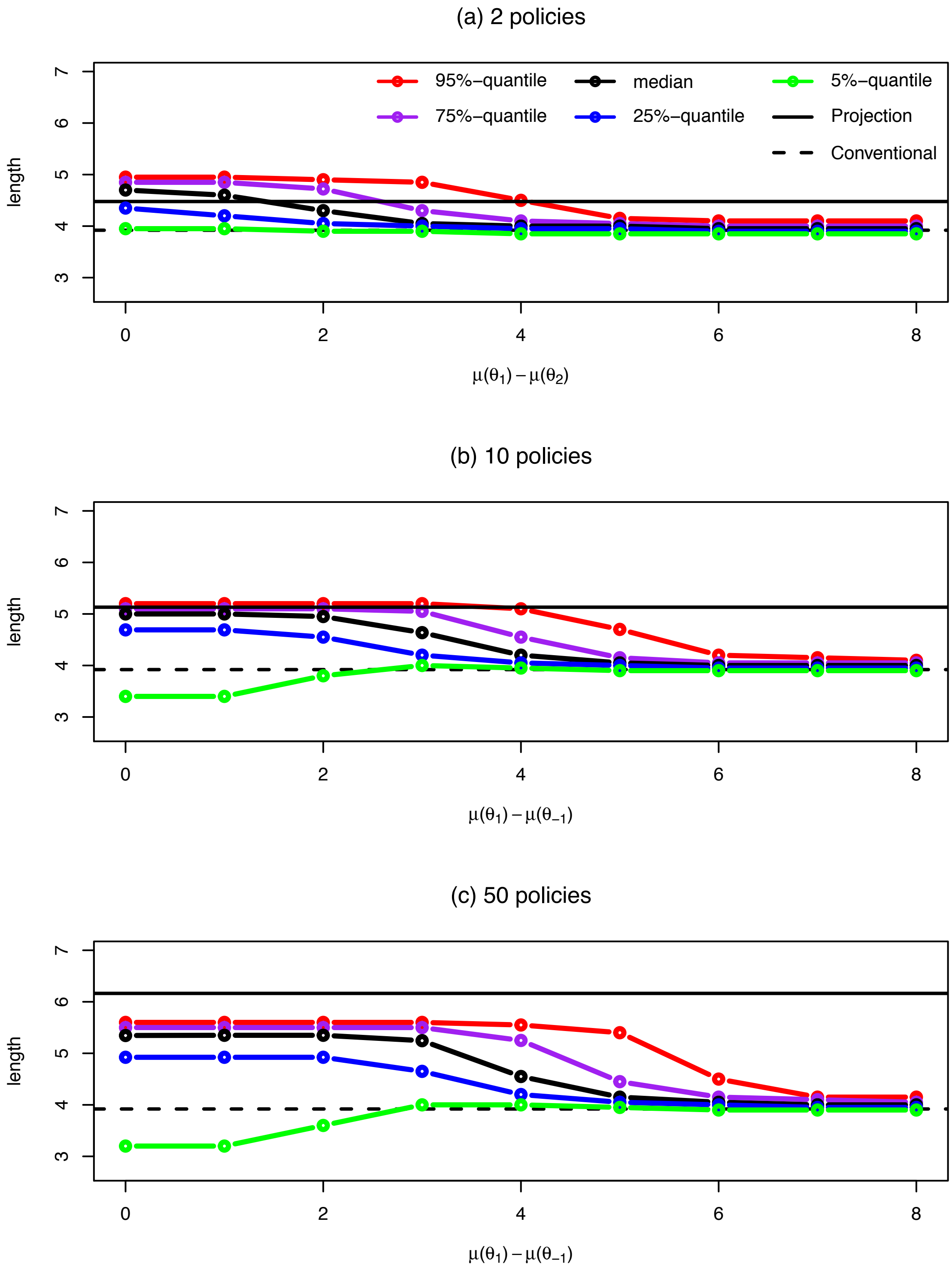

Figure 9: Quantiles of the length of 95% hybrid confidence intervals CS_{ET}^{H}, with $\beta = 0.005$.

