cemmap
centre for microdata methods and practice

# Inference on winners

Isaiah Andrews
Toru Kitagawa
Adam McCloskey

The Institute for Fiscal Studies
Department of Economics, UCL

**cemmap** working paper CWP73/18

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
An ESRC Research Centre

# Inference on Winners*

Isaiah Andrews† Toru Kitagawa‡ Adam McCloskey§

December 31, 2018

**Abstract**

Many empirical questions can be cast as inference on a parameter selected through optimization. For example, researchers may be interested in the effectiveness of the best policy found in a randomized trial, or the best-performing investment strategy based on historical data. Such settings give rise to a winner's curse, where conventional estimates are biased and conventional confidence intervals are unreliable. This paper develops optimal confidence sets and median-unbiased estimators that are valid conditional on the parameter selected and so overcome this winner's curse. If one requires validity only on average over target parameters that might have been selected, we develop hybrid procedures that combine conditional and projection confidence sets to offer further performance gains relative to existing alternatives.

KEYWORDS: WINNER'S CURSE, SELECTIVE INFERENCE

JEL CODES: C12, C13

## 1 Introduction

A wide range of empirical questions involve inference on target parameters selected through optimization over a finite set. In a randomized trial considering multiple treatments, for

*We would like to thank Tim Armstrong, Stéphane Bonhomme, Gregory Cox, Áureo de Paula, Patrick Kline, Hannes Leeb, Anna Mikusheva, Mikkel Plagborg-Møller, José Luis Montiel Olea, Jack Porter, Adam Rosen, Frank Schoerfheide, Jesse Shapiro, and participants at numerous seminars and conferences for helpful comments. Andrews gratefully aknowledges financial support from the NSF under grant number 1654234. Kitagawa gratefully aknowledges financial support from the ESRC through the ESRC Centre for Microdata Methods and Practice (CeMMAP) (grant number RES-589-28-0001) and the European Research Council (Starting grant No. 715940). Initial version posted May 10, 2018.

†Department of Economics, Harvard University, iandrews@fas.harvard.edu

‡CeMMAP and Department of Economics, University College London, t.kitagawa@ucl.ac.uk

§Department of Economics, University of Colorado, adam.mccloskey@colorado.edu

instance, one might want to learn about the true average effect of the treatment that performed best in the experiment. In finance, one might want to learn about the expected return of the trading strategy that performed best in a backtest. Perhaps less obviously, in threshold regression or tipping point models, researchers first estimate the location of a threshold by minimizing the sum of squared residuals and then seek to estimate the magnitude of the discontinuity taking the estimated threshold as given.

Estimators that do not account for data-driven selection of the target parameters can be badly biased, and conventional t-statistic-based confidence intervals may severely under-cover. To illustrate the problem, consider inference on the true average effect of the treatment that performed best in a randomized trial.[1] Since it ignores the data-driven selection of the treatment of interest, the conventional estimate for this average effect will be biased upwards. Similarly, the conventional confidence interval will under-cover, particularly when the number of treatments considered is large. This gives rise to a form of winner's curse, where follow-up trials will be systematically disappointing relative to what we would expect based on conventional estimates and confidence sets. This form of winner's curse has previously been discussed in contexts including genome-wide association studies (e.g. Zhong and Prentice, 2009; Ferguson et al., 2013) and online A/B tests (Lee and Shen, 2018).

This paper develops estimators and confidence sets that eliminate these biases and inference failures. There are two distinct perspectives from which to consider bias and coverage. The first conditions on the target parameter selected, for example on the identity of the best-performing treatment, while the second is unconditional and averages over possible target parameters. As we discuss in the next section, conditional validity is more demanding but may be desirable in some settings, for example when one wants to ensure validity conditional on the recommendation made to a policy maker. Both perspectives differ from inference on the effectiveness of the "true" best treatment, as in e.g. Chernozhukov et al. (2013) and Rai (2018), in that we consider inference on the

[1]Such a scenario seems to be empirically relevant, as a number of recently published randomized trials in economics either were designed with the intent of recommending a policy or represent a direct collaboration with a policy maker. For example, Khan et al. (2016) assesses how incentives for property tax collectors affect tax revenues in Pakistan, Banerjee et al. (2018) evaluates the efficacy of providing information cards to potential recipients of Indonesia's *Raskin* programme, and Duflo et al. (2018) collaborates with the Gujarat Pollution Control Board (an Indian regulator tasked with monitoring industrial emissions in the state) to evaluate how more frequent but randomized inspection of plants performs relative to discretionary inspection. Baird et al. (2016) finds that deworming Kenyan children had substantial beneficial effects on their health and labor market outcomes into adulthood, and Björkman Nyqvist and Jayachandran (2017) finds that providing parenting classes to Ugandan mothers has a greater impact on child outcomes than targeting these classes at fathers.

effectiveness of the (observed) best-performing treatment in the experiment rather than the (unobserved) best-performing treatment in the population.[2]

Considering first conditional inference, we derive optimal unbiased and equal-tailed confidence sets. Our results build on the rapidly growing literature on selective inference (e.g. Harris et al. (2016); Lee et al. (2016); Tian and Taylor (2016); Fithian et al. (2017)), which derives optimal conditional confidence sets in a range of other settings. We further observe that the results of Pfanzagl (1994) imply optimal median-unbiased estimators for conditional settings, which does not appear to have been previously noted in the selective inference literature. Hence, for settings where conditional validity is desired, we propose optimal inference procedures that eliminate the winner's curse noted above. We further show that in cases where this winner's curse does not arise (for instance because one treatment considered is vastly better than the others) our conditional procedures coincide with conventional ones. Hence, our corrections do not sacrifice efficiency in such cases.

A common alternative remedy for the biases we consider is sample splitting. In settings with independent observations, choosing the target parameter using the first part of the data and constructing estimates and confidence sets using the second part ensures unbiasedness of estimates and validity of conventional confidence sets conditional on the target parameter. Such conventional split-sample procedures can have undesirable properties, however. In particular, the target parameter is generally more variable than if constructed using the full data. Moreover, since only the second part of the data is used for inference, Fithian et al. (2017) show that conventional split-sample procedures are inadmissible within the class of procedures with the same target parameter. Motivated by this result, in the supplement to the paper we develop computationally tractable confidence sets and estimators that dominate conventional sample-splitting.

We next turn to unconditional inference. One approach to constructing unconditional confidence sets is projection, applied in various forms and settings by e.g. Romano and Wolf (2005), Berk et al. (2013), and Kitagawa and Tetenov (2018a). To obtain a projection confidence set, we form a simultaneous confidence band for all potential target parameters and take the implied set of values for the target parameter of interest. The resulting confidence sets have correct unconditional coverage but, unlike our conditional intervals, are wider than conventional confidence sets even when the latter are valid. On the other hand, we find in simulations that projection intervals outperform conditional intervals in

[2]See Dawid (1994) for an early discussion of this distinction, and an argument in favor of inference on the best-performing treatment in the experiment.

cases where there is substantial randomness in the target parameter, e.g. when there is not a clear best treatment.

Since neither conditional nor projection intervals perform well in all cases, we introduce hybrid confidence sets that combine conditioning and projection. These maintain most of the good performance of our conditional confidence intervals in cases for which the winner's curse does not arise but are subsets of (conservative) projection intervals by construction, limiting their maximal under-performance relative to projection confidence sets. We also introduce hybrid estimators that allow a controlled degree of bias while limiting the deviation from the conventional estimator.

We derive our main results in the context of a finite-sample normal model with an unknown mean vector and a known covariance matrix. This model can be viewed as an asymptotic approximation to non-normal finite sample problems where the optimal policy may not be obvious from the data. To formalize this connection, in the supplement to the paper we show that the procedures we derive are uniformly asymptotically valid over large classes of data-generating processes.

Since we are not aware of any other full-sample procedures that ensure validity conditional on the target parameter, our simulations focus on unconditional performance. The simulation designs are based on an empirical welfare maximization application from Kitagawa and Tetenov (2018b) and a threshold regression application from Card et al. (2008). In both settings, we find that while our conditional procedures exhibit good unconditional performance in cases where the objective function determining the target parameter has a well-separated optimum, their unconditional performance can be poor in other cases. By contrast, our hybrid procedures perform quite well: hybrid confidence sets are shorter than the previously available alternative (projection intervals) in all specifications, and are shorter than conditional intervals in all but the well-separated case (where they are nearly the same). Hybrid estimators eliminate nearly all the bias of conventional estimators, and are less dispersed than our exactly median unbiased estimators. These results show that while optimal conditional performance is attainable, conditional validity can come at the cost of unconditional performance. By combining conditional and projection approaches, our hybrid procedures yield better performance than either and offer a substantial improvement over existing alternatives.

While most of our simulation results focus on comparing our full-sample conditional and hybrid approaches to existing full-sample alternatives, Card et al. (2008) originally conducted inference based on a conventional split-sample approach. Hence, our simulations

based on Card et al. (2008) also compare conventional sample splitting procedures to our improved split-sample ones. We similarly find substantial performance improvements in these split-sample settings.

In this paper we focus on frequentist inference, and in particular on ensuring coverage and controlling bias under all parameter values. If one instead takes a Bayesian perspective then, as discussed by e.g. Dawid (1994), the selection issue does not arise since Bayesian inference conditions on the data and thus on any form of data-driven selection. One way to interpret this point is that e.g. the Bayes posterior median is median unbiased for the true parameter value under the prior. As highlighted by Dawid (1994), however, this property hinges crucially on the specification of the prior. If we consider frequentist performance in cases where the data are generated in a manner inconsistent with the prior, Bayes procedures may have large biases. In settings where we observe independent estimates for a large number of different parameters and are willing to assume that these parameters are drawn from some common unknown distribution, we can avoid this issue by adopting an empirical Bayes approach and estimating the prior (see Efron, 2011; Ferguson et al., 2013). Many settings, including our empirical welfare and threshold regression examples, lack this structure however, rendering this approach inapplicable.

It is important to emphasize that we take the rule for selecting the target parameter as given. In policy-evaluation contexts, for example, our goal is to evaluate the effectiveness of recommended policies taking the rule for selecting a recommendation as given, rather than to improve the rule. There are a number of reasons why valid confidence sets and median-unbiased estimates are of interest in such settings. One might be interested in understanding the true effectiveness of a selected policy for scientific reasons. Alternatively, one might want to assess uncertainty about the effect of a new policy for forecasting and risk management purposes. Finally, after a policy has been implemented or a follow-up trial conducted, one may want to test whether observed differences in efficacy can be explained solely by the winner's curse.

This paper is related to the literature on tests of superior predictive performance (e.g. White (2000); Hansen (2005); Romano and Wolf (2005)). This literature studies the problem of testing whether some strategy or policy beats a benchmark, while we consider the complementary question of inference on the effectiveness of the estimated "best" policy. Our conditional inference results combine naturally with the results of this literature, allowing one to condition inference on e.g. rejecting the null hypothesis that no policy outperforms a benchmark.

As mentioned above, our results are also closely related to the growing literature on selective inference. Fithian et al. (2017) describe a general conditioning approach applicable to a wide range of settings, while a rapidly growing literature including e.g. Harris et al. (2016); Lee et al. (2016); Tian and Taylor (2016) works out the details of this approach for a range of settings. Likewise, our analysis of conditional confidence sets examines the implications of the conditional approach in our setting. Our results are also related to the growing literature on unconditional post-selection inference, including e.g. Berk et al. (2013); Bachoc et al. (2017, 2018); Kuchibhotla et al. (2018). This literature considers analogs of our projection confidence sets for inference following model selection.

Beyond the new settings considered, we make two main theoretical contributions relative to the selective and post-selection inference literatures. First, when one only requires unconditional validity, we propose the class of hybrid inference and estimation procedures. We find that hybrid procedures offer large gains in unconditional performance relative both to conditional procedures and to existing unconditional alternatives. Second, for settings where conditional inference is desired, we observe that the same structure used to develop optimal conditional confidence sets also allows construction of optimal quantile unbiased estimators using the results of Pfanzagl (1994).[3]

In the next section, we begin by introducing the problem we consider and the techniques we propose in the context of a stylized example. Section 3 introduces the normal model in which we develop our main results, and shows how it arises as an asymptotic approximation to empirical welfare maximization and threshold regression examples. Section 4 develops our optimal conditional procedures, discusses their properties, and compares them to sample splitting. Section 5 introduces projection confidence intervals and our hybrid procedures. Finally, Sections 6 and 7 report results for simulations calibrated to empirical welfare maximization and threshold regression applications, respectively. The supplement to the paper collects proofs and other supporting material for the results in the main text, derives a computationally tractable split-sample approach that dominates conventional split-sample inference, shows that the finite sample results developed in the main text translate to uniform asymptotic results over large classes of data generating processes, and provides additional simulation results.

[3]Our asymptotic results are also novel relative to the literature. In particular, Tibshirani et al. (2018) establish uniform asymptotic validity for conditional confidence sets based on similar ideas to ours, but only under particular local sequences. We impose an analagous restriction for some of our asymptotic results but not others. See the supplement for details and further discussion.

## 2 A Stylized Example

We begin by illustrating the problem we consider, along with the solutions we propose, in a stylized example based on Manski (2004). In the treatment choice problem of Manski (2004) a treatment rule assigns treatments to subjects based on observable characteristics. Given a social welfare criterion and (quasi-)experimental data, Kitagawa and Tetenov (2018b) propose what they call empirical welfare maximization (EWM), which selects the treatment rule that maximizes the sample analog of the social welfare criterion over a class of candidate rules.

For simplicity suppose there are only two candidate policies: $\theta_1$ corresponding to "treat everyone" and $\theta_2$ corresponding to "treat no one." Suppose further that our social welfare function is the average of an outcome variable $Y$. If we have a sample of independent observations $i \in \{1,...,n\}$ from a randomized trial where a binary treatment $D_i \in \{0,1\}$ is randomly assigned to subjects with $Pr\{D_i = 1\} = d$, then as in Kitagawa and Tetenov (2018b) the scaled empirical welfare under $(\theta_1,\theta_2)$ is

$$(X_n(\theta_1),X_n(\theta_2)) = \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{D_i Y_i}{d}, \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(1-D_i)Y_i}{1-d} \right).$$

EWM selects the rule $\hat{\theta} = \text{argmax}_{\theta \in \{\theta_1,\theta_2\}} X_n(\theta)$.[4]

Kitagawa and Tetenov (2018b) show that the welfare from the policy selected by EWM converges to the optimal social welfare at the minimax optimal rate, providing a strong argument for this approach. Even after choosing a policy, we may want estimates and confidence intervals for its implied social welfare in order to learn about the size of the policy impact and communicate with stakeholders. For a fixed policy $\theta$, the empirical welfare $X_n(\theta)$ is unbiased for the true (scaled) social welfare $\mu_n(\theta)$ under the corresponding policy.[5] By contrast, the empirical welfare of the estimated optimal policy $X_n(\hat{\theta})$ is biased upwards relative to the true social welfare $\mu_n(\hat{\theta})$ since we are more likely to select a given policy when the empirical welfare over-estimates the true welfare. Likewise, confidence sets for $\mu_n(\hat{\theta})$ that ignore estimation of $\theta$ may cover $\mu_n(\hat{\theta})$ less often than we intend. This is a form of winner's curse: estimation error leads us to over-predict the benefits of our chosen policy and to misstate our uncertainty about its effectiveness.

[4] If the summands are instead weighted by sample propensity scores, we obtain Manski's conditional empirical success rule and the asymptotically optimal rules of Hirano and Porter (2009) with a symmetric loss.

[5] $X_n(\theta)$ is exactly mean-unbiased and asymptotically median-unbiased.

To simplify the analysis and develop corrected inference procedures, we turn to asymptotic approximations. Under mild conditions the central limit theorem implies that our estimates of social welfare are asymptotically normal:

$$\begin{pmatrix} X_n(\theta_1)-\mu_n(\theta_1) \\ X_n(\theta_2)-\mu_n(\theta_2) \end{pmatrix} \Rightarrow N\left(0, \begin{pmatrix} \Sigma(\theta_1) & \Sigma(\theta_1,\theta_2) \\ \Sigma(\theta_1,\theta_2) & \Sigma(\theta_2) \end{pmatrix}\right), \tag{1}$$

where the asymptotic variance $\Sigma$ can be consistently estimated while the scaled social welfare $\mu_n$ cannot be. To simplify the analysis, for this section only we assume that $\Sigma(\theta_1,\theta_2)=0$.[6] Motivated by (1), we abstract from approximation error and assume that we observe

$$\begin{pmatrix} X(\theta_1) \\ X(\theta_2) \end{pmatrix} \sim N\left(\begin{pmatrix} \mu(\theta_1) \\ \mu(\theta_2) \end{pmatrix}, \begin{pmatrix} \Sigma(\theta_1) & 0 \\ 0 & \Sigma(\theta_2) \end{pmatrix}\right)$$

for $\Sigma(\theta_1)$ and $\Sigma(\theta_2)$ known, and that $\hat{\theta}=\text{argmax}_{\theta\in\Theta}X(\theta)$ with $\Theta=\{\theta_1,\theta_2\}$.

As discussed above, $X(\hat{\theta})$ is biased upwards as an estimator of $\mu(\hat{\theta})$. This bias arises both conditional on $\hat{\theta}$ and unconditionally. To see this note that $\hat{\theta}=\theta_1$ if $X(\theta_1)>X(\theta_2)$, where ties occur with probability zero. Conditional on $\hat{\theta}=\theta_1$ and $X(\theta_2)$, $X(\theta_1)$ follows a normal distribution truncated below at $X(\theta_2)$. Since this holds for all $X(\theta_2)$, $X(\theta_1)$ has positive median bias conditional on $\hat{\theta}=\theta_1$:[7]

$$Pr_\mu\left\{X(\hat{\theta})\geq\mu(\hat{\theta})|\hat{\theta}=\theta_1\right\}>\frac{1}{2} \text{ for all } \mu. \tag{2}$$

Since the same argument holds for $\hat{\theta}=\theta_2$, $\hat{\theta}$ is likewise biased upwards unconditionally:

$$Pr_\mu\left\{X(\hat{\theta})\geq\mu(\hat{\theta})\right\}>\frac{1}{2} \text{ for all } \mu. \tag{3}$$

Note that (3) differs from (2) in that the target parameter is random. Unsurprisingly given this bias, the conventional confidence set which adds and subtracts a quantile of the standard normal distribution times the standard error need not have correct coverage.

To illustrate these issues, Figure 1 plots the coverage of conventional confidence sets, as well as the median bias of conventional estimates, in an example with $\Sigma(\theta_1)=\Sigma(\theta_2)=1$. For comparison we also consider cases with ten and fifty policies, $|\Theta|=10$ and $|\Theta|=50$, where

[6] One can show that $\Sigma(\theta_1,\theta_2)=-\mu(\theta_1)\mu(\theta_2)$, so this restriction arises naturally if one models $\mu$ as shrinking with the sample size to keep it on the same order as sampling uncertainty: $\mu_n=\frac{1}{\sqrt{n}}\mu^*$.