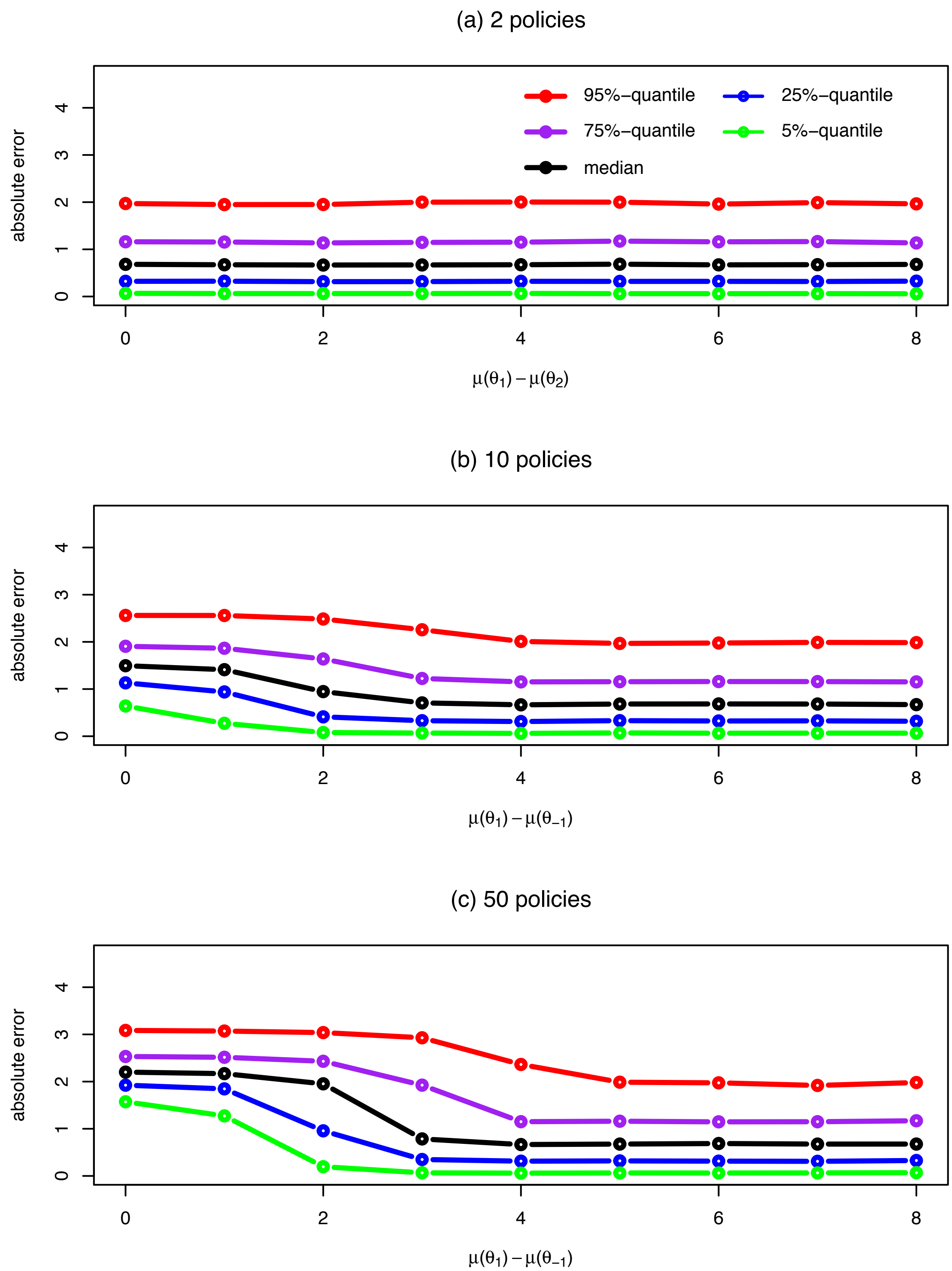

Figure 10: Quantiles of the absolute error of the conventional estimator (i.e. of $|X(\hat{\theta})-\mu(\hat{\theta})|$).