[7] It also has positive mean bias, but we focus on median bias for consistency with our later results.

we again set $\Sigma(\theta)=1$ for all $\theta$ and for ease of reporting assume that all the policies other than the first are equally effective: $\mu(\theta_2)=\mu(\theta_3)=...=\mu(\theta_{-1})$. The first panel of Figure 1 shows that while the conventional confidence set has reasonable coverage when there are only two policies, its coverage can fall substantially when $|\Theta|=10$ or $|\Theta|=50$.[8] The second panel shows that the median bias of the conventional estimator $\hat{\mu}=X(\hat{\theta})$, measured as the deviation of the exceedance probability $Pr_\mu\{X(\hat{\theta})\geq\mu(\hat{\theta})\}$ from $\frac{1}{2}$, can be quite large. The third panel shows that the same is true when we measure bias as the median of $X(\hat{\theta})-\mu(\hat{\theta})$. In all cases we find that performance is worse when we consider a larger number of policies, as is natural since a larger number of policies allows more scope for selection.

Our results correct these biases. Returning to the case with $|\Theta|=2$ for simplicity, let $F_{TN}(x(\theta_1);\mu(\theta_1),x(\theta_2))$ denote the (truncated normal) distribution function for $X(\theta_1)$ truncated below at $x(\theta_2)$ when the true social welfare for $\theta_1$ is $\mu(\theta_1)$. For fixed $x(\theta_1)>x(\theta_2)$ this function is strictly decreasing in $\mu(\theta_1)$, and for $\hat{\mu}_\alpha$ that solves $F_{TN}(X(\theta_1);\hat{\mu}_\alpha,X(\theta_2))=1-\alpha$, Proposition 1 below shows that

$$Pr_\mu\Big\{\hat{\mu}_\alpha\geq\mu(\hat{\theta})|\hat{\theta}=\theta_1\Big\}=\alpha \quad \text{for all } \mu.$$

Hence, $\hat{\mu}_\alpha$ is $\alpha$-quantile unbiased for $\mu(\hat{\theta})$ conditional on $\hat{\theta}=\theta_1$, and the analogous statement holds conditional on $\hat{\theta}=\theta_2$. Indeed, Proposition 1 shows that $\hat{\mu}_\alpha$ is the optimal $\alpha$-quantile unbiased estimator conditional on $\hat{\theta}$.

Using this result, we can eliminate the biases discussed above. The estimator $\hat{\mu}_{1/2}$ is median unbiased and the equal-tailed confidence interval $CS_{ET}=\left[\hat{\mu}_{\alpha/2},\hat{\mu}_{1-\alpha/2}\right]$ has conditional coverage $1-\alpha$, where we say that a confidence set $CS$ has conditional coverage $1-\alpha$ if

$$Pr\Big\{\mu(\hat{\theta})\in CS|\hat{\theta}=\theta_j\Big\}\geq 1-\alpha \text{ for } j\in\{1,2\} \text{ and all } \mu. \tag{4}$$

While the equal-tailed confidence interval is easy to compute, there are other confidence sets available in this setting. As in Lehmann and Scheffé (1955) and Fithian et al. (2017) it is possible to construct a uniformly most accurate unbiased (UMAU) confidence set, $CS_U$, conditional on $\hat{\theta}$. To construct $CS_U$, we collect the parameter values not rejected by a uniformly most powerful unbiased test conditional on $\hat{\theta}$. While straightforward to implement, the exact form of this test is somewhat involved and so is deferred to Section 4 below. The equal-tailed confidence set $CS_{ET}$ is not unbiased, so there is not a clear ranking between $CS_{ET}$ and $CS_U$.

[8]For example, these could correspond to cases where we consider "treat no one" along with nine or forty nine different treatment assignment rules, respectively.

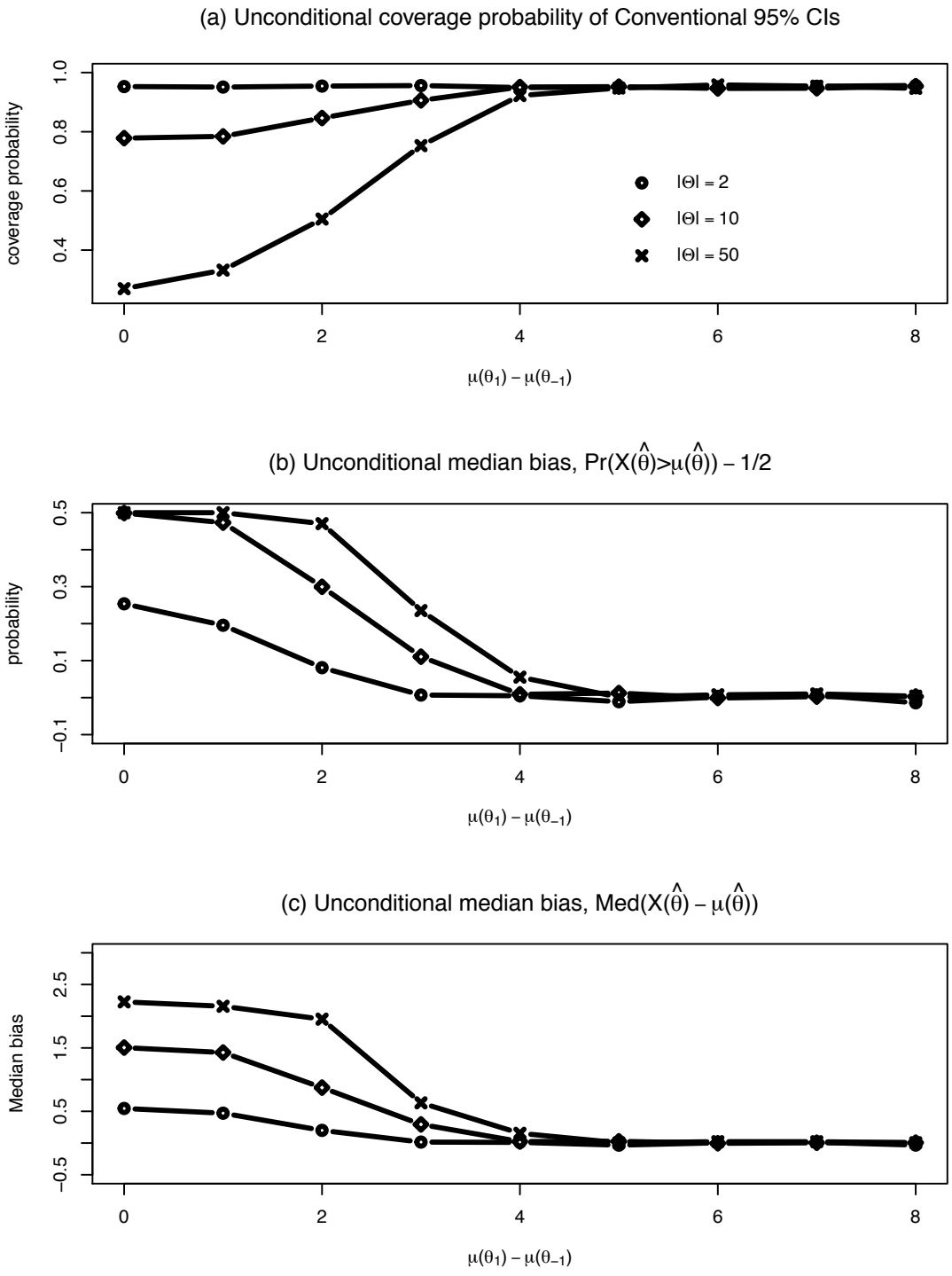


**Figure 1:** Performance of conventional procedures in examples with 2, 10, and 50 policies.

The law of iterated expectations implies that $CS_{ET}$ and $CS_U$ have unconditional coverage $1-\alpha$ as well:

$$Pr_\mu\Big\{\mu(\hat{\theta})\in CS\Big\}\geq 1-\alpha \text{ for all } \mu. \tag{5}$$

Unconditional coverage is easier to attain, so relaxing the coverage requirement from (4) to (5) may yield tighter confidence sets in some cases. Conditional and unconditional coverage requirements address different questions, however, and which is more appropriate depends on the problem at hand. In the EWM problem, for instance, a policy maker who is told the recommended policy $\hat{\theta}$ along with a confidence interval may want the confidence interval to be valid conditional on the recommendation, which is precisely the conditional coverage requirement (4). In particular, this ensures that if one considers repeated instances in which EWM recommends a particular course of action (e.g. departure from the status quo), reported confidence sets will in fact cover the true effects a fraction $1-\alpha$ of the time. On the other hand, if we only want to ensure that our confidence sets cover the true value with probability at least $1-\alpha$ on average across the distribution of recommendations, it suffices to impose the unconditional requirement (5).

We are unaware of alternative procedures that ensure conditional coverage (4).[9] For unconditional coverage (5), however, Kitagawa and Tetenov (2018a) propose an unconditional confidence set based on projecting a simultaneous confidence band for $\mu$ to obtain a confidence set for $\mu(\hat{\theta})$. In particular, let $c_\alpha$ denote the $1-\alpha$ quantile of $\max_j|\xi_j|$ for $\xi=(\xi_1,\xi_2)'\sim N(0,I_2)$ a two-dimensional standard normal random vector. If we define $CS_P$ as

$$CS_P=\left[Y(\hat{\theta})-c_\alpha\sqrt{\Sigma(\hat{\theta})},Y(\hat{\theta})+c_\alpha\sqrt{\Sigma(\hat{\theta})}\right],$$

this set has correct unconditional coverage (5).

Figure 2 plots the median (unconditional) length of 95% confidence sets $CS_{ET}$, $CS_U$, and $CS_P$, along with the conventional confidence set, again in cases with $|\Theta|\in\{2,10,50\}$. We focus on median length, rather than mean length, because the results for Kivaranovic and Leeb (2018) imply that both $CS_{ET}$ and $CS_U$ have infinite expected length.[10] As Figure 2 illustrates, the median lengths of $CS_{ET}$ and $CS_U$ are shorter than the (nonrandom) length of $CS_P$ when $|\mu(\theta_1)-\mu(\theta_{-1})|$ exceeds four, and converges to the length of the

[9] As noted in the introduction and further discussed in Section 4.3 below, split-sample confidence intervals also have conditional coverage but change the definition of $\hat{\theta}$.

[10] While Kivaranovic and Leeb (2018) do not consider the behavior of unbiased confidence sets, one can show that the expected length of the level $1-\alpha$ unbiased confidence set is bounded below by that of the level $1-2\alpha$ equal-tailed confidence set.

conventional interval as $|\mu(\theta_1)-\mu(\theta_{-1})|$ tends to infinity. When $|\mu(\theta_1)-\mu(\theta_{-1})|$ is small, on the other hand, $CS_{ET}$ and $CS_U$ can be substantially wider than $CS_P$. Both features become more pronounced as we increase the number of policies considered, and are still more pronounced for higher quantiles of the length distribution. To illustrate, Figure 3 plots the 95th percentile of the distribution of length in the case with $|\Theta|=50$ policies, while results for other quantiles and specifications are reported in Section E of the supplement.

In Figure 4 we plot the median absolute error $Med_\mu\Big(|\hat{\mu}-\mu(\hat{\theta})|\Big)$ for different estimators, and find that the median-unbiased estimator likewise exhibits larger median absolute error than the conventional estimator $X(\hat{\theta})$ when $|\mu(\theta_1)-\mu(\theta_{-1})|$ is small.[11] This feature is again more pronounced as we increase the number of policies considered, or if we consider higher quantiles as in Section E of the supplement.

Recall that $CS_U$ is the optimal unbiased confidence set, while the endpoints of $CS_{ET}$ are optimal quantile unbiased estimators. So long as we impose correct conditional coverage (4) and unbiasedness, there is therefore no scope to improve unconditional performance. If we instead require only correct unconditional coverage (5), improved performance is possible.

To improve performance, we consider hybrid confidence sets $CS^H_{ET}$ and $CS^H_U$. As detailed in Section 5.2 below, these confidence sets are constructed analogously to $CS_{ET}$ and $CS_U$, but further condition on the event that the true social welfare falls in the level $1-\beta$ projection interval $CS^\beta_P$ for $\beta<\alpha$. This ensures that the hybrid confidence sets are never longer than the level $1-\beta$ projection interval, and so both limits the performance deterioration when $|\mu(\theta_1)-\mu(\theta_{-1})|$ is small and ensures that the expected length of hybrid confidence sets is always finite. These hybrid confidence sets have correct unconditional coverage (5), but do not in general have correct conditional coverage (4). By relaxing the conditional coverage requirement, however, we obtain major improvements in unconditional performance, as illustrated in Figure 2. In particular, we see that in the cases with 10 and 50 policies, the hybrid confidence sets have shorter median length than the unconditional interval $CS_P$ for all parameter values considered. The gains relative to conditional confidence sets are large for many parameter values, and are still more pronounced for higher quantiles of the length distribution, as in Figure 3 and Section E of the supplement. In Figure 4 we report results for a hybrid estimation procedure based on a similar approach (detailed in Section 5.3 below), and again find substantial performance improvements.

The improved unconditional performance of the hybrid confidence sets is achieved by

[11] The proof of Proposition 1 of Kivaranovic and Leeb (2018) implies that the mean absolute error of the median unbiased estimator is infinite.

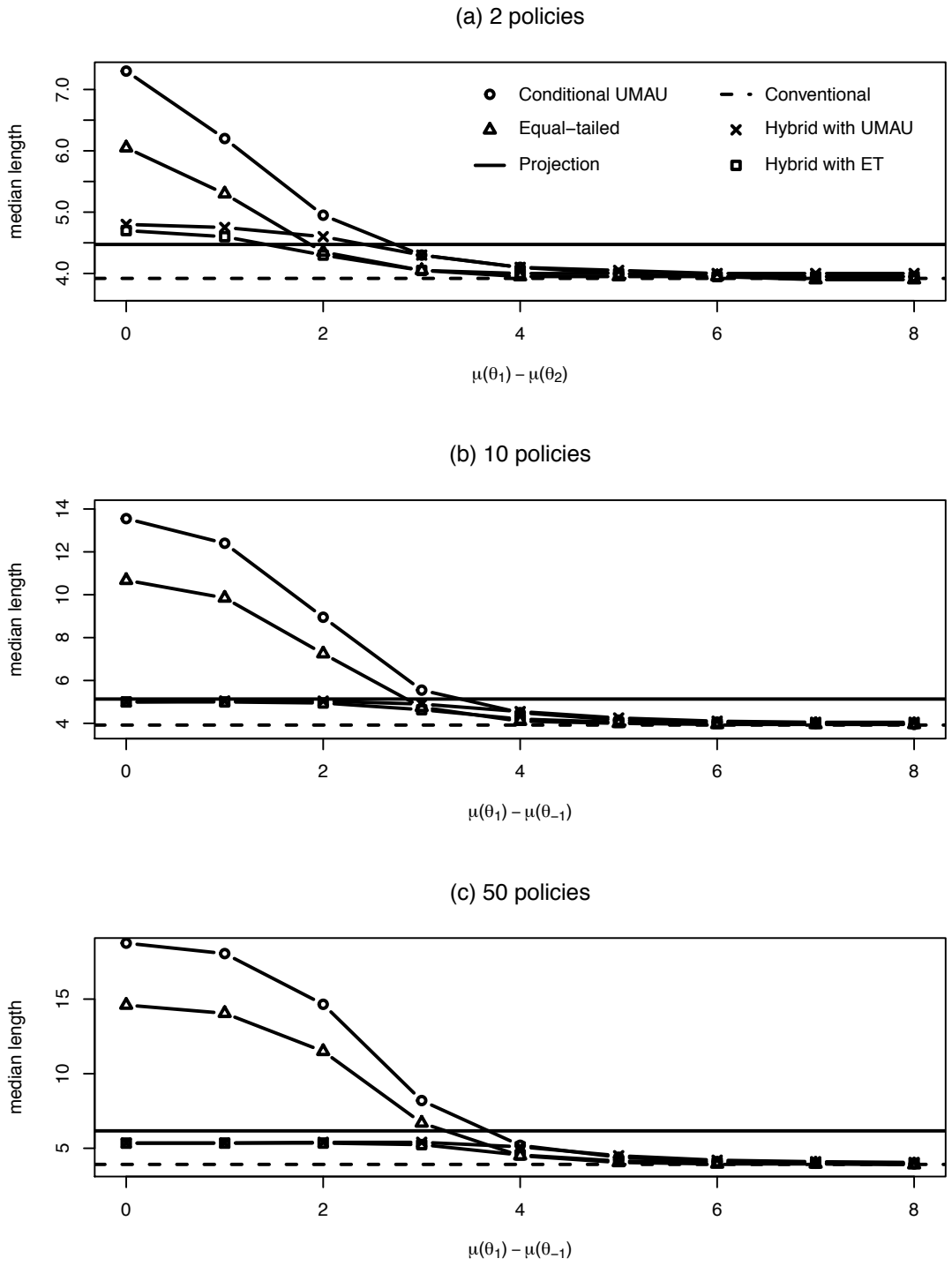


**Figure 2:** Median length of confidence sets for $\mu(\hat{\theta})$ in cases with 2, 10, and 50 policies.

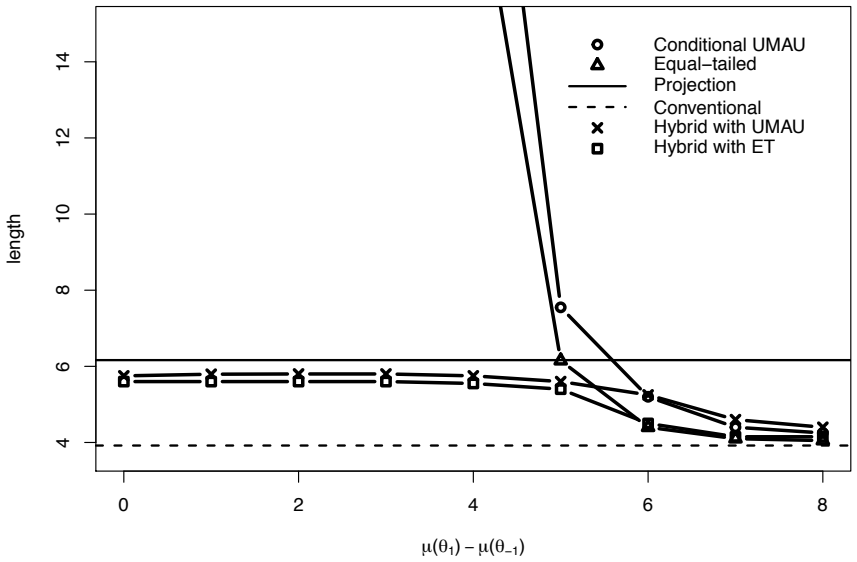


**Figure 3:** 95th percentile of length of confidence sets for $\mu(\hat{\theta})$ in case with 50 policies.

requiring only unconditional, rather than conditional, coverage. To illustrate, Figure 5 plots the conditional coverage given $\hat{\theta}=\theta_1$ in the case with two policies. As expected, the conditional intervals have correct conditional coverage, while coverage distortions appear for the hybrid and projection intervals when $\mu(\theta_1)\ll\mu(\theta_2)$. In this case $\hat{\theta}=\theta_2$ with high probability but the data will nonetheless sometimes realize $\hat{\theta}=\theta_1$. Conditional on this event, $X(\theta_1)$ will be far away from $\mu(\theta_1)$ with high probability, so projection and hybrid confidence sets under-cover.