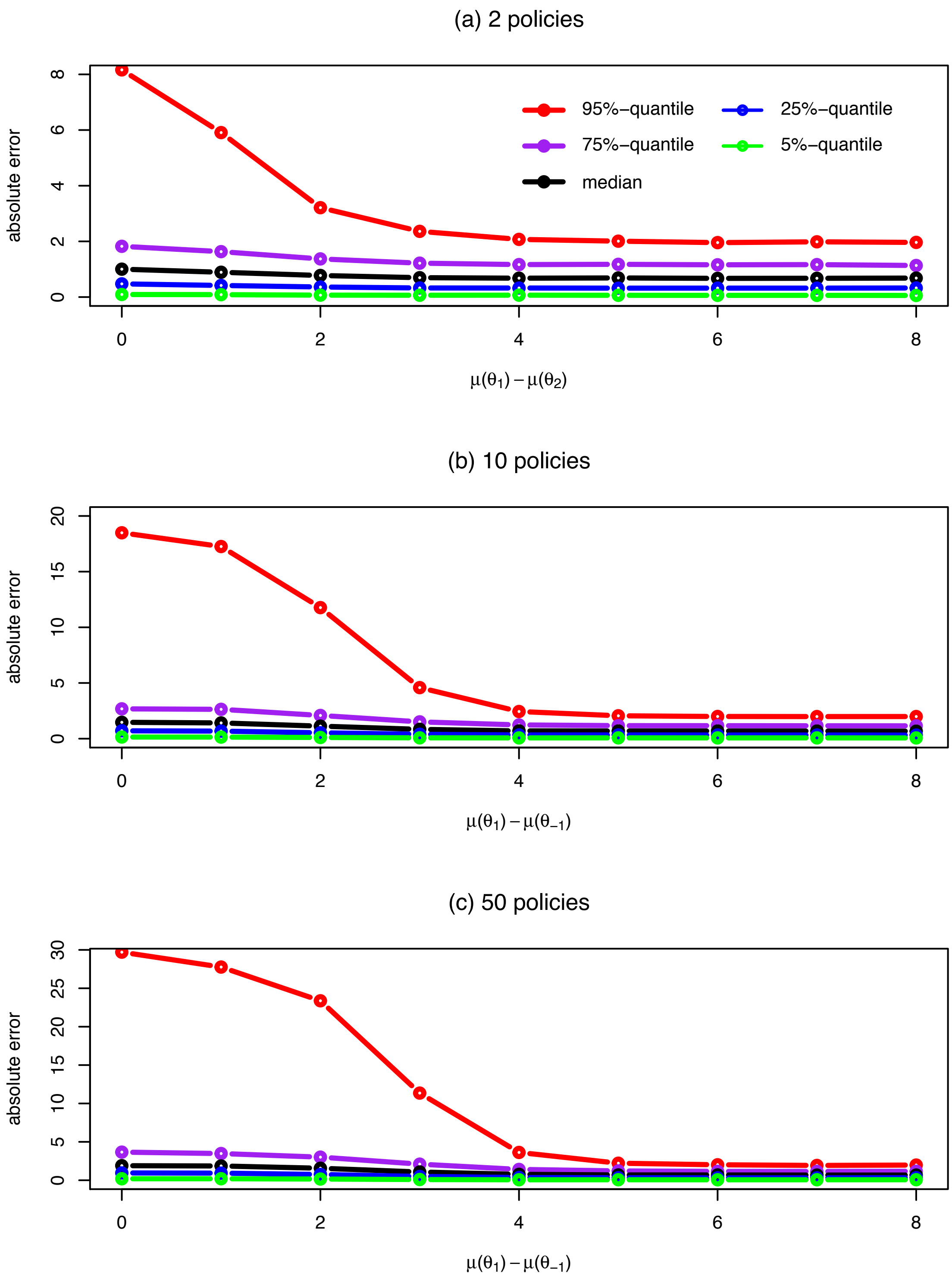

Figure 11: Quantiles of the absolute error of the conditionally optimal median unbiased estimator (i.e. of $|\hat{\mu}_{1/2} - \mu(\hat{\theta})|$).

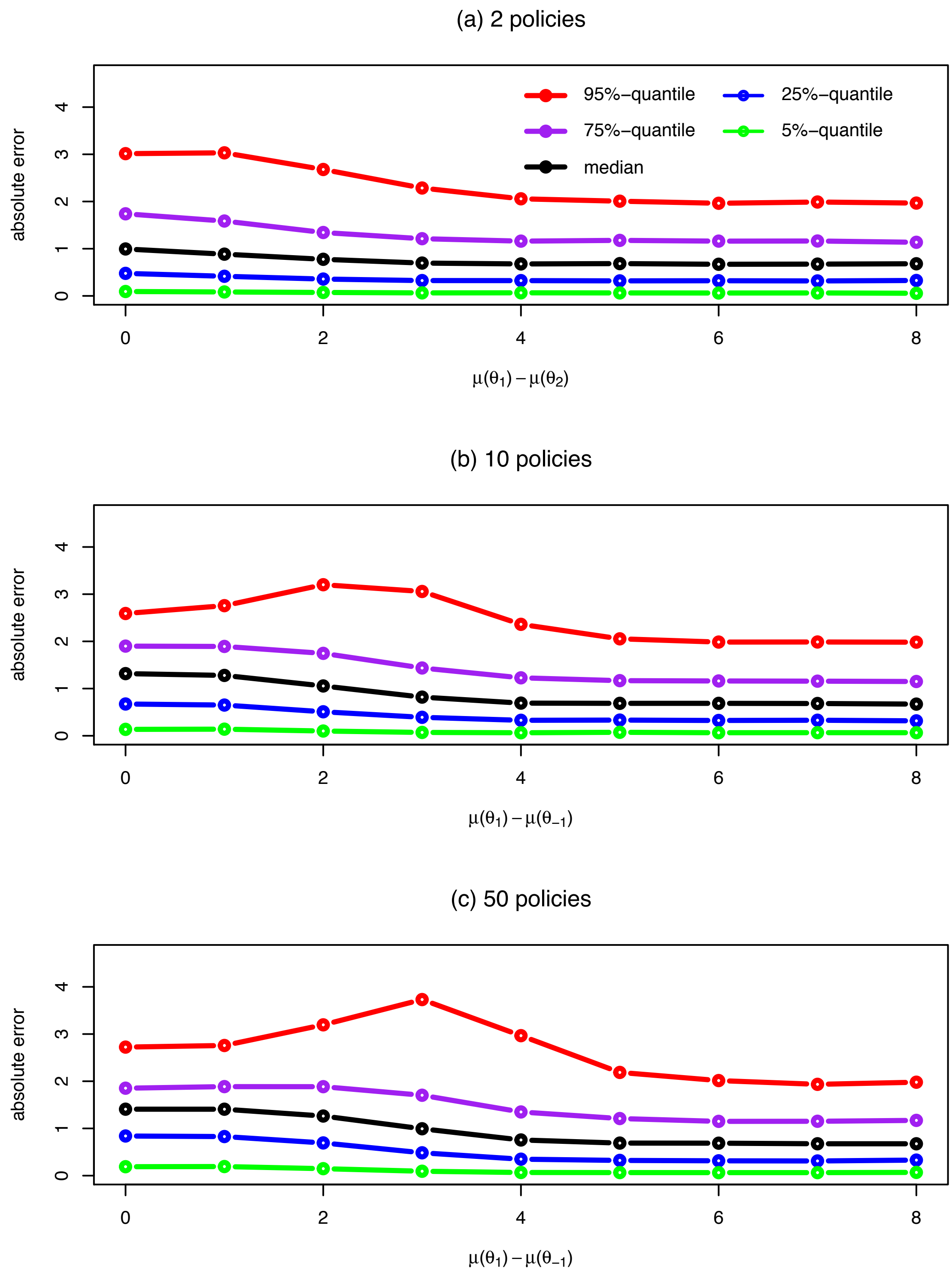

Figure 12: Quantiles of the absolute error of the hybrid estimator (i.e. of $|\hat{\mu}^H_{1/2} - \mu(\hat{\theta})|$) with $\beta = 0.005$.