## 3 Setting

This section introduces our general setting, which extends the stylized example of the previous section in several directions. We assume that we observe normal random vectors $\big(X(\theta)',Y(\theta)\big)'$ for $\theta\in\Theta$ where $\Theta$ is a finite set, $X(\theta)\in\mathbb{R}^{d_X}$, and $Y(\theta)\in\mathbb{R}$. In particular, for $\Theta=\big\{\theta_1,...,\theta_{|\Theta|}\big\}$, let $X=\Big(X(\theta_1)',...,X\big(\theta_{|\Theta|}\big)'\Big)'$ and $Y=\big(Y(\theta_1),...,Y\big(\theta_{|\Theta|}\big)\big)'$. Then

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N(\mu,\Sigma) \tag{6}$$

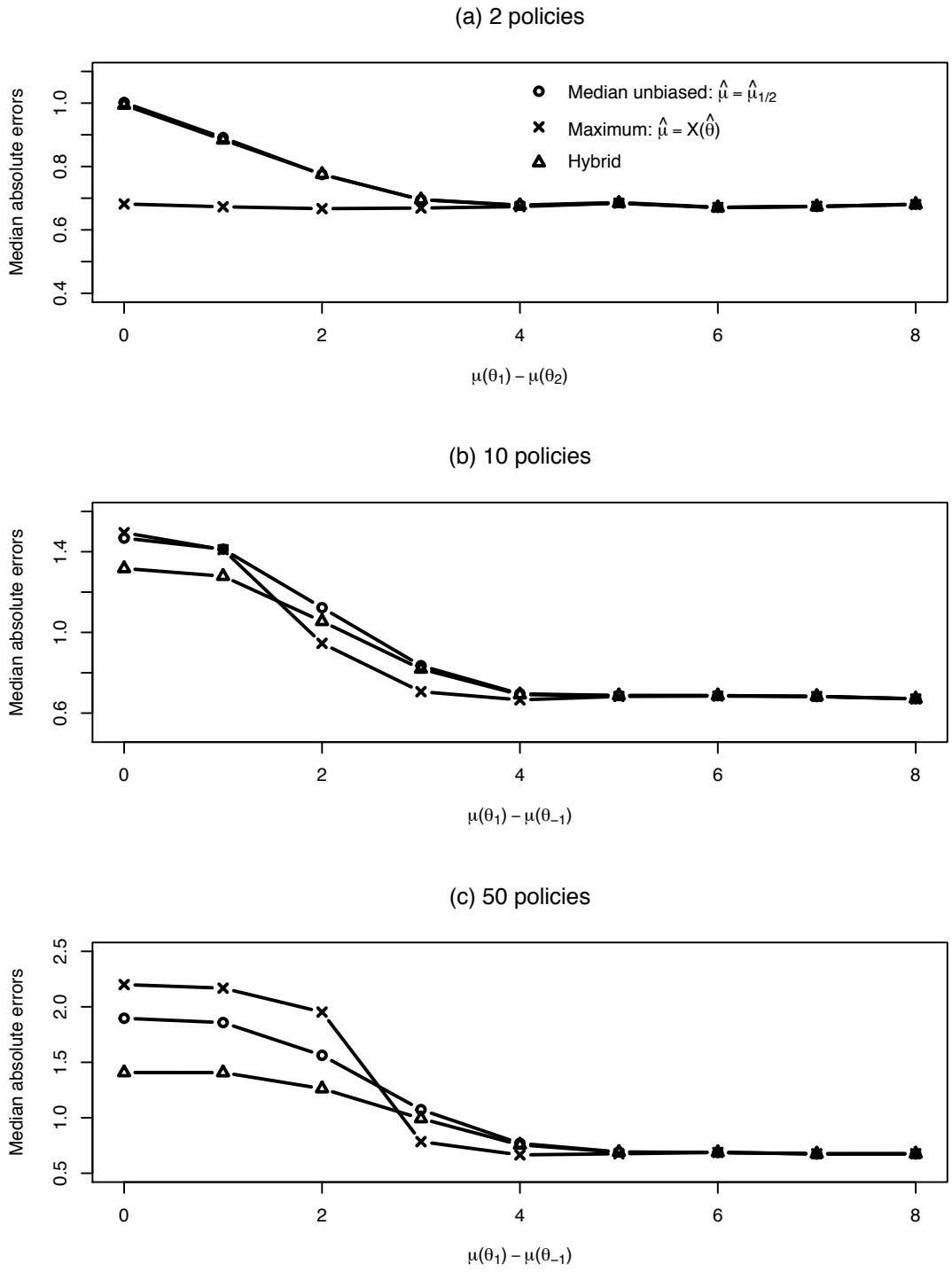


**Figure 4:** Median absolute error of estimators of $\mu(\hat{\theta})$ in cases with 2, 10, and 50 policies.

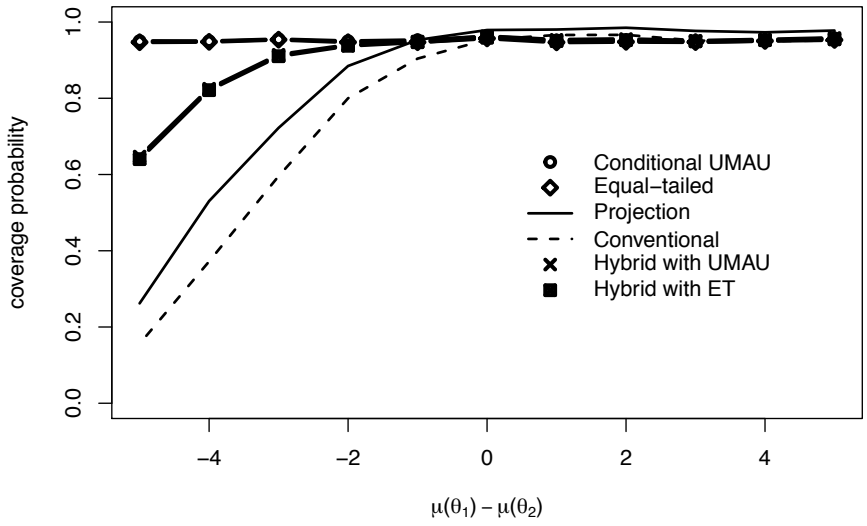


**Figure 5:** Coverage conditional on $\hat{\theta}=\theta_1$ in case with two policies.

for

$$E\left[\begin{pmatrix} X(\theta) \\ Y(\theta) \end{pmatrix}\right]=\mu(\theta)=\begin{pmatrix} \mu_X(\theta) \\ \mu_Y(\theta) \end{pmatrix},$$

$$\Sigma(\theta,\tilde{\theta})=\begin{pmatrix} \Sigma_X(\theta,\tilde{\theta}) & \Sigma_{XY}(\theta,\tilde{\theta}) \\ \Sigma_{YX}(\theta,\tilde{\theta}) & \Sigma_Y(\theta,\tilde{\theta}) \end{pmatrix}=Cov\left(\begin{pmatrix} X(\theta) \\ Y(\theta) \end{pmatrix},\begin{pmatrix} X(\tilde{\theta}) \\ Y(\tilde{\theta}) \end{pmatrix}\right).$$

We assume that $\Sigma$ is known, while $\mu$ is unknown and unrestricted unless noted otherwise. For brevity of notation, we abbreviate $\Sigma(\theta,\theta)$ to $\Sigma(\theta)$. We will show that this model arises naturally as an asymptotic approximation. We assume throughout that $\Sigma_Y(\theta)>0$ for all $\theta\in\Theta$, since the inference problem we study is trivial when $\Sigma_Y(\theta)=0$.

We are interested in inference on $\mu_Y(\hat{\theta})$, where $\hat{\theta}$ is determined based on $X$. We define $\hat{\theta}$ through either the *level maximization* problem where (for $d_X=1$)

$$\hat{\theta}=\underset{\theta\in\Theta}{\operatorname{argmax}}\, X(\theta), \tag{7}$$

or the *norm maximization* problem where (for $d_X\geq 1$)

$$\hat{\theta}=\underset{\theta\in\Theta}{\operatorname{argmax}}\, \|X(\theta)\|, \tag{8}$$

with $\|\cdot\|$ denoting the Euclidean norm.[12] We will again be interested in constructing confidence sets for $\mu_Y(\hat{\theta})$ that are valid either conditional on the value of $\hat{\theta}$ or unconditionally, as well as median-unbiased estimates. We may also want to condition on some additional event $\hat{\gamma}=\tilde{\gamma}$, for $\hat{\gamma}=\gamma(X)$ a function of $X$ which takes values in the finite set $\Gamma$. In such cases, we aim to construct confidence sets for $\mu_Y(\hat{\theta})$ that are valid conditional on the pair $(\hat{\theta},\hat{\gamma})$. Examples of such additional conditioning events are discussed below.

In the remainder of this section, we show how this class of problems arises in examples and discuss the choice between conditional and unconditional confidence sets in each case. We first revisit the EWM problem in a more general setting and show that it gives rise to the level maximization problem (7) asymptotically. We then discuss threshold regression models and show that they reduce to the norm maximization problem (8) asymptotically. We also briefly discuss other examples giving rise to level and norm maximization problems, and note that finite sample results for level and norm maximization in the normal model (6) translate to uniform asymptotic results over large classes of models.

**Empirical Welfare Maximization** As in the last section, we aim to select a welfare-maximizing treatment rule from a set of policies $\Theta$ in the EWM problem of Kitagawa and Tetenov (2018b). Let us assume that we have a sample of independent observations $i\in\{1,...,n\}$ from a randomized trial where treatment is randomly assigned conditional on observables $C_i$ with $Pr\{D_i=1|C_i\}=d(C_i)$. We consider policies that assign units to treatment based on the observables, where rule $\theta$ assigns $i$ to treatment if and only if $C_i\in\mathcal{C}_\theta$. The scaled empirical welfare under policy $\theta$ is[13]

$$X_n(\theta)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{Y_iD_i}{d(C_i)}1\{C_i\in\mathcal{C}_\theta\}+\frac{Y_i(1-D_i)}{1-d(C_i)}1\{C_i\notin\mathcal{C}_\theta\}\right).$$

EWM again selects the policy that maximizes empirical welfare: $\hat{\theta}_n=\text{argmax}_{\theta\in\Theta}X_n(\theta)$.

The definition of $Y_n$ in this setting depends on the object of interest. We may be interested in the overall social welfare, in which case we can define $Y_n=X_n$. Alternatively we could be interested in social welfare relative to the baseline of no treatment, in which case we can define $Y_n(\theta)$ as the difference in scaled empirical welfare between policy $\theta$ and

[12]For simplicity of notation we will assume $\hat{\theta}$ is unique almost surely unless noted otherwise. Our conditional analysis does not rely on this assumption, however: see footnote 20 below.

[13]Kitagawa and Tetenov (2018b) primarily consider welfare relative to the baseline of no treatment, which yields the same optimal policy.

the policy that treats no one, which we denote by $\theta=0$:

$$Y_n(\theta)=X_n(\theta)-X_n(0)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{Y_iD_i}{d(C_i)}-\frac{Y_i(1-D_i)}{1-d(C_i)}\right]1\{C_i\in\mathcal{C}_\theta\}.$$

Likewise, we might be interested in the social welfare for a particular subgroup defined by the observables, say $\mathcal{S}$, in which case we can take

$$Y_n(\theta)=\frac{\sqrt{n}\sum_{i=1}^{n}\left(\frac{Y_iD_i}{d(C_i)}1\{C_i\in\mathcal{S}\cap\mathcal{C}_\theta\}+\frac{Y_i(1-D_i)}{1-d(C_i)}1\{C_i\in\mathcal{S}\setminus\mathcal{C}_\theta\}\right)}{\sum_{i=1}^{n}1\{C_i\in\mathcal{S}\}}.$$

For $\mu_{X,n}$ and $\mu_{Y,n}$ the true scaled social welfare corresponding to $X_n$ and $Y_n$,

$$\begin{pmatrix} X_n-\mu_{X,n} \\ Y_n-\mu_{Y,n} \end{pmatrix}\Rightarrow N(0,\Sigma) \tag{9}$$

under mild conditions, where the covariance $\Sigma$ will depend on the data generating process and the definition of $Y_n$ but is consistently estimable. By contrast, the scaling of $X_n$ and $Y_n$ means that $\mu_{X,n}$ and $\mu_{Y,n}$ are not consistently estimable. As in the last section, this suggests the asymptotic problem where we observe normal random vectors $(X,Y)$ as in (6) with $\Sigma$ known and $\hat{\theta}$ defined as in (7), the level maximization problem.[14]

As argued in the last section, if a policy maker is given a recommended policy $\hat{\theta}$ as well as a confidence set for $\mu_Y(\hat{\theta})$, it is natural to require that the confidence set be valid conditional on the recommendation. It may also be natural to condition on additional variables. For example, if a recommendation is made only when we reject the null hypothesis that no policy in $\Theta$ improves outcomes over the base case of no treatment, $H_0:\max_{\theta\in\Theta}\mu(\theta)\leq\mu(0)$, then it is also natural to condition inference on this rejection.[15] To cover this case we can define $\hat{\gamma}=\gamma(X)$ as a dummy for rejection of $H_0$. If on the other hand we care only about performance on average across a range of recommendations, we need only impose unconditional coverage. $\triangle$

The level maximization problem arises in a number of other settings as well. For example, selecting the "best" policy from a collection considered in A/B tests is closely

[14]Under mild regularity conditions, (9) also holds in settings where the empirical welfare involves estimated propensity scores and/or estimated outcome regressions, e.g., the hybrid procedures of Kitagawa and Tetenov (2018b) and the doubly robust welfare estimators of Athey and Wager (2018).

[15]In the case of $|\Theta|=2$, conditioning on this rejection can be interpreted as conditioning on the event that the decision criterion of Tetenov (2012) supports the same policy.

related to EWM. Further afield, the literature on tests of superior predictive performance (c.f. White (2000); Hansen (2005); Romano and Wolf (2005)) considers the problem of testing whether some trading strategies or forecasting rules amongst a candiate set beat a benchmark. If we define $X_n = Y_n$ as the vector of performance measures for different strategies, $X_n$ is asymptotically normal under mild conditions (see e.g. Romano and Wolf (2005)). If one wants to form a confidence set for the performance of the "best" strategy based on $X_n$ (perhaps also conditioning on the result of a test for superior performance), this reduces to our level maximization problem asymptotically.

Another example comes from Bhattacharya (2009) and Graham et al. (2014), who consider the problem of optimally matching individuals to maximize peer effects. For $X_n$ again a scaled objective function, the results of Bhattacharya (2009) show that his problem reduces to level maximization asymptotically when one considers a finite set of assignments. More broadly, any time we consider $M$-estimation with a finite parameter space and are interested in the value of the population objective or some other function at the estimated optimal value, this falls into our level maximization framework under mild conditions.

We next discuss an example of threshold regression estimation, showing that it gives rise to our norm-maximization problem asymptotically.

**Threshold Regression Estimation** Suppose we observe data on an outcome $Y_i$, a threshold regressor $Q_i$ and a $k$-dimensional vector of regressors $C_i$ for $i \in \{1,...,n\}$. We assume there is a linear but potentially regressor-dependent relationship between $Y_i$ and $C_i$:

$$Y_i = C_i'(\beta + \varphi_n(Q_i)) + U_i, \tag{10}$$

where $Q_i \in \mathbb{R}$ and the residuals $U_i$ are orthogonal to $Q_i$ and $C_i$. Similarly to Elliott and Müller (2014) and Wang (2018), the function $\varphi_n : \mathbb{R} \to \mathbb{R}^k$ determines the value of the regressor-dependent coefficient $\beta + \varphi_n(Q_i)$. This model nests the traditional threshold regression model (see e.g. Hansen (2000) and references therein) by taking

$$\varphi_n(Q_i) = 1(Q_i > \theta)\delta, \tag{11}$$

where $\theta \in \mathbb{R}$ is the "true" threshold. The threshold model (11) is often used as a parsimonious approximation to a more general linear regression model with regressor-dependent coefficents. For example, Card et al. (2008) use the threshold model to approximate a theoretical model with smoothly-varying regressor-dependent coefficients. See also the

motivations for this model discussed in Hansen (1997, 2000).

Since the threshold regression model is widely used in practice, we consider a researcher who fits the model (11). To allow the possibility of misspecification, however, we assume only that the data is generated by (10). To provide a good asymptotic approximation to finite sample behavior, we follow Elliott and Müller (2007, 2014) and Wang (2018) and model parameter instability as on the same order as sampling uncertainty, with $\varphi_n(Q_i) = \frac{1}{\sqrt{n}} g(Q_i)$ for a fixed function $g$. We further assume that

$$\frac{1}{n}\sum_{i=1}^{n} C_i C_i' 1(Q_i \leq \theta) \to_p \Sigma_C(\theta), \quad \frac{1}{n}\sum_{i=1}^{n} C_i C_i' g(Q_i) 1(Q_i \leq \theta) \to_p \Sigma_{Cg}(\theta), \tag{12}$$

and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} C_i U_i 1(Q_i \leq \theta) \Rightarrow G(\theta), \tag{13}$$

all uniformly in $\theta \in \mathbb{R}$. Here $\Sigma_C : \mathbb{R} \to \mathbb{R}^{k \times k}$ is a consistently-estimable matrix-valued function and $\Sigma_C(\theta)$ is full rank for all $\theta$ in the interior of the support of $Q_i$, $\Sigma_{Cg} : \mathbb{R} \to \mathbb{R}^k$ is a vector-valued function, and $G(\cdot)$ is a $k$-dimensional mean zero Gaussian process with a consistently estimable covariance function that is positive definite when evaluated at points in the interior of the support of $Q_i$. Conditions (12) and (13) are analogous to Conditions 1(ii) and 1(iv) of Elliott and Müller (2007) for structural break models in a time-series setting. See Wang (2018) for sufficient conditions that give rise to (12) and (13).

The standard threshold estimator $\hat{\theta}_n$ chooses $\theta$ to minimize the sum of squared residuals in an OLS regression of $Y_i$ on $C_i$ and $\mathbf{1}(Q_i > \theta) C_i$ across a finite grid of thresholds $\Theta$.[16] For

$$X_n(\theta) = \begin{pmatrix} (\sum_{i=1}^n C_i C_i' 1(Q_i \leq \theta))^{-\frac{1}{2}} (\sum_{i=1}^n C_i \eta_i 1(Q_i \leq \theta)) \\ (\sum_{i=1}^n C_i C_i' 1(Q_i > \theta))^{-\frac{1}{2}} (\sum_{i=1}^n C_i \eta_i 1(Q_i > \theta)) \end{pmatrix},$$

with $\eta_i \equiv U_i + n^{-1/2} C_i' g(Q_i)$, arguments analogous to those in the proof of Proposition 1 in Elliott and Müller (2007) imply that $\hat{\theta}_n = \text{argmax}_{\theta \in \Theta} \|X_n(\theta)\| + o_p(1)$, where $o_p(1)$ is an asymptotically negligible term. Hence, $\hat{\theta}_n$ is asymptotically equivalent to the solution to a particular norm-maximization problem (8).

Suppose we are interested in the approximate change in the $j$th parameter $\delta_j = e_j' \delta$,

[16] Note that finiteness of $\Theta$ is without loss of generality if $Q_i$ is finitely-supported, but that we otherwise limit attention to a finite collection of thresholds.

where $e_j$ is the $j^{th}$ standard basis vector.[17] In practice it is common to estimate $\delta$ by least squares imposing the estimated threshold $\hat{\theta}_n$. When the threshold regression model (11) is misspecified, however, there is neither a "true" threshold $\theta$ nor a "true" change coefficient $\delta$. Instead, the population regression coefficient $\delta(\theta)$ imposing threshold $\theta$ depends on $\theta$. Thus, for threshold $\theta$, the coefficient of interest is $\delta_j(\theta)$. Denote the OLS estimate imposing threshold $\theta$ by $\hat{\delta}_j(\theta)$ and define $Y_n(\theta) = \sqrt{n}\hat{\delta}_j(\theta)$. If we define $\mu_{Y,n}(\theta) = \sqrt{n}\delta_j(\theta)$ as the scaled coefficient of interest and $\mu_{X,n}(\theta)$ as the population analog of $X_n(\theta)$, Section B.2 of the supplement shows that

$$\begin{pmatrix} X_n(\theta) - \mu_{X,n}(\theta) \\ Y_n(\theta) - \mu_{Y,n}(\theta) \end{pmatrix} \Rightarrow N(0, \Sigma(\theta)) \tag{14}$$

uniformly over a parameter space $\Theta$ contained in the interior of the support of $Q_i$, where the covariance matrix $\Sigma(\theta)$ is consistently estimable but $\mu_{X,n}(\theta)$ and $\mu_{Y,n}(\theta)$ are not. As before, this suggests the asymptotic problem (6) where we now define $\hat{\theta}$ through norm maximization (8).

Since the estimated threshold $\hat{\theta}$ is random and the parameter of interest $\delta_j(\theta)$ depends on $\theta$, it is important to account for this randomness in our inference procedures. In particular, it may be appealing to condition inference on the estimated threshold $\hat{\theta}$, since we only seek to conduct inference on $\delta_j(\tilde{\theta})$ when $\hat{\theta} = \tilde{\theta}$. It may also be natural to condition inference on additional variables. For example, if we report a confidence set for the change coefficient $\delta_j(\hat{\theta})$ only when we reject the null hypothesis of parameter constancy, $H_0 : \varphi_n(\theta) = 0$ for all $\theta$, it is natural to condition inference on this rejection. As above, this can be accomplished by defining $\hat{\gamma} = \gamma(X)$ as a dummy for rejection of $H_0$, and conditioning inference on $(\hat{\theta}, \hat{\gamma})$. Even if we only desire coverage of $\delta_j(\hat{\theta})$ on average over the distribution of $\hat{\theta}$, and so prefer to consider unconditional confidence sets, accounting for the randomness of $\hat{\theta}$ remains important. If on the other hand we are confident that the threshold model is correctly specified, so that (11) holds in the data, it will typically be more appealing to focus on inference for the "true" parameters as in Elliott and Müller (2014) and Wang (2018). $\triangle$

An analogous analysis applies to estimation and inference in the traditional structural break model (see e.g. Hansen (2001) and Perron (2006) and references therein) under local asymptotics as in Elliott and Müller (2007, 2014). Moreover, while our discussion of threshold regression estimation focuses on the linear model (10), Elliott and Müller

[17] By changing the definition of $Y_n$ below, our results likewise apply to the pre-change parameters $\beta_j$ and the post-change parameters $\beta_j + \delta_j$, amongst other possible objects of interest.

(2014) show that structural break estimation in nonlinear models with time-varying parameters gives rise to the same asymptotic problem. Hence, our results apply in that setting as well. Likewise, Wang (2018) shows that the same asymptotic problem arises in nonlinear threshold models.[18] Further afield, one could generalize our approach to consider norm-minimization rather than norm-maximization, and so derive results for GMM-type problems with finite parameter spaces.

**Uniform Asymptotic Validity** We have shown that the emprical welfare maximization and threshold regression problems asymptotically resemble level and norm maximization based on the finite-sample normal model (6). Section D of the supplement builds on this connection and shows that if we consider classes of data generating processes such that $(X_n,Y_n)$ are uniformly well-approximated by the normal model (6), we have a uniformly consistent estimator $\widehat{\Sigma}_n$ for $\Sigma$, and $\Sigma$ satisfies mild regularity conditions, our finite-sample results in the normal model (6) translate to uniform asymptotic results. These unifomity results apply to level maximization settings without any restrictions on the behavior of $(\mu_{X,n},\mu_{Y,n})$. In norm maxmization settings, by contrast, we limit attention to $(\mu_{X,n},\mu_{Y,n})$ lying in bounded sets, since this is the context for which the asymptotic results of Elliott and Müller (2007, 2014) and Wang (2018) imply an asymptotic norm-maximization representation.[19]

# 4 Conditional Inference

This section develops conditional inference procedures for our general setting. We seek confidence sets with correct coverage conditional on $(\hat{\theta},\hat{\gamma})$,

$$Pr_\mu\Big\{\mu_Y(\hat{\theta})\in CS|\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma}\Big\}\geq 1-\alpha \text{ for all } \tilde{\theta}\in\Theta,\ \tilde{\gamma}\in\Gamma, \text{ and all } \mu. \tag{15}$$

As in the stylized example of Section 2, we consider both equal-tailed and uniformly most accurate unbiased confidence sets.[20] We also derive optimal conditionally $\alpha$-quantile-unbiased estimators, which for $\alpha\in(0,1)$ satisfy

$$Pr_\mu\Big\{\hat{\mu}_\alpha\geq\mu_Y(\hat{\theta})|\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma}\Big\}=\alpha \text{ for all } \tilde{\theta}\in\Theta,\ \tilde{\gamma}\in\Gamma, \text{ and all } \mu. \tag{16}$$

[18] In a manuscript circulated after the initial public version of this paper, Hyun et al. (2018) consider the related problem of conditional inference for changepoint detection, but the changepoint estimation methods they consider cannot be cast as norm-maximization, so their results do not overlap with ours.

[19] If one instead considers cases where $(\mu_{X,n},\mu_{Y,n})$ diverges, as occurs for example in threshold regression with non-vanishing parameter instability, the problem reduces to level-maximization asymptotically.

[20] If $\hat{\theta}$ is not unique we change the conditioning event $\hat{\theta}=\tilde{\theta}$ to $\tilde{\theta}\in\operatorname{argmax} X(\theta)$ or $\tilde{\theta}\in\operatorname{argmax}\|X(\theta)\|$ for the level and norm maximization problems, respectively.

Our conditional procedures depend on the conditioning events of interest. We analyze these conditioning events for our general level and norm maximization settings, and illustrate them in our EWM and threshold regression examples. We then discuss conventional sample splitting as an alternative conditional approach and briefly discuss the construction of dominating procedures. Finally, we show that our conditional procedures converge to conventional ones when $Pr_\mu\left\{\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma}\right\}\to 1$ so the latter are valid.

### 4.1 Optimal Conditional Inference

Since $\hat{\theta}$ and $\hat{\gamma}$ are functions of $X$, we can re-write the conditioning event in terms of the sample space of $X$ as $\left\{X:\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma}\right\}=\mathcal{X}(\tilde{\theta},\tilde{\gamma})$. Thus, for conditional inference we are interested in the distribution of $(X,Y)$ conditional on $X\in\mathcal{X}(\tilde{\theta},\tilde{\gamma})$. Our results below imply that under mild conditions, the elements of $Y$ other than $Y(\tilde{\theta})$ do not help in constructing a quantile-unbiased estimate or unbiased confidence set for $\mu_Y(\hat{\theta})$ conditional on $X\in\mathcal{X}(\tilde{\theta},\tilde{\gamma})$. Hence, we limit attention to the conditional distribution of $(X,Y(\tilde{\theta}))$ given $X\in\mathcal{X}(\tilde{\theta},\tilde{\gamma})$.

Since $(X,Y(\tilde{\theta}))$ is jointly normal unconditionally, it has a multivariate truncated normal distribution conditional on $X\in\mathcal{X}(\tilde{\theta},\tilde{\gamma})$. Correlation between $X$ and $Y(\tilde{\theta})$ implies that the conditional distribution of $Y(\tilde{\theta})$ depends on both the parameter of interest $\mu_Y(\hat{\theta})$ and $\mu_X$. To eliminate dependence on the nuisance parameter $\mu_X$, we condition on a sufficient statistic. Without truncation and for any fixed $\mu_Y(\tilde{\theta})$, a minimal sufficient statistic for $\mu_X$ is

$$Z_{\tilde{\theta}}=X-\Big(\Sigma_{XY}(\cdot,\tilde{\theta})/\Sigma_Y(\tilde{\theta})\Big)Y(\tilde{\theta}), \tag{17}$$

where we use $\Sigma_{XY}(\cdot,\tilde{\theta})$ to denote $Cov(X,Y(\tilde{\theta}))$. $Z_{\tilde{\theta}}$ corresponds to the part of $X$ that is (unconditionally) orthogonal to $Y(\tilde{\theta})$ which, since $(X,Y(\tilde{\theta}))$ are jointly normal, means that $Z_{\tilde{\theta}}$ and $Y(\tilde{\theta})$ are independent. Truncation breaks this independence, but $Z_{\tilde{\theta}}$ remains minimal sufficient for $\mu_X$. The conditional distribution of $Y(\hat{\theta})$ given $\left\{\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma},Z_{\tilde{\theta}}=z\right\}$ is truncated normal:

$$Y(\hat{\theta})|\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma},Z=z\sim\xi|\xi\in\mathcal{Y}(\tilde{\theta},\tilde{\gamma},z), \tag{18}$$

where $\xi\sim N\Big(\mu_Y(\tilde{\theta}),\Sigma_Y(\tilde{\theta})\Big)$ is normally distributed and

$$\mathcal{Y}(\tilde{\theta},\tilde{\gamma},z)=\left\{y:z+\Big(\Sigma_{XY}(\cdot,\tilde{\theta})/\Sigma_Y(\tilde{\theta})\Big)y\in\mathcal{X}(\tilde{\theta},\tilde{\gamma})\right\} \tag{19}$$

is the set of values for $Y(\tilde{\theta})$ such that the implied $X$ falls in $\mathcal{X}(\tilde{\theta},\tilde{\gamma})$ given $Z_{\tilde{\theta}}=z$. Thus, conditional on $\hat{\theta}=\tilde{\theta}$, $\hat{\gamma}=\tilde{\gamma}$, and $Z_{\tilde{\theta}}=z$, $Y(\hat{\theta})$ follows a one-dimensional truncated normal

distribution with truncation set $\mathcal{Y}(\tilde{\theta},\tilde{\gamma},z)$.

Using this result, it is straightforward to construct quantile-unbiased estimators for $\mu_Y(\hat{\theta})$. Let $F_{TN}(y;\mu_Y(\tilde{\theta}),\tilde{\theta},\tilde{\gamma},z)$ denote the distribution function for the truncated normal distribution (18). This distribution function is strictly decreasing in $\mu_Y(\tilde{\theta})$. Define $\hat{\mu}_\alpha$ as the unique solution to

$$F_{TN}(Y(\hat{\theta});\hat{\mu}_\alpha,\tilde{\theta},\tilde{\gamma},Z_{\tilde{\theta}})=1-\alpha. \tag{20}$$

Proposition 1 below shows that $\hat{\mu}_\alpha$ is conditionally $\alpha$-quantile-unbiased in the sense of (16), so $\hat{\mu}_{\frac{1}{2}}$ is median-unbiased while the equal-tailed interval $CS_{ET}=\left[\hat{\mu}_{\alpha/2},\hat{\mu}_{1-\alpha/2}\right]$ has conditional coverage $1-\alpha$. Moreover, results in Pfanzagl (1979) and Pfanzagl (1994) on quantile-unbiased estimation in exponential families imply that $\hat{\mu}_\alpha$ is optimal in the class of quantile-unbiased estimators.

To establish optimality, we add the following assumption:

**Assumption 1**

*If $\Sigma=Cov((X',Y')')$ has full rank, then the parameter space for $\mu$ is open and convex. Otherwise, there exists some $\mu^*$ such that the parameter space for $\mu$ is an open convex subset of $\left\{\mu^*+\Sigma^{\frac{1}{2}}v:v\in\mathbb{R}^{\dim(X,Y)}\right\}$ where $\Sigma^{\frac{1}{2}}$ is the symmetric square root of $\Sigma$.*

This assumption requires that the parameter space for $\mu$ be sufficiently rich.[21] When $\Sigma$ is degenerate (for example when $X$ and $Y$ are perfectly correlated as in the EWM example with $X=Y$), this assumption further implies that $(X,Y)$ have the same support for all values of $\mu$. This rules out cases in which some a pair of parameter values $\mu_1$, $\mu_2$ can be perfectly distinguished based on the data. Under this assumption, $\hat{\mu}_\alpha$ is an optimal quantile-unbiased estimator.

**Proposition 1**

*Let $\hat{\mu}_\alpha$ be the unique solution of (20). $\hat{\mu}_\alpha$ is conditionally $\alpha$-quantile-unbiased in the sense of (16). If Assumption 1 holds, then $\hat{\mu}_\alpha$ is the uniformly most concentrated $\alpha$-quantile-unbiased estimator in that for any other conditionally $\alpha$-quantile-unbiased estimator $\hat{\mu}^*_\alpha$ and any loss function $L\Big(d,\mu_Y(\tilde{\theta})\Big)$ that attains its minimum at $d=\mu_Y(\tilde{\theta})$ and is quasiconvex in $d$ for all $\mu_Y(\tilde{\theta})$,*

$$E_\mu\left[L\Big(\hat{\mu}_\alpha,\mu_Y(\tilde{\theta})\Big)|\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma}\right]\leq E_\mu\left[L\Big(\hat{\mu}^*_\alpha,\mu_Y(\tilde{\theta})\Big)|\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma}\right]$$

[21] The assumption that the parameter space is open can be relaxed at the cost of complicating the statements below.

*for all* $\mu$ *and all* $\tilde{\theta}\in\Theta$, $\tilde{\gamma}\in\Gamma$.

Proposition 1 shows that $\hat{\mu}_\alpha$ is optimal in the strong sense that it has lower risk (expected loss) than any other quantile-unbiased estimator for a large class of loss functions.

Rather than considering equal-tailed intervals, we can alternatively consider unbiased confidence sets. Following Lehmann and Romano (2005), we say that a level $1-\alpha$ two-sided confidence set $CS$ is unbiased if its probability of covering any given false parameter value is bounded above by $1-\alpha$. Likewise, a one sided lower (upper) confidence set is unbiased if its probability of covering a false parameter value above (below) the true value is bounded above by $1-\alpha$. Using the duality between tests and confidence sets, a level $1-\alpha$ confidence set $CS$ is unbiased if and only if $\phi(\mu_{Y,0})=1\{\mu_{Y,0}\notin CS\}$ is an unbiased test for the corresponding family of hypotheses.[22] The results of Lehmann and Scheffé (1955) applied in our setting imply that optimal unbiased tests conditional on $\left\{\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma}\right\}$ are the same as optimal unbiased tests conditional on $\left\{\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma},Z_{\tilde{\theta}}=z_{\tilde{\theta}}\right\}$. These optimal tests take a simple form.

Define a size $\alpha$ test of the two-sided hypothesis $H_0:\mu_Y(\tilde{\theta})=\mu_{Y,0}$ as

$$\phi_{TS,\alpha}(\mu_{Y,0})=1\left\{Y(\tilde{\theta})\notin[c_l(Z_{\tilde{\theta}}),c_u(Z_{\tilde{\theta}})]\right\} \tag{21}$$

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

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

for $\zeta$ that follows a truncated normal distribution

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

Likewise, define a size $\alpha$ test of the one-sided hypothesis $H_0:\mu_Y(\tilde{\theta})\geq\mu_{Y,0}$ as

$$\phi_{OS-,\alpha}(\mu_{Y,0})=1\left\{F_{TN}(Y(\tilde{\theta});\mu_{Y,0},\tilde{\theta},\tilde{\gamma},z)\leq\alpha\right\} \tag{22}$$

and a test of $H_0:\mu_Y(\tilde{\theta})\leq\mu_{Y,0}$ as

$$\phi_{OS+,\alpha}(\mu_{Y,0})=1\left\{F_{TN}(Y(\tilde{\theta});\mu_{Y,0},\tilde{\theta},\tilde{\gamma},z)\geq1-\alpha\right\}. \tag{23}$$

[22] That is, $H_0:\mu_Y(\tilde{\theta})=\mu_{Y,0}$ for a two-sided confidence set, $H_0:\mu_Y(\tilde{\theta})\geq\mu_{Y,0}$ for a lower confidence set and $H_0:\mu_Y(\tilde{\theta})\leq\mu_{Y,0}$ for an upper confidence set.

**Proposition 2**

*If Assumption 1 holds, $\phi_{TS,\alpha}$, $\phi_{OS-,\alpha}$, and $\phi_{OS+,\alpha}$ are uniformly most powerful unbiased size $\alpha$ tests of their respective null hypotheses conditional on $\hat{\theta}=\tilde{\theta}$ and $\hat{\gamma}=\tilde{\gamma}$.*

To form uniformly most accurate unbiased confidence sets we collect the values not rejected by these tests. The two-sided uniformly most accurate unbiased confidence set is $CS_U=\{\mu_{Y,0}:\phi_{TS,\alpha}(\mu_{Y,0})=0\}$. $CS_U$ is unbiased and has conditional coverage $1-\alpha$ by construction. Likewise, we can form lower and upper one-sided uniformly most accurate unbiased confidence intervals as $CS_{U,-}=\{\mu_{Y,0}:\phi_{OS-,\alpha}(\mu_{Y,0})=0\}=(-\infty,\hat{\mu}_{1-\alpha}]$, and $CS_{U,+}=\{\mu_{Y,0}:\phi_{OS+,\alpha}(\mu_{Y,0})=0\}=[\hat{\mu}_{\alpha},\infty)$, respectively. Hence, we can view $CS_{ET}$ as the intersection of level $1-\frac{\alpha}{2}$ uniformly most accurate unbiased upper and lower confidence intervals. Unfortunately, no such simplification is generally available for $CS_U$, though Lemma 5.5.1 of Lehmann and Romano (2005) guarantees that this set is an interval.

### 4.2 Conditioning Sets

Thus far we have left the conditioning events $\mathcal{X}(\tilde{\theta},\tilde{\gamma})$ and $\mathcal{Y}(\tilde{\theta},\tilde{\gamma},z)$ abstract. To implement our conditional procedures, however, we need tractable representations of $\mathcal{Y}(\tilde{\theta},\tilde{\gamma},z)$. We first derive the form of this conditioning event for the level maximization problem (7) and the norm maximization problem (8) without additional conditioning variables $\hat{\gamma}$. We then discuss the effect of adding conditioning variables and illustrate in our examples.