EWM data-calibrated designs described in Section 6 of the main text. Looking at the upper quantiles in Table 4, we can see that the conditional confidence intervals CS_{ET} and CS_U can become very wide when the maximal element of μ_X is not well-separated from the others. On the other hand, Table 5 shows that the hybrid approach is very successful at mitigating this problem. Indeed, CS_{ET}^H and CS_U^H dominate CS_P across nearly all quantiles and simulation designs considered. Table 6 reports the same quantiles of the studentized absolute errors of $\hat{\mu}_{\frac{1}{2}}$, $\hat{\mu}_{\frac{1}{2}}^H$ and $Y(\hat{\theta})$. Here we can see that, although the hybrid estimator $\hat{\mu}_{\frac{1}{2}}^H$ does not dominate the conventional estimator $Y(\hat{\theta})$ according to this performance measure, it does dominate $\hat{\mu}_{\frac{1}{2}}$ across all quantiles and DGPs considered. This dominance is especially pronounced at higher quantiles. The underlying message here is a bit more nuanced than that which applies to the confidence intervals: when minimal bias is desired, $\hat{\mu}_{\frac{1}{2}}^H$ is the preferred estimator.

Table 4: Ratios of Length Quantiles Relative to CS_P

	CS_{ET} 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}
Class of Threshold Policies										
(i)	0.75	1.32	1.17	1.97	8.88	0.75	1.48	1.27	1.94	7.17
(ii)	0.74	0.75	0.75	0.75	0.76	0.74	0.75	0.75	0.75	0.75
(iii)	0.74	0.74	0.82	1.22	3.30	0.74	0.76	0.93	1.45	3.65
Class of Interval Policies										
(i)	1.11	1.41	1.54	2.31	10.78	1.27	1.54	1.65	1.91	8.72
(ii)	0.63	0.63	0.63	0.64	0.64	0.63	0.63	0.64	0.64	0.64
(iii)	0.66	0.71	0.78	1.14	4.39	0.70	0.76	0.88	1.36	3.61

Table 5: Ratios of Length Quantiles Relative to CS_P

	CS_{ET}^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}
Class of Threshold Policies										
(i)	0.76	0.85	0.63	0.93	0.99	0.76	0.77	0.64	0.95	1.01
(ii)	0.76	0.76	0.76	0.77	0.77	0.76	0.76	0.76	0.76	0.77
(iii)	0.77	0.78	0.84	0.92	0.98	0.79	0.81	0.89	0.96	1.00
Class of Interval Policies										
(i)	0.75	0.76	0.77	0.85	0.88	0.63	0.74	0.76	0.86	0.89
(ii)	0.64	0.65	0.65	0.65	0.65	0.64	0.65	0.65	0.65	0.65
(iii)	0.67	0.72	0.76	0.85	0.89	0.69	0.76	0.81	0.88	0.92

Table 6: 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}
Class of Threshold Policies															
(i)	0.11	0.54	1.11	2.01	10.65	0.11	0.53	1.10	1.91	3.04	0.11	0.47	0.88	1.36	2.14
(ii)	0.06	0.31	0.67	1.15	1.97	0.06	0.31	0.67	1.15	1.97	0.06	0.31	0.67	1.16	1.97
(iii)	0.08	0.36	0.80	1.43	3.60	0.08	0.36	0.79	1.43	2.90	0.06	0.31	0.67	1.15	1.93
Class of Interval Policies															
(i)	0.14	0.68	1.42	2.61	17.51	0.14	0.67	1.39	2.21	3.07	0.52	0.94	1.30	1.75	2.49
(ii)	0.06	0.31	0.65	1.13	1.92	0.06	0.31	0.65	1.13	1.92	0.06	0.31	0.65	1.14	1.92
(iii)	0.08	0.40	0.86	1.57	5.15	0.08	0.40	0.86	1.57	3.46	0.07	0.32	0.69	1.16	1.96