In level maximization problems without additional conditioning variables, we are interested in inference conditional on $X\in\mathcal{X}(\tilde{\theta})$ for $\mathcal{X}(\tilde{\theta})=\left\{X:X(\tilde{\theta})=\max_{\theta\in\Theta}X(\theta)\right\}$. The following result, based on Lemma 5.1 of Lee et al. (2016), derives $\mathcal{Y}(\tilde{\theta},z)$ in this setting.

**Proposition 3**

*Let $\Sigma_{XY}(\tilde{\theta})=Cov(X(\tilde{\theta}),Y(\tilde{\theta}))$. Define*

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

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

*and*

$$\mathcal{V}(\tilde{\theta},Z_{\tilde{\theta}})=\min_{\theta\in\Theta:\Sigma_{XY}(\tilde{\theta})=\Sigma_{XY}(\tilde{\theta},\theta)}-\left(Z_{\tilde{\theta}}(\theta)-Z_{\tilde{\theta}}(\tilde{\theta})\right).$$

*If $\mathcal{V}(\tilde{\theta},z)\geq 0$, then $\mathcal{Y}(\tilde{\theta},z)=\left[\mathcal{L}(\tilde{\theta},z),\mathcal{U}(\tilde{\theta},z)\right]$. If $\mathcal{V}(\tilde{\theta},z)<0$, then $\mathcal{Y}(\tilde{\theta},z)=\emptyset$.*

Thus, the conditioning event $\mathcal{Y}(\tilde{\theta},z)$ is an interval bounded above and below by easy-to-calculate functions of $z$. While we must have $\mathcal{V}(\tilde{\theta},z)\geq 0$ for this interval to be non-empty, $Pr_{\mu}\Big\{\mathcal{V}(\hat{\theta},Z_{\hat{\theta}})<0\Big\}=0$ for all $\mu$ so this constraint holds almost surely when we consider the value $\hat{\theta}$ observed in the data. Hence, in applications we can safely ignore this constraint and calculate only $\mathcal{L}(\hat{\theta},Z_{\hat{\theta}})$ and $\mathcal{U}(\hat{\theta},Z_{\hat{\theta}})$.

The norm maximization conditioning event is $\mathcal{X}(\tilde{\theta})=\Big\{X:\|X(\tilde{\theta})\|=\max_{\theta\in\Theta}\|X(\theta)\|\Big\}$. This conditioning event involves nonlinear constraints so the results of Lee et al. (2016) do not apply. The expression for $\mathcal{Y}(\tilde{\theta},z)$ is more involved, but remains easy to calculate.

**Proposition 4**

*Define*

$$A(\tilde{\theta},\theta)=\Sigma_Y(\tilde{\theta})^{-2}\sum_{i=1}^{d_X}\Big[\Sigma_{XY,i}(\tilde{\theta})^2-\Sigma_{XY,i}(\theta,\tilde{\theta})^2\Big],$$

$$B_Z(\tilde{\theta},\theta)=2\Sigma_Y(\tilde{\theta})^{-1}\sum_{i=1}^{d_X}\Big[\Sigma_{XY,i}(\tilde{\theta})Z_{\tilde{\theta},i}(\tilde{\theta})-\Sigma_{XY,i}(\theta,\tilde{\theta})Z_{\tilde{\theta},i}(\theta)\Big],$$

$$C_Z(\tilde{\theta},\theta)=\sum_{i=1}^{d_X}\Big[Z_{\tilde{\theta},i}(\tilde{\theta})^2-Z_{\tilde{\theta},i}(\theta)^2\Big].$$

*For*

$$D_Z(\tilde{\theta},\theta)=B_Z(\tilde{\theta},\theta)^2-4A(\tilde{\theta},\theta)C_Z(\tilde{\theta},\theta),\ H_Z(\tilde{\theta},\theta)=\frac{-C_Z(\tilde{\theta},\theta)}{B_Z(\tilde{\theta},\theta)},$$

$$G_Z(\tilde{\theta},\theta)=\frac{-B_Z(\tilde{\theta},\theta)-\sqrt{D_Z(\tilde{\theta},\theta)}}{2A(\tilde{\theta},\theta)},\ \textit{and}\ K_Z(\tilde{\theta},\theta)=\frac{-B_Z(\tilde{\theta},\theta)+\sqrt{D_Z(\tilde{\theta},\theta)}}{2A(\tilde{\theta},\theta)},$$

*define*

$$\ell_Z^1(\tilde{\theta})=\max\Bigg\{\max_{\theta\in\Theta:A(\tilde{\theta},\theta)<0,D_Z(\tilde{\theta},\theta)\geq 0}G_Z(\tilde{\theta},\theta),\max_{\theta\in\Theta:A(\tilde{\theta},\theta)=0,B_Z(\tilde{\theta},\theta)>0}H_Z(\tilde{\theta},\theta)\Bigg\},$$

$$\ell_Z^2(\tilde{\theta},\theta)=\max\Bigg\{\max_{\theta\in\Theta:A(\tilde{\theta},\theta)<0,D_Z(\tilde{\theta},\theta)\geq 0}G_Z(\tilde{\theta},\theta),\max_{\theta\in\Theta:A(\tilde{\theta},\theta)=0,B_Z(\tilde{\theta},\theta)>0}H_Z(\tilde{\theta},\theta),G_Z(\tilde{\theta},\theta)\Bigg\},$$

$$u_Z^1(\tilde{\theta},\theta)=\min\Bigg\{\min_{\theta\in\Theta:A(\tilde{\theta},\theta)<0,D_Z(\tilde{\theta},\theta)\geq 0}K_Z(\tilde{\theta},\theta),\min_{\theta\in\Theta:A(\tilde{\theta},\theta)=0,B_Z(\tilde{\theta},\theta)<0}H_Z(\tilde{\theta},\theta),K_Z(\tilde{\theta},\theta)\Bigg\},$$

$$u_Z^2(\tilde{\theta})=\min\Bigg\{\min_{\theta\in\Theta:A(\tilde{\theta},\theta)<0,D_Z(\tilde{\theta},\theta)\geq 0}K_Z(\tilde{\theta},\theta),\min_{\theta\in\Theta:A(\tilde{\theta},\theta)=0,B_Z(\tilde{\theta},\theta)<0}H_Z(\tilde{\theta},\theta)\Bigg\},$$

*and*

$$\mathcal{V}(\tilde{\theta},Z_{\tilde{\theta}})=\min_{\theta\in\Theta:A(\tilde{\theta},\theta)=B_Z(\tilde{\theta},\theta)=0 \text{ or } D_Z(\tilde{\theta},\theta)<0} C_Z(\tilde{\theta},\theta).$$

*If* $\mathcal{V}(\tilde{\theta},Z_{\tilde{\theta}})\geq 0$ *then*

$$\mathcal{Y}(\tilde{\theta},Z_{\tilde{\theta}})=\bigcap_{\theta\in\Theta:A(\tilde{\theta},\theta)>0,D_Z(\tilde{\theta},\theta)\geq 0}\Big[\ell_Z^1(\tilde{\theta}),u_Z^1(\tilde{\theta},\theta)\Big]\cup\Big[\ell_Z^2(\tilde{\theta},\theta),u_Z^2(\tilde{\theta})\Big].$$

*If* $\mathcal{V}(\tilde{\theta},Z_{\tilde{\theta}})<0$, *then* $\mathcal{Y}(\tilde{\theta},Z_{\tilde{\theta}})=\emptyset$.

While the expression for $\mathcal{Y}(\tilde{\theta},z)$ in this setting is long, it is easy to calculate in practice and can be expressed as a finite union of intervals using DeMorgan's laws. As before, $Pr_\mu\Big\{\mathcal{V}(\hat{\theta},Z_{\hat{\theta}})<0\Big\}=0$ for all $\mu$ so we can ignore this constraint in applications.

Our derivations have so far assumed we have no additional conditioning variables $\hat{\gamma}$. If we also condition on $\hat{\gamma}=\tilde{\gamma}$, then for $\mathcal{X}_\gamma(\tilde{\gamma})=\{X:\gamma(X)=\tilde{\gamma}\}$, we can write $\mathcal{X}(\tilde{\theta},\tilde{\gamma})=\mathcal{X}(\tilde{\theta})\cap\mathcal{X}_\gamma(\tilde{\gamma})$. Likewise, for $\mathcal{Y}_\gamma(\tilde{\gamma},z)$ defined analogously to (19), $\mathcal{Y}(\tilde{\theta},\tilde{\gamma},z)=\mathcal{Y}(\tilde{\theta},z)\cap\mathcal{Y}_\gamma(\tilde{\gamma},z)$. The form of $\mathcal{X}_\gamma(\tilde{\gamma})$ and $\mathcal{Y}_\gamma(\tilde{\gamma},z)$ depends on the conditioning variables $\hat{\gamma}$ considered. To illustrate we next discuss the effect of conditioning on the outcomes of pretests in our EWM and threshold regression examples.

**Empirical Welfare Maximization (continued)** Suppose that we report estimates and confidence sets for welfare only if the improvement in empirical welfare from the estimated optimal policy over a baseline policy $\theta=0$ exceeds a threshold $c$, i.e. $X(\hat{\theta})-X(0)\geq c$. For instance, we might report results only when the test of White (2000) rejects the null that no policy has performance exceeding the baseline, $H_0:\max_{\theta\in\Theta}\mu_X(\theta)\leq\mu_X(0)$. This implies that we report results only if $X(\hat{\theta})-X(0)\geq c$ for $c$ a critical value depending on $\Sigma$. We can set $\gamma(X)=1\Big\{X(\hat{\theta})-X(0)\geq c\Big\}$ and it is natural to condition inference on $\hat{\gamma}=1$.

Assuming $\Sigma_{XY}(\tilde{\theta})-\Sigma_{XY}(\tilde{\theta},0)>0$ for simplicity, the conditioning event in this setting is $\mathcal{X}_\gamma(1)=\Big\{X:X(\hat{\theta})-X(0)\geq c\Big\}$ and one can show that

$$\mathcal{Y}_\gamma(1,Z_{\tilde{\theta}})=\left\{y:y\geq\frac{\Sigma_Y(\tilde{\theta})\Big(c-Z_{\tilde{\theta}}(\tilde{\theta})+Z_{\tilde{\theta}}(0)\Big)}{\Sigma_{XY}(\tilde{\theta})-\Sigma_{XY}(\tilde{\theta},0)}\right\}.$$

See Section B.1 of the supplement for details, as well as expressions for other values of $\Sigma_{XY}(\tilde{\theta})-\Sigma_{XY}(\tilde{\theta},0)$. In the present case, provided $\mathcal{V}(\tilde{\theta},Z_{\tilde{\theta}})\geq 0$, $\mathcal{Y}(\tilde{\theta},1,Z_{\tilde{\theta}})=$

$\left[\mathcal{L}^*(\tilde{\theta},Z_{\tilde{\theta}}),\mathcal{U}(\tilde{\theta},Z_{\tilde{\theta}})\right]$, where $\mathcal{U}(\tilde{\theta},Z_{\tilde{\theta}})$ is the upper bound derived in Proposition 3 while

$$\mathcal{L}^*(\tilde{\theta},Z_{\tilde{\theta}})=\max\left\{\mathcal{L}(\tilde{\theta},Z_{\tilde{\theta}}),\frac{\Sigma_Y(\tilde{\theta})\left(c-Z_{\tilde{\theta}}(\tilde{\theta})+Z_{\tilde{\theta}}(0)\right)}{\Sigma_{XY}(\tilde{\theta})-\Sigma_{XY}(\tilde{\theta},0)}\right\},$$

for $\mathcal{L}(\tilde{\theta},Z_{\tilde{\theta}})$ defined as in Proposition 3. Hence, when $\Sigma_{XY}(\tilde{\theta})-\Sigma_{XY}(\tilde{\theta},0)>0$, conditoning on $\hat{\gamma}=1$ simply modifies the lower bound $\mathcal{L}(\tilde{\theta},Z_{\tilde{\theta}})$. Likewise, when $\Sigma_{XY}(\tilde{\theta})-\Sigma_{XY}(\tilde{\theta},0)<0$ or $\Sigma_{XY}(\tilde{\theta})-\Sigma_{XY}(\tilde{\theta},0)=0$, conditioning on $\hat{\gamma}=1$ modifies $\mathcal{U}(\tilde{\theta},Z_{\tilde{\theta}})$ and $\mathcal{V}(\tilde{\theta},Z_{\tilde{\theta}})$, respectively. $\triangle$

**Threshold Regression Estimation (continued)** Suppose that we report estimates and confidence sets for the change parameter $\delta_j(\hat{\theta})$ only if we reject the null hypothesis of no threshold, $H_0:\delta(\theta)=0$ for all $\theta\in\Theta$. Suppose, in particular, that we test this hypothesis with the sup-Wald test of Andrews (1993). Analogous results to those shown in Elliott and Müller (2014) provide that in our setting, such a test rejects asymptotically if and only if $\|X(\hat{\theta})\|>c$ for a critical value $c$ that depends on $\Sigma$. We can set $\gamma(X)=1\left\{\|X(\hat{\theta})\|>c\right\}$ and it is again natural to condition inference on $\hat{\gamma}=1$.

In this setting $\mathcal{X}_\gamma(1)=\left\{X:\|X(\hat{\theta})\|>c\right\}$. As before, the expressions for the conditioning sets are involved but straightforward to compute. In particular, for $\bar{\mathcal{V}}(Z_{\tilde{\theta}})$, $\bar{\mathcal{L}}(Z_{\tilde{\theta}})$, and $\bar{\mathcal{U}}(Z_{\tilde{\theta}})$ defined in Section B.2 of the supplement, if $\bar{\mathcal{V}}(Z_{\tilde{\theta}})\geq 0$ then $\mathcal{Y}_\gamma(1,Z_{\tilde{\theta}})=\left(\bar{\mathcal{L}}(Z_{\tilde{\theta}}),\bar{\mathcal{U}}(Z_{\tilde{\theta}})\right)^c$, where $\mathcal{S}^c$ denotes the complement of a generic set $\mathcal{S}$. Thus,

$$\mathcal{Y}(\tilde{\theta},1,Z_{\tilde{\theta}})=\left(\bar{\mathcal{L}}(Z_{\tilde{\theta}}),\bar{\mathcal{U}}(Z_{\tilde{\theta}})\right)^c\cap\bigcap_{\theta\in\Theta:A(\tilde{\theta},\theta)>0,D_Z(\tilde{\theta},\theta)\geq 0}\left[\ell_Z^1(\tilde{\theta}),u_Z^1(\tilde{\theta},\theta)\right]\cup\left[\ell_Z^2(\tilde{\theta},\theta),u_Z^2(\tilde{\theta})\right]$$

when $\min\left\{\mathcal{V}(\tilde{\theta},Z_{\tilde{\theta}}),\bar{\mathcal{V}}(Z_{\tilde{\theta}})\right\}\geq 0$. Details and expressions under other realizations of $\bar{\mathcal{V}}(Z_{\tilde{\theta}})$ can be found in Section B.2 of the supplement. $\triangle$

As these example illustrate, it is straightforward to incorporate additional conditioning variables $\hat{\gamma}$ in both the level and norm maximization problems provided one can characterize the set $\mathcal{Y}_\gamma(\tilde{\gamma},z)$. While such characterizations are easy to obtain in many cases, they depend on the conditioning variable considered and must be derived on a case-by-case basis.

### 4.3 Comparison to Sample Splitting

A common remedy in practice for the problems we study is to split the sample. If we have iid observations and select $\hat{\theta}^1$ based on the first half of the data, conventional estimates and confidence intervals for $\mu_Y(\hat{\theta}^1)$ that use only the second half of the data will be

(conditionally) valid. Hence, it is natural to ask how our conditioning approach compares to this conventional sample splitting approach.

For ease of exposition, in this section we focus on even sample splits. Asymptotically, such splits yield a pair of independent and identically distributed normal draws $(X^1,Y^1)$ and $(X^2,Y^2)$, both of which follow (6), albeit with a different scaling for $(\mu,\Sigma)$ than in the full-sample case.[23] Sample splitting procedures calculate $\hat{\theta}^1$ as in (7) and (8) for level and norm maximization, respectively, replacing $X$ by $X^1$. Inference on $\mu_Y(\hat{\theta}^1)$ is then conducted using $(X^2,Y^2)$. In particular, the conventional 95% sample-splitting confidence interval for $\mu_Y(\hat{\theta}^1)$,

$$\left[Y^2(\hat{\theta}^1)-1.96\sqrt{\Sigma_Y(\hat{\theta}^1)},Y^2(\hat{\theta}^1)+1.96\sqrt{\Sigma_Y(\hat{\theta}^1)}\right],$$

has correct (conditional) coverage and $Y^2(\hat{\theta}^1)$ is a median-unbiased estimator for $\mu_Y(\hat{\theta}^1)$.

While conventional sample splitting resolves the inference problem, this comes at a cost. First, $\hat{\theta}^1$ is based on less data than in the full-sample case, which is unappealing since a policy recommendation estimated with a smaller sample size leads to a lower expected welfare (see, e.g., Theorems 2.1 and 2.2 in Kitagawa and Tetenov (2018b)). Moreover, even after conditioning on $\hat{\theta}^1$, the full-sample average $\frac{1}{2}(X^1,Y^1)+\frac{1}{2}(X^2,Y^2)$ remains a minimal sufficient statistic for $\mu$. Hence, using only $(X^2,Y^2)$ for inference sacrifices information.

Fithian et al. (2017) formalize this point and show that conventional sample splitting tests (and thus confidence sets) are inadmissible.[24] Motivated by this result, in Section C of the supplement we derive optimal confidence sets and estimates that are valid conditional on $\hat{\theta}^1$. These optimal split-sample procedures involve truncated normal distributions which are difficult to compute, however, so we also propose computationally straightforward alternatives. These alternatives dominate conventional split-sample methods, but are in turn dominated by the (computationally intractable) optimal split-sample procedures. Nevertheless, these computationally straightforward alternative procedures dominate their conventional counterparts by a substantial margin in simulations calibrated to Card et al. (2008) and reported in Section 7.

Splitting the sample changes the target parameter from $\mu_Y(\hat{\theta})$ to $\mu_Y(\hat{\theta}^1)$, so split-sample

[23]Section C of the supplement considers cases with general sample splits and describes the scaling for $(\mu,\Sigma)$. Intuitively, the scope for improvement over conventional split-sample inference is increasing in the fraction of the data used to construct $X_1$.

[24]Corollary 1 of Fithian et al. (2017) applied in our setting shows that for any sample splitting test based on $Y^2$, there exists a test that uses the full data and has weakly higher power against all alternatives and strictly higher power against some alternatives.

approaches are not directly comparable to our full-sample conditioning approach developed above. Nonetheless, while conventional sample splitting methods are dominated, calculating $\hat{\theta}^1$ based on only part of the data may increase the amount of information available for inference and so allow tighter confidence intervals. Thus, depending on how we weight noisier choices of $\theta$ against more precise inference on $\mu_Y(\hat{\theta})$, it may be helpful to split the sample and use a procedure that dominates conventional split-sample inference. See Tian and Taylor (2016) and Tian et al. (2016) for related discussions.

### 4.4 Behavior When $Pr_\mu\left\{\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma}\right\}$ is Large

As discussed in Section 2, if we ignore selection and compute the conventional (or "naive") estimator $\hat{\mu}_N=Y(\hat{\theta})$ and the conventional confidence set

$$CS_N=\left[Y(\hat{\theta})-c_{\alpha/2,N}\sqrt{\Sigma_Y(\hat{\theta})},Y(\hat{\theta})+c_{\alpha/2,N}\sqrt{\Sigma_Y(\hat{\theta})}\right] \tag{24}$$

where $c_{\alpha,N}$ is the $1-\alpha$-quantile of the standard normal distribution, $\hat{\mu}_N$ is biased and $CS_N$ has incorrect coverage conditional on $\hat{\theta}=\tilde{\theta}$, $\hat{\gamma}=\tilde{\gamma}$. These biases are mild when $Pr_\mu\left\{\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma}\right\}$ is close to one, however, since in this case the conditional distribution is close to the unconditional one. Intuitively, $Pr_\mu\left\{\hat{\theta}=\tilde{\theta}\right\}$ is close to one for some $\tilde{\theta}$ when $\mu_X(\theta)$ or $\|\mu_X(\theta)\|$ has a well-separated maximum in the level and norm maximization problems, respectively. This section shows that our procedures converge to conventional ones in this case.

In particular, suppose first that for some sequence of values $\mu_{Y,m}$ and $z_{\tilde{\theta},m}$ the probability that $\hat{\theta}=\tilde{\theta}$ and $\hat{\gamma}=\tilde{\gamma}$, conditional on $Z_{\tilde{\theta}}=z_{\tilde{\theta},m}$, converges to one as $m\to\infty$. Then our conditional confidence sets and estimates converge to the usual confidence sets and estimates.

**Lemma 1**

*Consider any sequence of values $\mu_{Y,m}$ and $z_{\tilde{\theta},m}$ such that $Pr_{\mu_{Y,m}}\left\{\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma}|Z_{\tilde{\theta}}=z_{\tilde{\theta},m}\right\}\to 1$. Then under $\mu_{Y,m}$, conditional on $\left\{\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma},Z_{\tilde{\theta}}=z_{\tilde{\theta},m}\right\}$ we have $CS_U\to_p CS_N$, $CS_{ET}\to_p CS_N$, and $\hat{\mu}_{\frac{1}{2}}\to_p Y(\tilde{\theta})$, where for confidence sets $\to_p$ denotes convergence in probability of the endpoints.*

Lemma 1 discusses probabilities conditional on $Z_{\tilde{\theta}}$. If we consider a sequence of values $\mu_m$ such that $Pr_{\mu_m}\left\{\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma}\right\}\to_p 1$, the same result holds when conditioning only on $\left\{\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma}\right\}$ and unconditionally.

**Proposition 5**

*Consider any sequence of values $\mu_m$ such that $Pr_{\mu_m}\left\{\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma}\right\}\to 1$. Then under $\mu_m$,*

*we have* $CS_U \to_p CS_N$, $CS_{ET} \to_p CS_N$, *and* $\hat{\mu}_{\frac{1}{2}} \to_p Y(\tilde{\theta})$ *both conditional on* $\left\{\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma}\right\}$ *and unconditionally.*

These results provide an additional argument for using our procedures: they remain valid when conventional procedures fail, but coincide with conventional procedures when the latter are valid. On the other hand, as we saw in Section 2, there are cases where our conditional procedures have poor unconditional performance.

# 5 Unconditional Inference

Rather than requiring validity conditional on $(\hat{\theta},\hat{\gamma})$ we can instead require coverage only on average, yielding the unconditional coverage requirement

$$Pr\Big\{\mu(\hat{\theta}) \in CS\Big\} \geq 1-\alpha \text{ for all } \mu. \tag{25}$$

All confidence sets with correct conditional coverage in the sense of (15) also have correct unconditional coverage provided $\hat{\theta}$ is unique with probability one.

**Proposition 6**
*Suppose that* $\hat{\theta}$ *is unique with probability one for all* $\mu$. *Then any confidence set* $CS$ *with correct conditional coverage (15) also has correct unconditional coverage (25).*

Uniqueness of $\hat{\theta}$ implies that the conditioning events $\mathcal{X}(\tilde{\theta},\tilde{\gamma})$ partition the support of $X$ with measure zero overlap. The result then follows from the law of iterated expectations.

A sufficient condition for almost sure uniqueness of $\hat{\theta}$ is that $\Sigma_X$ has full rank. A weaker sufficient condition is given in the next lemma. Cox (2018) gives sufficient conditions for uniqueness of a global optimum in a much wider class of problems.

**Lemma 2**
*Suppose that for all* $\theta, \tilde{\theta} \in \Theta$ *such that* $\theta \neq \tilde{\theta}$, *either* $Var\Big(X(\theta)|X(\tilde{\theta})\Big) \neq 0$ *or* $Var\Big(X(\tilde{\theta})|X(\theta)\Big) \neq 0$. *Then* $\hat{\theta}$ *is unique with probability one for all* $\mu$.

While the conditional confidence sets derived in the last section are unconditionally valid, unconditional coverage is less demanding than conditional coverage. Hence, if we are only concerned with unconditional coverage, relaxing the coverage requirement from (15) to (25) may allow us to obtain shorter confidence sets in some settings.

In this section we explore the benefits of such a relaxation. We begin by introducing unconditional confidence sets based on projections of simultaneous confidence bands for

$\mu$. We then introduce hybrid confidence sets that combine projection confidence sets with conditioning arguments. We do not know of estimators for $\mu_Y(\hat{\theta})$ that are unconditionally $\alpha$-quantile-unbiased but not conditionally unbiased, but introduce hybrid estimators which substantially reduce variability at the cost of permitting a small unconditional bias.

### 5.1 Projection Confidence Sets

One approach to obtain an unconditional confidence set for $\mu_Y(\hat{\theta})$ is to start with a joint confidence set for $\mu$ and project on the dimension corresponding to $\hat{\theta}$. This approach was used by Kitagawa and Tetenov (2018a) for inference in EWM, and by Romano and Wolf (2005) in the context of multiple testing. This approach has also been used in a large and growing statistics literature on post-selection inference including e.g. Berk et al. (2013), Bachoc et al. (2017), Kuchibhotla et al. (2018), and Bachoc et al. (2018). Laber and Murphy (2011) consider a variant of projection for inference on the generalization error of an estimated classifier, obtaining a smaller critical value via a first-stage pretest with a divergent critical value.

To formally describe the projection approach, let $c_\alpha$ denote the $1-\alpha$ quantile of $\max_\theta |\xi(\theta)|/\sqrt{\Sigma_Y(\theta)}$ for $\xi \sim N(0,\Sigma_Y)$. If we define

$$CS_\mu = \left\{\mu : |Y(\theta) - \mu_Y(\theta)| \leq c_\alpha \sqrt{\Sigma_Y(\theta)} \text{ for all } \theta \in \Theta\right\},$$

then $CS_\mu$ is a level $1-\alpha$ confidence set for $\mu$.[25] If we then define

$$CS_P = \left\{\tilde{\mu}_Y(\hat{\theta}) : \exists \mu \in CS_\mu \text{ such that } \mu_Y(\hat{\theta}) = \tilde{\mu}_Y(\hat{\theta})\right\} = \left[Y(\hat{\theta}) - c_\alpha\sqrt{\Sigma_Y(\hat{\theta})}, Y(\hat{\theta}) + c_\alpha\sqrt{\Sigma_Y(\hat{\theta})}\right]$$

as the projection of $CS_\mu$ on the parameter space for $\mu_Y(\hat{\theta})$, then since $\mu \in CS_\mu$ implies $\mu_Y(\hat{\theta}) \in CS_P$, $CS_P$ satisfies the unconditional coverage requirement (25). As noted in Section 2, however, $CS_P$ does not generally have correct conditional coverage.

The width of the confidence set $CS_P$ depends on the variance $\Sigma_Y(\hat{\theta})$ but does not otherwise depend on the data. To account for the randomness of $\hat{\theta}$, the critical value $c_\alpha$ is larger than the conventional two-sided normal critical value. This means that $CS_P$ will be conservative in cases where $\hat{\theta}$ takes a given value $\tilde{\theta}$ with high probability. To improve performance in this case, we next consider hybrid confidence sets.

[25] Note that we consider a studentized confidence band that adjusts the width based on $\Sigma_Y(\hat{\theta})$, while Kitagawa and Tetenov (2018a) consider an unstudentized band. Romano and Wolf (2005) argue for studentization in a closely related problem.

### 5.2 Hybrid Confidence Sets

As shown in Section 2, conditional and projection confidence sets each have good unconditional performance in some cases, but neither is fully satisfactory. Hybrid confidence sets combine these procedures to obtain good performance over a wide range of parameter values.

Hybrid confidence sets are constructed to be subsets of the level $1-\beta$ projection confidence set $CS_P^\beta$ for $0 \leq \beta < \alpha$. A hybrid confidence set collects the values $\mu_{Y,0} \in CS_P^\beta$ not rejected by a hybrid test. Like our conditional tests, hybrid tests of $H_0 : \mu_Y(\tilde{\theta}) = \mu_{Y,0}$ condition on $\left\{\hat{\theta} = \tilde{\theta}, \hat{\gamma} = \tilde{\gamma}\right\}$, but they further condition on the event that the null value is contained in the projection confidence set, i.e. $\mu_{Y,0} \in CS_P^\beta$. This changes the conditioning event to

$$\mathcal{Y}^H(\tilde{\theta},\tilde{\gamma},\mu_{Y,0},z) = \mathcal{Y}(\tilde{\theta},\tilde{\gamma},z) \cap \left[\mu_{Y,0} - c_\beta\sqrt{\Sigma_Y(\tilde{\theta})}, \mu_{Y,0} + c_\beta\sqrt{\Sigma_Y(\tilde{\theta})}\right]$$

for $c_\beta$ as defined in Section 5.1.

Similarly to our conditional confidence sets, we construct hybrid confidence sets by inverting both equal-tailed and uniformly most powerful unbiased hybrid tests. To construct the equal-tailed test, we define $\phi^H_{OS-,\alpha}$ and $\phi^H_{OS+,\alpha}$ analogously to $\phi_{OS-,\alpha}$ and $\phi_{OS+,\alpha}$ in (22) and (23), respectively, using the conditioning event $\mathcal{Y}^H(\tilde{\theta},\tilde{\gamma},\mu_{Y,0},Z_{\tilde{\theta}})$ rather than $\mathcal{Y}(\tilde{\theta},\tilde{\gamma},Z_{\tilde{\theta}})$. The equal-tailed hybrid test of $H_0 : \mu_Y(\tilde{\theta}) = \mu_{Y,0}$ is

$$\phi^H_{ET,\alpha}(\mu_{Y,0}) = \max\left\{\phi^H_{OS-,\alpha/2}(\mu_{Y,0}), \phi^H_{OS+,\alpha/2}(\mu_{Y,0})\right\},$$

which rejects if either of the upper or lower size $\alpha/2$ one-sided tests rejects. The level $1-\alpha$ equal-tailed hybrid confidence set is $CS^H_{ET} = \left\{\mu_{Y,0} \in CS_P^\beta : \phi^H_{ET,\frac{\alpha-\beta}{1-\beta}}(\mu_{Y,0}) = 0\right\}$, which collects the set of values in $CS_P^\beta$ which are not rejected by $\phi^H_{ET,\frac{\alpha-\beta}{1-\beta}}$.

To form a hybrid confidence set based on inverting unbiased tests, we likewise define $\phi^H_{TS,\alpha}$ analogously to $\phi_{TS,\alpha}$ in (21), using the conditioning event $\mathcal{Y}^H(\tilde{\theta},\tilde{\gamma},\mu_{Y,0},Z_{\tilde{\theta}})$ rather than $\mathcal{Y}(\tilde{\theta},\tilde{\gamma},Z_{\tilde{\theta}})$. By the results of Proposition 2, we know that $\phi^H_{TS,\alpha}(\mu_{Y,0})$ is the uniformly most powerful level $\alpha$ unbiased test of $H_0 : \mu_Y(\tilde{\theta}) = \mu_{Y,0}$ conditional on $\left\{\hat{\theta} = \tilde{\theta}, \hat{\gamma} = \tilde{\gamma}, \mu_{Y,0} \in CS_P^\beta\right\}$. The corresponding level $1-\alpha$ confidence set is then $CS^H_U = \left\{\mu_{Y,0} \in CS_P^\beta : \phi^H_{U,\frac{\alpha-\beta}{1-\beta}}(\mu_{Y,0}) = 0\right\}$.

For $\beta = 0$ the hybrid confidence sets coincide with the conditional confidence sets $CS_{ET}$ and $CS_U$. For $\beta > 0$ on the other hand, the hybrid confidence sets are contained in $CS_P^\beta$ and the level of hybrid tests that condition on $\left\{\hat{\theta} = \tilde{\theta}, \hat{\gamma} = \tilde{\gamma}, \mu_{Y,0} \in CS_P^\beta\right\}$ are correspondingly

adjusted to $\frac{\alpha-\beta}{1-\beta}$. This adjustment is necessary because the true value $\mu_Y(\hat{\theta})$ sometimes falls outside $CS_P^\beta$, and if we do not account for this our hybrid confidence sets may under-cover. With this adjustment, however, hybrid confidence sets have coverage at least $1-\alpha$ both conditionally and unconditionally.

**Proposition 7**

*The hybrid confidence sets $CS_{ET}^H$ and $CS_U^H$ have conditional coverage $\frac{1-\alpha}{1-\beta}$:*

$$Pr_\mu\left\{\mu(\tilde{\theta})\in CS_{ET}^H|\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma},\mu_Y(\tilde{\theta})\in CS_P^\beta\right\}=\frac{1-\alpha}{1-\beta},$$

$$Pr_\mu\left\{\mu(\tilde{\theta})\in CS_U^H|\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma},\mu_Y(\tilde{\theta})\in CS_P^\beta\right\}=\frac{1-\alpha}{1-\beta},$$

*for all $\tilde{\theta}\in\Theta$, $\tilde{\gamma}\in\Gamma$, and all $\mu$. Moreover, provided $\hat{\theta}$ is unique with probability one for all $\mu$, both confidence sets have unconditional coverage between $1-\alpha$ and $\frac{1-\alpha}{1-\beta}\leq 1-\alpha+\beta$:*

$$\inf_\mu Pr_\mu\left\{\mu(\hat{\theta})\in CS_{ET}^H\right\}\geq 1-\alpha,\ \ \sup_\mu Pr_\mu\left\{\mu(\hat{\theta})\in CS_{ET}^H\right\}\leq\frac{1-\alpha}{1-\beta},$$

$$\inf_\mu Pr_\mu\left\{\mu(\hat{\theta})\in CS_U^H\right\}\geq 1-\alpha,\ \ \sup_\mu Pr_\mu\left\{\mu(\hat{\theta})\in CS_U^H\right\}\leq\frac{1-\alpha}{1-\beta}.$$

Hybrid confidence sets strike a balance between the conditional and projection approaches. The maximal length of hybrid confidence sets is bounded above by the length of $CS_P^\beta$. For small $\beta$, hybrid confidence sets will be close to conditional confidence sets and thus to the conventional confidence set when $\left\{\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma}\right\}$ with high probability. However, for $\beta>0$, hybrid confidence sets do not fully converge to conventional confidence intervals as $Pr_\mu\left\{\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma}\right\}\to 1$.[26] Nevertheless, in our simulations we find the performance of the hybrid and conditional approaches to be quite similar in these well-separated cases.

While hybrid confidence sets combine the conditional and projection approaches, they can yield overall performance more appealing than either. In Section 2 we found that hybrid confidence sets had a shorter median length for many parameter values than did either the conditional or projection approaches used in isolation. Our simulation results in Sections 6 and 7 below provide further evidence of outperformance in realistic settings.

[26]Indeed, one can directly choose $\beta$ to yield a given maximal power loss for the hybrid tests relative to conditional tests in the well-separated case. Such a choice of $\beta$ will depend on $\Sigma$, however. For simplicity we instead use $\beta=\alpha/10$ in our simulations. Romano et al. (2014) and McCloskey (2017) find this choice to perform well in two different settings when using a Bonferroni correction.

It is worth contrasting our hybrid approach with more conventional Bonferroni corrections as in e.g. Romano et al. (2014); McCloskey (2017). A simple Bonferroni approach for our setting intersects a level $1-\beta$ projection confidence interval $CS_P^\beta$ with a level $1-\alpha+\beta$ conditional interval that conditions only on $\left\{\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma}\right\}$. Bonferroni intervals differ from our hybrid approach in two respects. First, they use a level $1-\alpha+\beta$ conditional confidence interval, while the hybrid approach uses a level $\frac{1-\alpha}{1-\beta}$ conditional interval, where $\frac{1-\alpha}{1-\beta}\leq 1-\alpha+\beta$. Second, the conditional interval used by the Bonferroni approach does not condition on $\mu_Y(\tilde{\theta})\in CS_P^\beta$, while that used by the hybrid approach does. Consequently, hybrid confidence sets never contains the endpoints of $CS_P^\beta$, while the same is not true of Bonferroni intervals.

### 5.3 Hybrid Estimators

The simulation results of Section 2 showed that our median-unbiased estimator can sometimes be much more dispersed than the conventional estimator $\hat{\mu}=Y(\hat{\theta})$. While we do not know of an alternative approach to construct exactly median-unbiased estimators in our setting, a version of our hybrid approach yields estimators that control both median bias and dispersion relative to $\hat{\mu}=Y(\hat{\theta})$.

To construct hybrid estimators we again condition on both $\left\{\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma}\right\}$ and $\mu_Y(\tilde{\theta})\in CS_P^\beta$. Conditional on these events and $Z_{\tilde{\theta}}=z$, we know that $Y(\tilde{\theta})$ again lies in $\mathcal{Y}^H(\tilde{\theta},\tilde{\gamma},\mu_Y(\tilde{\theta}),z)$. Let $F_{TN}^H(y;\mu_Y(\tilde{\theta}),\tilde{\theta},\tilde{\gamma},z)$ denote the conditional distribution function of $Y(\tilde{\theta})$, and define $\hat{\mu}_\alpha^H$ to solve $F_{TN}^H(Y(\hat{\theta});\hat{\mu}_\alpha^H,\hat{\theta},\hat{\gamma},Z_{\tilde{\theta}})=1-\alpha$.