Table 7: Unconditional Coverage Probability with Estimated Variance Matrix

DGP	CS_{ET}	CS_U	CS_{ET}^H	CS_U^H	CS_P	CS_N
	Class of Threshold Policies					
(i)	0.944	0.945	0.948	0.948	0.984	0.916
(ii)	0.95	0.95	0.954	0.953	0.990	0.95
(iii)	0.946	0.946	0.950	0.951	0.991	0.948
	Class of Interval Policies					
(i)	0.948	0.950	0.952	0.954	0.989	0.821
(ii)	0.953	0.953	0.956	0.957	0.997	0.952
(iii)	0.947	0.947	0.953	0.953	0.997	0.948

Table 8: Length of Confidence Sets Relative to CS_P in EWM Simulations with Estimated Variance Matrix

DGP	Median Length Relative to CS_P				Probability Longer than CS_P			
	CS_{ET}	CS_U	CS_{ET}^H	CS_U^H	CS_{ET}	CS_U	CS_{ET}^H	CS_U^H
	Class of Threshold Policies							
(i)	1.16	1.27	0.65	0.66	0.70	0.79	0.05	0.3
(ii)	0.74	0.74	0.76	0.76	0	0	0	0
(iii)	0.83	0.92	0.83	0.90	0.32	0.42	0.03	0.24
	Class of Interval Policies							
(i)	1.51	1.60	0.74	0.73	0.78	0.88	0	0
(ii)	0.63	0.63	0.65	0.65	0	0	0	0
(iii)	0.78	0.89	0.77	0.82	0.33	0.42	0	0

Supplement References

Andrews, D. W. K., Cheng, X., and Guggenberger, P. (2018). Generic results for establishing the asymptotic size of confidence sets and tests. Forthcoming in *Journal of Econometrics*.

Table 9: Bias and Median Absolute Error of Point Estimators with Estimated Variance Matrix

DGP	$Pr_\mu\left\{\hat{\mu}>\mu_X(\hat{\theta})\right\}-\frac{1}{2}$			$Med_\mu\left(\frac{\hat{\mu}-\mu_X(\hat{\theta})}{\sqrt{\Sigma_X(\hat{\theta})}}\right)$			$Med_\mu\left(\frac{\lvert\hat{\mu}-\mu_X(\hat{\theta})\rvert}{\sqrt{\Sigma_X(\hat{\theta})}}\right)$		
	$\hat{\mu}_{\frac{1}{2}}$	$\hat{\mu}^H_{\frac{1}{2}}$	$X(\hat{\theta})$	$\hat{\mu}_{\frac{1}{2}}$	$\hat{\mu}^H_{\frac{1}{2}}$	$X(\hat{\theta})$	$\hat{\mu}_{\frac{1}{2}}$	$\hat{\mu}^H_{\frac{1}{2}}$	$X(\hat{\theta})$
Class of Threshold Policies									
(i)	-0.005	-0.004	0.397	-0.02	-0.02	0.82	1.12	1.11	0.86
(ii)	0.009	0.009	0.009	0.02	0.02	0.02	0.67	0.67	0.67
(iii)	0.006	0.006	0.104	0.02	0.02	0.26	0.80	0.79	0.67
Class of Interval Policies									
(i)	0.006	0.006	0.500	0.04	0.04	1.30	1.42	1.39	1.30
(ii)	0.009	0.009	0.009	0.02	0.02	0.02	0.65	0.65	0.65
(iii)	0.003	0.003	0.150	0.01	0.01	0.36	0.85	0.85	0.67