**Proposition 8**
*For $\alpha\in(0,1)$, $\hat{\mu}_\alpha^H$ is unique and $\hat{\mu}_\alpha^H\in CS_P^\beta$. If $\hat{\theta}$ is unique almost surely for all $\mu$, $\hat{\mu}_\alpha^H$ is $\alpha$-quantile-unbiased conditional on $\mu_Y(\hat{\theta})\in CS_P^\beta$:*

$$Pr_\mu\left\{\hat{\mu}_\alpha^H\geq\mu_Y(\hat{\theta})|\mu_Y(\hat{\theta})\in CS_P^\beta\right\}=\alpha \text{ for all } \mu.$$

Proposition 8 implies several notable properties for the hybrid estimator. First, since $Pr_\mu\left\{\mu_Y(\hat{\theta})\in CS_P^\beta\right\}\geq 1-\beta$ by construction, one can show that

$$\left|Pr_\mu\left\{\hat{\mu}_\alpha^H\geq\mu_Y(\hat{\theta})\right\}-\alpha\right|\leq\beta\cdot\max\{\alpha,1-\alpha\} \text{ for all } \mu.$$

This implies that the absolute median bias of $\hat{\mu}_{\frac{1}{2}}^H$ (measured as the deviation of the exceedance probability from 1/2) is bounded above by $\beta/2$. On the other hand, since $\hat{\mu}_{\frac{1}{2}}^H\in CS_P^\beta$ we have $\left|\hat{\mu}_{\frac{1}{2}}^H-Y(\hat{\theta})\right|\leq c_\beta\sqrt{\Sigma_Y(\tilde{\theta})}$, so the difference between $\hat{\mu}_{\frac{1}{2}}^H$ and the conventional esti-

mator $Y(\hat{\theta})$ is bounded above by half the width of $CS_P^\beta$. As $\beta$ varies, the hybrid estimator interpolates between the median-unbiased estimator $\hat{\mu}_{\frac{1}{2}}$ and the conventional estimator $Y(\hat{\theta})$.

# 6 Simulations: Empirical Welfare Maximization

Our first set of simulations considers the EWM setting introduced in Section 3. We calibrate our simulations to experimental data from the National Job Training Partnership Act (JTPA) Study, which was previously used by Kitagawa and Tetenov (2018b) to study empirical welfare maximization. For a detailed description of the study see Bloom et al. (1997).

We have data on $n=11{,}204$ individuals $i$ and the treatment $D_i$ is binary; $D_i=1$ indicates assignment to a job training program and $D_i=0$ indicates non-assignment. The probability of assignment is constant: $d(c)=\Pr(D_i=1|C_i=c)=2/3$. We consider rules that allocate treatment based on years of education $C_i$. In the data, $C$ takes integer values ranging from 6 to 18 years. As in Section 3, rule $\theta$ assigns $i$ to treatment if and only if $C_i \in \mathcal{C}_\theta$.

We consider two classes of policies. The first, which we call threshold policies, treat all individuals with fewer than $\theta$ years of education: $\mathcal{C}_\theta=\{C:C\leq\theta\}$. The second, which we call interval policies, treat all individuals with between $\theta_l$ and $\theta_u$ years of education: $\mathcal{C}_\theta=\{C:\theta_l\leq C\leq\theta_u\}$, where a policy $\theta$ consists of a $(\theta_l,\theta_u)$ pair. The total number of policies $|\Theta|$ is equal to 13 and 91 for the threshold and interval cases, respectively. We define $X_n(\theta)$ as a scaled estimate for the increase in income from policy $\theta$ relative to the baseline of no treatment. For $Y_i$ individual income measured in hundreds of thousands of dollars,

$$X_n(\theta)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{Y_iD_i}{d(C_i)}1\{C_i\in\mathcal{C}_\theta\}-\frac{Y_i(1-D_i)}{1-d(C_i)}1\{C_i\notin\mathcal{C}_\theta\}\right),$$

and we consider inference on the average increase in income, so $Y_n=X_n$.

For our simulations, we focus on the asymptotic problem and draw normal vectors $X$ with known variance $\Sigma_X$ equal to a (consistent) estimate for the asymptotic variance of $X_n$ based on the JTPA data and take $\hat{\theta}=\text{argmax}_\theta X(\theta)$. The object of interest is thus $\mu_X(\hat{\theta})$. The mean vector $\mu_{X,n}$ of $X_n$ is not consistently estimable due to the $\sqrt{n}$ scaling, so we consider three specifications for the mean $\mu_X$ of X. Specification (i) sets $\mu_X=0$, so all policies yield the same welfare as the baseline of no treatment. Specification (ii) sets $\mu_X=(0,-10^5,...,-10^5)$, so one policy is vastly more effective than the others. Finally, specification (iii) sets $\mu_X=X_n$ for $X_n$ calculated in the JTPA data. Intuitively, we expect that specification (i) will be unfavorable to conditional confidence sets since in Section 2 these performed poorly when all policies were equally effective. Specification (ii) should

be favorable to conditional confidence sets since in this case $\hat{\theta}$ selects one policy with high probability, and the results of Section 4.4 apply. Finally, specification (iii) is calibrated to the data and it is not obvious which approaches will perform well in this setting.

To the best of our knowledge our conditional confidence sets are the only known procedures available with correct conditional coverage given $\hat{\theta}$. Hence, we focus on unconditional performance and compare the conditional confidence sets $CS_{ET}$ and $CS_U$ and the hybrid confidence sets $CS^H_{ET}$ and $CS^H_U$ to the projection confidence set $CS_P$. The conditional and hybrid confidence sets are novel to this paper, but (unstudentized) projection confidence sets were previously considered for this problem by Kitagawa and Tetenov (2018a). We take $\alpha = 0.05$ in all cases and so consider 95% confidence sets. For hybrid confidence sets we set $\beta = \alpha/10 = .005$. All reported results are based on $10^4$ simulation draws.

Table 1 reports the unconditional coverage $Pr_\mu\{\mu_X(\hat{\theta}) \in CS\}$ of all five confidence sets, along with the conventional confidence set $CS_N$ as in (24). As expected, all confidence sets other than $CS_N$ have correct coverage in all settings considered. The conditional confidence sets are exact, with coverage equal to 95% up to simulation error. By contrast, hybrid confidence sets tend to be slightly conservative, and projection confidence sets are often quite conservative, with coverage close to one when we consider interval policies.

**Table 1:** Unconditional Coverage Probability

| DGP | $CS_{ET}$ | $CS_U$ | $CS^H_{ET}$ | $CS^H_U$ | $CS_P$ | $CS_N$ |
|---|---|---|---|---|---|---|
| Class of Threshold Policies | | | | | | |
| (i) | 0.949 | 0.950 | 0.952 | 0.953 | 0.986 | 0.922 |
| (ii) | 0.952 | 0.952 | 0.956 | 0.956 | 0.991 | 0.952 |
| (iii) | 0.95 | 0.95 | 0.955 | 0.955 | 0.992 | 0.952 |
| Class of Interval Policies | | | | | | |
| (i) | 0.952 | 0.949 | 0.956 | 0.953 | 0.992 | 0.837 |
| (ii) | 0.95 | 0.951 | 0.954 | 0.954 | 0.998 | 0.950 |
| (iii) | 0.951 | 0.95 | 0.954 | 0.955 | 0.998 | 0.948 |

We next compare the length of confidence sets. Projection confidence sets were proposed in the previous literature and their length is proportional to the standard error $\sqrt{\Sigma_X(\hat{\theta})}$ for the welfare of the estimated optimal policy. Hence, $CS_P$ provides a natural benchmark against which to compare the length of our new confidence sets. In Table 2 we compare our new confidence sets to this benchmark in two ways, first reporting the median lengths of $CS_{ET}$, $CS_U$, $CS^H_{ET}$, and $CS^H_U$ relative to $CS_P$ (that is, the ratio of the median of their lengths), and then reporting the fraction of simulation draws for which

our new confidence sets are longer than $CS_P$.

Focusing first on specification (i) for which $\mu_X = 0$, we see that conditional confidence sets are longer than $CS_P$ according to both measures in the threshold and interval policy specifications. Hence, as expected, this case is unfavorable to these confidence sets. By contrast, our hybrid confidence sets are shorter than the projection sets both in median length and in the substantial majority of simulation draws. Turning next to specification (ii) for which $\mu_X$ has a well-separated maximum, we see that, as expected, conditional confidence sets are much shorter than projection confidence sets. Hybrid confidence sets perform nearly as well. Finally in specification (iii) for which $\mu_X$ is calibrated to the data, we see that the performance of the conditional sets is between its performance in cases (i) and (ii), and that hybrid confidence sets again perform best.

Overall, these simulation results favor the hybrid confidence sets relative to both the conditional and projection sets. The benefits of hybrid confidence sets are still more pronounced if we consider higher quantiles of the length distribution, reported in Section F of the supplement. We do not find a strong advantage for either $CS^H_{ET}$ or $CS^H_U$, though when the two differ $CS^H_{ET}$ typically performs better. Since $CS^H_{ET}$ is also typically easier to calculate, these simulation results suggest using $CS^H_{ET}$ in this setting.

**Table 2:** Length of Confidence Sets Relative to $CS_P$ in EWM Simulations

| DGP | Median Length Relative to $CS_P$ | | | | Probability Longer than $CS_P$ | | | |
|---|---|---|---|---|---|---|---|---|
| | $CS_{ET}$ | $CS_U$ | $CS^H_{ET}$ | $CS^H_U$ | $CS_{ET}$ | $CS_U$ | $CS^H_{ET}$ | $CS^H_U$ |
| Class of Threshold Policies | | | | | | | | |
| (i) | 1.17 | 1.27 | 0.63 | 0.64 | 0.71 | 0.80 | 0.04 | 0.35 |
| (ii) | 0.75 | 0.75 | 0.76 | 0.76 | 0 | 0 | 0 | 0 |
| (iii) | 0.84 | 0.93 | 0.84 | 0.89 | 0.33 | 0.43 | 0 | 0.19 |
| Class of Interval Policies | | | | | | | | |
| (i) | 1.54 | 1.65 | 0.77 | 0.76 | 0.79 | 0.88 | 0 | 0 |
| (ii) | 0.63 | 0.64 | 0.65 | 0.65 | 0 | 0 | 0 | 0 |
| (iii) | 0.78 | 0.88 | 0.76 | 0.81 | 0.32 | 0.42 | 0 | 0 |

We next consider the properties of our point estimators. The initial columns of Table 3 report the simulated median bias of our median unbiased estimator $\hat{\mu}_{\frac{1}{2}}$, our hybrid estimator $\hat{\mu}^H_{\frac{1}{2}}$, and the conventional estimator $X(\hat{\theta})$, measured both as the difference in the exceedance probability from $\frac{1}{2}$ and as the median studentized estimation error. The hybrid estimator is quite close to being median unbiased. By constrast, the conventional estimator exhibits substantial bias when $\mu_X$ does not have a well-separated maximum.

The final three columns of Table 3 report the median absolute studentized error for the estimators considered. These results show that the median unbiased estimator $\hat{\mu}_{\frac{1}{2}}$ has a larger median absolute error than the conventional estimator $X(\hat{\theta})$ in all designs except the well-separated case (ii), where all three estimators perform similarly. The hybrid estimator $\hat{\mu}^H_{\frac{1}{2}}$ likewise has a larger median absolute error than the conventional estimator. Additional results reported in Section F of the supplement show that the hybrid estimator substantially outperforms the median unbiased estimator when one considers higher quantiles of absolute error.

**Table 3:** Bias and Median Absolute Error of Point Estimators

| DGP | $Pr_\mu\left\{\hat{\mu} > \mu_X(\hat{\theta})\right\} - \frac{1}{2}$ | | | $Med_\mu\left(\frac{\hat{\mu}-\mu_X(\hat{\theta})}{\sqrt{\Sigma_X(\hat{\theta})}}\right)$ | | | $Med_\mu\left(\frac{\lvert\hat{\mu}-\mu_X(\hat{\theta})\rvert}{\sqrt{\Sigma_X(\hat{\theta})}}\right)$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\hat{\mu}_{\frac{1}{2}}$ | $\hat{\mu}^H_{\frac{1}{2}}$ | $X(\hat{\theta})$ | $\hat{\mu}_{\frac{1}{2}}$ | $\hat{\mu}^H_{\frac{1}{2}}$ | $X(\hat{\theta})$ | $\hat{\mu}_{\frac{1}{2}}$ | $\hat{\mu}^H_{\frac{1}{2}}$ | $X\left(\hat{\theta}\right)$ |
| Class of Threshold Policies | | | | | | | | | |
| (i) | -0.007 | -0.007 | 0.391 | -0.02 | -0.02 | 0.82 | 1.11 | 1.10 | 0.88 |
| (ii) | -0.001 | 0.001 | 0.001 | 0 | 0 | 0 | 0.67 | 0.67 | 0.67 |
| (iii) | -0.001 | -0.001 | 0.104 | 0 | 0 | 0.25 | 0.80 | 0.79 | 0.67 |
| Class of Interval Policies | | | | | | | | | |
| (i) | 0 | 0.003 | 0.5 | 0 | 0.02 | 1.3 | 1.42 | 1.39 | 1.30 |
| (ii) | -0.002 | 0.001 | 0.001 | 0 | 0 | 0 | 0.65 | 0.65 | 0.66 |
| (iii) | 0 | 0.001 | 0.148 | 0 | 0 | 0.35 | 0.86 | 0.86 | 0.69 |

The results of this section confirm our theoretical findings. Conditional confidence sets and estimators perform well when the optimal policy is well-separated but can otherwise underperform existing alternatives. Hybrid confidence sets outperform existing alternatives in all cases, nearly matching conditional confidence sets in well-separated cases while maintaining much better performance in other settings. Finally, hybrid estimators eliminate almost all median bias while obtaining a substantially smaller median absolute error than the exact median-unbiased estimator. Hence, we find strong evidence favoring our hybrid confidence sets relative to the available alternatives and evidence favoring our hybrid estimators if bias reduction is desired.

## 7 Simulations: Tipping Point Estimation

Our second set of simulation results is based on the tipping point model of Card et al. (2008), a leading application of the threshold regression model discussed throughout this paper as a running example. Card et al. (2008) study the evolution of neighborhood

composition as a function of minority population share. In particular, for $Y_i$ the normalized change in the white population of census tract $i$ between 1980 and 1990, $C_i$ a vector of controls, and $Q_i$ the minority share in 1980, Card et al. (2008) consider the specification

$$Y_i = \beta + C_i'\alpha + \delta 1\{Q_i > \theta\} + U_i,$$

which allows the white population share to change discontinuously when the minority share exceeds some threshold $\theta$. They then fit this model, including the break point $\theta$, by least squares. See Card et al. (2008) for details on the data and motivation. We consider data from Chicago and Los Angeles with $n = 1{,}820$ and $n = 2{,}035$ observations, respectively, estimating the model separately in each city.[27]

Results in Wang (2018) show that if we model the coefficient $\delta$ as on the same order as sampling uncertainty, this threshold regression model satisfies the high-level conditions (12)–(13) we introduced in Section 3. Hence, we can immediately apply our results for the norm-maximization problem to the present setting. Specifically, we define $X_n$ as discussed in Section 3 and $\hat{\theta}_n$ is again asymptotically equivalent to the solution to a norm-maximization problem $\text{argmax}_{\theta \in \Theta} \|X(\theta)\|$.[28] We define $Y_n(\theta) = \sqrt{n}\hat{\delta}(\theta)$ to be proportional to the estimated change coefficient imposing tipping point $\theta$, so we again consider the problem of inference on the change coefficient while acknowledging randomness in the estimated threshold.

Our simulations draw normal random vectors $(X, Y)$ from the limiting normal model derived in Section 3. This model depends on the function $\Sigma_C$ and the covariance function of $G$ in Section 3 which we (consistently) estimate from the Card et al. (2008) data. It also depends on the function $\Sigma_{cg}(\cdot)$. Since this is not consistently estimable, we consider three specifications. Specification (i) assumes there is no coefficient change, corresponding to $\delta = 0$. Specification (ii) assumes that there is a single large change, setting $\delta = -100\%$ and taking the true threshold to equal the estimate in the Card et al. (2008) data. Finally, specification (iii) calibrates $\Sigma_{cg}(\cdot)$ to the data, corresponding to the analog of model (10) where the intercept term in the regression may depend arbitrarily upon a neighborhood's minority share. This specification implies that the break model is misspecified but as discussed above, our approach remains applicable in this case, unlike the results of Wang

[27] We focus on these cities following Wang (2017), a previous version of Wang (2018), since Card et al. (2008) note that their tipping point estimation method appears more appropriate for larger cities.

[28] While Card et al. (2008) optimize over all possible tipping points between 5% and 60%, consistent with our theoretical results we limit attention to a finite set of thresholds. In particular, we consider 100 evenly-spaced quantiles of the minority share, and then further restrict attention to thresholds between 5% and 60%. We also tried several other discretization schemes and found very similar results in all cases.

(2018). Indeed, Card et al. (2008) acknowledge that the tipping point model only approximates their underlying theoretical model of neighborhood ethnic composition, so misspecification seems likely in this setting.

We again focus on the unconditional performance of our proposed procedures along with existing alternatives. All reported results are based on $10^4$ simulation draws. Table 4 reports coverage for the confidence sets $CS_{ET}$, $CS_U$, $CS^H_{ET}$, $CS^H_U$, and $CS_P$, along with the conventional confidence set $CS_N$. As for the simulations calibrated to the EWM application, we see that all confidence sets other than $CS_N$ have correct coverage, $CS_P$ often over-covers, the conditional confidence sets have exact coverage and the hybrid confidence sets exhibit minimal over-coverage. In this application, the conventional confidence set $CS_N$ severely under-covers for some simulation designs.

**Table 4:** Unconditional Coverage Probability

| DGP | $CS_{ET}$ | $CS_U$ | $CS^H_{ET}$ | $CS^H_U$ | $CS_P$ | $CS_N$ |
|---|---|---|---|---|---|---|
| Chicago Data Calibration | | | | | | |
| (i) | 0.948 | 0.95 | 0.949 | 0.949 | 0.95 | 0.750 |
| (ii) | 0.951 | 0.95 | 0.956 | 0.955 | 0.994 | 0.951 |
| (iii) | 0.947 | 0.946 | 0.951 | 0.951 | 0.990 | 0.934 |
| Los Angeles Data Calibration | | | | | | |
| (i) | 0.949 | 0.948 | 0.949 | 0.948 | 0.95 | 0.615 |
| (ii) | 0.952 | 0.952 | 0.956 | 0.956 | 0.996 | 0.952 |
| (iii) | 0.951 | 0.949 | 0.955 | 0.954 | 0.996 | 0.95 |

Table 5 compares the lengths of our confidence sets to that of $CS_P$. For each confidence set we again report both median length relative to $CS_P$ and the frequency with which the confidence set is longer than $CS_P$. Here we see that the conditional confidence sets can be relatively long, while the hybrid confidence sets provide marked performance improvements across the specifications considered. Similarly to the simulation exercises of the previous section, the benefits of the hybrid confidence sets can become even more pronounced at different length quantiles. See Section G of the supplemental appendix. Remarkably, neither of the hybrid confidence sets is longer than $CS_P$ in any simulation draw across all specifications examined. The overall message is similar to that of the previous section: hybrid confidence sets possess clear advantages for unconditional inference and $CS^H_{ET}$ seems to be the most compelling option, especially given its computational simplicity.

Finally, we consider the properties of our point estimators. The initial columns of Table 6 report median bias measured both with the deviation of the exceedance proba-

**Table 5:** Length of Confidence Sets Relative to $CS_P$ in Tipping Point Simulations

| | Median Length Relative to $CS_P$ | | | | Probability Longer than $CS_P$ | | | |
|---|---|---|---|---|---|---|---|---|
| | $CS_{ET}$ | $CS_U$ | $CS^H_{ET}$ | $CS^H_U$ | $CS_{ET}$ | $CS_U$ | $CS^H_{ET}$ | $CS^H_U$ |
| Chicago Data Calibration | | | | | | | | |
| (i) | 1.33 | 1.38 | 0.94 | 0.94 | 0.83 | 0.89 | 0 | 0 |
| (ii) | 0.72 | 0.72 | 0.74 | 0.74 | 0 | 0 | 0 | 0 |
| (iii) | 0.82 | 0.93 | 0.82 | 0.87 | 0.35 | 0.44 | 0 | 0 |
| Los Angeles Data Calibration | | | | | | | | |
| (i) | 1.26 | 1.29 | 0.86 | 0.85 | 0.58 | 0.62 | 0 | 0 |
| (ii) | 0.68 | 0.68 | 0.69 | 0.69 | 0 | 0 | 0 | 0 |
| (iii) | 0.68 | 0.70 | 0.70 | 0.72 | 0.15 | 0.19 | 0 | 0 |

bility from $\frac{1}{2}$ and with the studentized median estimation error. We again see that $\hat{\mu}_{\frac{1}{2}}$ is median-unbiased (up to simulation error) and that $\hat{\mu}^H_{\frac{1}{2}}$ exhibits minimal median bias. By contrast, in specification (i) the conventional estimator $Y(\hat{\theta})$ has substantial median bias as measured by the studentized median estimation error, though very little as measured by the exceedance probability. This latter feature reflects the fact that the density of $Y(\hat{\theta})-\mu_Y(\hat{\theta})$ has very little mass near zero in this specification.

Turning to median absolute studentized error, we see that all estimators perform similarly when the series has a single large break. By contrast, the median unbiased estimator $\hat{\mu}_{\frac{1}{2}}$ performs better than the conventional estimator $Y(\hat{\theta})$ in specification (i) (no break) but performs worse in specification (iii). The hybrid estimator is weakly better than the unbiased estimator in all cases, with perfomance gains in case (i) and equal performance in the other two cases. Again, the performance gains are more pronounced if one considers higher quantiles of the absolute error distribution, as reported in Section G of the supplement.

### 7.1 Split-Sample Procedures

While we have so far compared the performance of our conditional and hybrid procedures to the projection confidence set $CS_P$ and conventional estimator $Y(\hat{\theta})$, Card et al. (2008) instead adopt a sample-splitting approach, using two-thirds of the data to select the break-date and a third of the data for inference. In this section we compare the performance of this conventional split-sample procedure to that of the implementable split-sample alternative developed in Section C of the supplement. We consider the same calibrations to the Card et al. (2008) data as above and choose the sample split as in Card et al. (2008).

Table 7 compares the conventional split-sample confidence set $CS_{SS}$ and estimator $Y^2(\hat{\theta}^1)$ used by Card et al. (2008) to our (equal-tailed) alternative split-sample confidence

**Table 6:** Bias and Median Absolute Error in Tipping Point Simulations

| | $Pr_\mu\left\{\hat{\mu}>\mu_Y(\hat{\theta})\right\}-\frac{1}{2}$ | | | $Med_\mu\left(\frac{\hat{\mu}-\mu_Y(\hat{\theta})}{\sqrt{\Sigma_Y(\hat{\theta})}}\right)$ | | | $Med_\mu\left(\left\lvert\frac{\hat{\mu}-\mu_Y(\hat{\theta})}{\sqrt{\Sigma_Y(\hat{\theta})}}\right\rvert\right)$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\hat{\mu}_{\frac{1}{2}}$ | $\hat{\mu}^H_{\frac{1}{2}}$ | $Y(\hat{\theta})$ | $\hat{\mu}_{\frac{1}{2}}$ | $\hat{\mu}^H_{\frac{1}{2}}$ | $Y(\hat{\theta})$ | $\hat{\mu}_{\frac{1}{2}}$ | $\hat{\mu}^H_{\frac{1}{2}}$ | $Y(\hat{\theta})$ |
| Chicago Data Calibration | | | | | | | | | |
| (i) | 0 | 0 | 0.01 | -0.01 | 0.01 | 0.64 | 1.51 | 1.38 | 1.52 |
| (ii) | -0.01 | -0.01 | -0.01 | -0.03 | -0.03 | -0.03 | 0.66 | 0.66 | 0.66 |
| (iii) | -0.01 | -0.01 | -0.15 | -0.03 | -0.03 | -0.37 | 0.83 | 0.83 | 0.71 |
| Los Angeles Data Calibration | | | | | | | | | |
| (i) | 0 | 0 | 0 | 0 | 0 | -0.8 | 1.38 | 1.29 | 1.80 |
| (ii) | 0 | 0 | 0 | 0.01 | 0.01 | 0.01 | 0.67 | 0.67 | 0.67 |
| (iii) | 0 | 0 | 0.006 | 0 | -0.01 | -.016 | 0.74 | 0.74 | 0.68 |

set $C^A_{SS}$ and median-unbiased estimator $\hat{\mu}^A_{\frac{1}{2},SS}$. See Section C of the supplement for definitions. These results clearly reflect the dominance of our alternative split-sample procedures, with substantial performance improvements for both confidence sets and estimators across all calibrations. These improvements are largest in the well-separated case (ii), but are nearly as large in the data-calibrated case (iii). Section G of the supplement provides ratios of the $5^{th}$, $25^{th}$, $50^{th}$, $75^{th}$ and $95^{th}$ quantiles of the lengths of $CS^A_{SS}$ relative to the those of $CS_{SS}$ as well as the quantiles of $\left|\hat{\mu}-\mu_Y(\hat{\theta}^1)\right|/\sqrt{\Sigma_Y(\hat{\theta}^1)}$ for $\hat{\mu}=\hat{\mu}^A_{\frac{1}{2},SS}$ and $\hat{\mu}=Y^2(\hat{\theta}^1)$. There, our new split-sample procedures can be seen to dominate the conventional ones across all quantiles and simulation designs considered, often by very wide margins.

**Table 7:** Performance Measures of Split-Sample Procedures

| | Median Length Relative to $CS_{SS}$ | $Med_\mu\left(\frac{\lvert\hat{\mu}-\mu_Y(\hat{\theta}^1)\rvert}{\sqrt{\Sigma_Y(\hat{\theta}^1)}}\right)$ | |
|---|---|---|---|
| DGP | $CS^A_{SS}$ | $\hat{\mu}^A_{\frac{1}{2},SS}$ | $Y^2(\hat{\theta}^1)$ |
| Chicago Data Calibration | | | |
| (i) | 0.83 | 0.57 | 0.67 |
| (ii) | 0.58 | 0.38 | 0.66 |
| (iii) | 0.64 | 0.44 | 0.67 |
| Los Angeles Data Calibration | | | |
| (i) | 0.78 | 0.55 | 0.69 |
| (ii) | 0.58 | 0.39 | 0.67 |
| (iii) | 0.59 | 0.42 | 0.68 |

## 8 Conclusion

This paper considers a form of the winner's curse that arises when we select a target parameter for inference based on optimization. We propose confidence sets and quantile unbiased estimators for the target parameter that are optimal conditional on its selection. We hence recommend our conditional inference procedures when it is appropriate to remove uncertainty about the choice of target parameters from inferential statements. These conditionally valid procedures are also unconditionally valid, but we find that they sometimes have unappealing (unconditional) performance relative to existing alternatives. If one is satisfied with correct unconditional coverage and (in the case of estimation) a small, controlled degree of bias, we propose hybrid inference and estimation procedures which combine conditioning with projection confidence sets. Examining performance in simulations calibrated to empirical welfare maximization and tipping point applications, we find that our hybrid approach performs well in both cases.

Our results suggest a range of opportunities for future work. First, rather than considering inference on $\mu_Y(\hat{\theta})$, under suitable assumptions one could build on our results to forecast $Y(\hat{\theta})$. Alternatively, while conditional and projection confidence sets have antecedents in the literature on inference after model selection, including in Berk et al. (2013) and Fithian et al. (2017), there is no analog of our hybrid approach in this literature. Our very positive simulation results for the hybrid approach in the present setting suggest that this approach might yield appealing performance in a range of post-selection-inference settings. Even if a fully conditional approach is desired in the post-selection problem, as in Fithian et al. (2017), one could consider the analog of our optimal median-unbiased estimates that condition on the selected model. Finally, the problem of estimating the value of a dynamic treatment rule (c.f. Chakraborty and Murphy, 2014; Han, 2018) is closely related to our level-maxmization setting, so it seems likely that our results could prove to be useful there as well.

## References


Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):821–856.

Athey, S. and Wager, S. (2018). Efficient policy learning. *arXiv preprint*. arXiv 1702.0289.

Bachoc, F., Leeb, H., and Potscher, B. (2018). Valid confidence intervals for post-model-selection predictors. *Annals of Statistics (forthcoming)*.

Bachoc, F., Preinerstorfer, D., and Steinberger, L. (2017). Uniformly valid confidence intervals post-model-selection. Unpublished Manuscript.

Baird, S., Hicks, J. H., Kremer, M., and Miguel, E. (2016). Worms at work: Long-run impacts of a child health investment. *The Quarterly Journal of Economics*, 131(4):1637–1680.

Banerjee, A., Hanna, R., Kyle, J., Olken, B. A., and Sumarto, S. (2018). Tangible information and citizen empowerment: Identification cards and food subsidy programs in Indonesia. *Journal of Political Economy*, 126(2):451–491.

Berk, R., Brown, L., Buja, A., Zhang, K., and Zhao, L. (2013). Valid post-selection inference. *Annals of Statistics*, 41(2):802–831.

Bhattacharya, D. (2009). Inferring optimal peer assignment from experimental data. *Journal of the American Statistical Association*, 104:486–500.

Björkman Nyqvist, M. and Jayachandran, S. (2017). Mothers care more, but fathers decide: Educating parents about child health in Uganda. *American Economic Review*, 107(5):496–500.

Bloom, H. S., Orr, L. L., Bell, S. H., Cave, G., Doolittle, F., Lin, W., and Bos, J. M. (1997). The benefits and costs of jtpa title ii-a programs: Key findings from the national job training partnership act study. *Journal of Human Resources*, 32(3):549–576.

Card, D., Mas, A., and Rothstein, J. (2008). Tipping and the dynamics of segregation. *Quarterly Journal of Economics*, 123:177–216.

Chakraborty, B. and Murphy, S. (2014). Dynamic treatment regimes. *Annual Review of Statistics and Its Applications*, 1:447–64.

Chernozhukov, V., Lee, S., and Rosen, A. (2013). Intersection bounds: Estimation and inference. *Econometrica*, 81(2):667–737.

Cox, G. (2018). Almost sure uniqueness of a global minimum without convexity. Unpublished Manuscript.

Dawid, A. P. (1994). Selection paradoxes in bayesian inference. *Institute of Mathematical Statistics Lecture Notes - Monograph Series*, 24:211–220.

Duflo, E., Greenstone, M., Pande, R., and Ryan, N. (2018). The value of regulatory discretion: Estimates from environmental inspections in India. *Econometrica*, 86(6):2123–2160.

Efron, B. (2011). Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614.

Elliott, G. and Müller, U. K. (2007). Confidence sets for the date of a single break in linear time series regressions. *Journal of Econometrics*, 141(2):1196–1218.

Elliott, G. and Müller, U. K. (2014). Pre and post break parameter inference. *Journal of Econometrics*, 180:141–157.

Ferguson, J. P., Cho, J. H., Yang, C., and Zhao, H. (2013). Empirical bayes correction for the winner's curse in genetic association studies. *Genetic Epidemiology*, 37(1):60–68.

Fithian, W., Sun, D., and Taylor, J. (2017). Optimal inference after model selection. *arXiv*.

Graham, B. S., Imbens, G. W., and Ridder, G. (2014). Complementarity and aggregate implications of assortative matching: A nonparametric analysis. *Quantitative Economics*, 5:29–66.

Han, S. (2018). Identification in nonparametric models for dynamic treatment effects. Unpublished Manuscript.

Hansen, B. E. (1997). Inference in TAR models. *Studies in Nonlinear Dynamics and Econometrics*, 2:1–14.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica*, 68:575–603.

Hansen, B. E. (2001). The new econometrics of structural change: Dating breaks in u.s. labor productivity. *Journal of Economic Perspectives*, 15(4):117–128.

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business and Economic Statistics*, 23(4):365–380.

Harris, X. T., Panigrahi, S., Markovic, J., Bi, N., and Taylor, J. (2016). Selective sampling after solving a convex problem. *arXiv*.

Hirano, K. and Porter, J. R. (2009). Asymptotics for statistical treatment rules. *Econometrica*, 77:1683–1701.

Hyun, S., Lin, K., G'Sell, M., and Tibshirani, R. J. (2018). Post-selection inference for changepoint detection algorithms with application to copy number variation data. Unpublished Manuscript.

Khan, A. Q., Khwaja, A. I., and Olken, B. A. (2016). Tax farming redux: Experimental evidence on performance pay for tax collectors. *The Quarterly Journal of Economics*, 131(1):219–271.

Kitagawa, T. and Tetenov, A. (2018a). Supplement to "who should be treated? empirical welfare maximization methods for treatment choice". *Econometrica*.

Kitagawa, T. and Tetenov, A. (2018b). Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica*, 86(2):591–616.

Kivaranovic, D. and Leeb, H. (2018). Expected length of post-model-selection confidence intervals conditional on polyhedral constraints. Unpublished Manuscript.

Kuchibhotla, A., Brown, L., Buja, A., George, E., and Zhao, L. (2018). A model free perspective for linear regression: Uniform-in-model bounds on post selection inference. Unpublished Manuscript.

Laber, E. and Murphy, S. (2011). Adaptive confidence intervals for the test error in classification. *Journal of the American Statistical Association*, 106(495):904–913.

Lee, J. D., Sun, D. L., Sun, Y., and Taylor, J. E. (2016). Exact post-selection inference, with application to the LASSO. *Annals of Statistics*, 44:907–927.

Lee, M. R. and Shen, M. (2018). Winner's curse: Bias estimation for total effects of features in online controlled experiments. In *KDD*.

Lehmann, E. and Scheffé, H. (1955). Completeness, similar regions, and unbiased estimation: Part ii. *Sankhyā: The Indian Journal of Statistics*, 15(3):219–236.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. Springer, third ed. edition.

Manski, C. F. (2004). Statistical treatment rules for heterogeneous populations. *Econometrica*, 72(4):1221–1246.

McCloskey, A. (2017). Bonferroni-based size-correction for nonstandard testing problems. *Journal of Econometrics*, 200(1):17–35.

Perron, P. (2006). Dealing with structural breaks. In *Palgrave Handbook of Econometrics*, volume 1: Econometric Theory, pages 278–352. Palgrave.

Pfanzagl, J. (1979). On optimal median unbiased estimators in the presence of nuisance parameters. *Annals of Statistics*, 7(1):187–193.

Pfanzagl, J. (1994). *Parametric Statistical Theory*. De Gruyter.

Rai, Y. (2018). Statistical inference for treatment assignment policies. Unpublished Manuscript.

Romano, J. P., Shaikh, A., and Wolf, M. (2014). A practical two-step method for testing moment inequalities. *Econometrica*, 82(5):1979–2002.

Romano, J. P. and Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4):1237–1282.

Tetenov, A. (2012). Statistical treatment choice based on asymmetric minimax regret criteria. *Journal of Econometrics*, 166(1):157–165.

Tian, X., Bi, N., and Taylor, T. (2016). MAGIC: a general, powerful and tractable method for selective inference. *arXiv*.

Tian, X. and Taylor, J. (2016). Selective inference with a randomized response. *Annals of Statistics (forthcoming)*.

Tibshirani, R. J., Rinaldo, A., Tibshirani, R., and Wasserman, L. (2018). Uniform asymptotic inference and the bootstrap after model selection. *The Annals of Statistics*, 46(3):1255–1287.

Wang, Y. (2017). Inference in the threshold model. Unpublished Manuscript.

Wang, Y. (2018). Inference in the change point model with time-varying moments. Unpublished Manuscript.

White, H. (2000). A reality check for data snooping. *Econometrica*, 68(5):1097–1126.

Zhong, H. and Prentice, R. (2009). Correcting "winner's curse" in odds ratios from genomewide association findings for major complex human diseases. *Genetic Epidemiology*, 34(1):78–91.

Supplement to the paper

# Inference on Winners

Isaiah Andrews Toru Kitagawa Adam McCloskey

December 31, 2018

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 and threshold regression examples 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. Section F reports additional simulations results for the EWM simulations discussed in Section 6 of the paper. Finally, Section G reports additional simulation results for the threshold regression simulations discussed in Section 7 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

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

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 $(\mu_X,\mu_Y)$, 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 $\Lambda$ 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

$$\frac{1\left\{\left(\tilde{z} + \Sigma_{XY}(\cdot,\tilde{\theta})\Sigma_Y(\tilde{\theta})^{-1}\tilde{y}\right) \in \mathcal{X}(\tilde{\theta},\tilde{\gamma})\right\}}{Pr_{\tilde{\mu}_Z,\tilde{\mu}_Y}\left\{X \in \mathcal{X}(\tilde{\theta},\tilde{\gamma})\right\}} f_{N,\tilde{Z}}(\tilde{z};\tilde{\mu}_Z) f_{N,\tilde{Y}}(\tilde{y};\tilde{\mu}_Y) f_{N,Y^*}(\tilde{y}^*;\mu^*_Y). \tag{26}$$

The density (26) 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 $(\mu_X,\mu_Y)$ 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 $\Theta$ as $\{\theta_1,\theta_2,...,\theta_{|\Theta|}\}$, 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 $\eta$ 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 (17) 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 $j$th 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. □

**Proof of Proposition 4** Note the following equivalence of events:

$$\begin{aligned}
\{\hat\theta=\tilde\theta\} &= \left\{\sum_{i=1}^{d_X} X_i(\tilde\theta)^2 \geq \sum_{i=1}^{d_X} X_i(\theta)^2 \ \forall\theta\in\Theta\right\} \\
&= \left\{\sum_{i=1}^{d_X}\Big[Z_{\tilde\theta,i}(\tilde\theta)+\Sigma_{XY,i}(\tilde\theta)\Sigma_Y(\tilde\theta)^{-1}Y(\tilde\theta)\Big]^2 \right. \\
&\qquad \left. \geq \sum_{i=1}^{d_X}\Big[Z_{\tilde\theta,i}(\theta)+\Sigma_{XY,i}(\theta,\tilde\theta)\Sigma_Y(\tilde\theta)^{-1}Y(\tilde\theta)\Big]^2 \ \forall\theta\in\Theta\right\} \\
&= \Big\{A(\tilde\theta,\theta)Y(\tilde\theta)^2+B_Z(\tilde\theta,\theta)Y(\tilde\theta)+C_Z(\tilde\theta,\theta)\geq 0 \ \forall\theta\in\Theta\Big\}, \qquad (27)
\end{aligned}$$

for $A(\tilde\theta,\theta)$, $B_Z(\tilde\theta,\theta)$, and $C_Z(\tilde\theta,\theta)$ as defined in the statement of the proposition.

By the quadratic formula, (27) is equivalent to the event

$$\left\{\frac{-B_Z(\tilde\theta,\theta)-\sqrt{D_Z(\tilde\theta,\theta)}}{2A(\tilde\theta,\theta)} \leq Y(\tilde\theta) \leq \frac{-B_Z(\tilde\theta,\theta)+\sqrt{D_Z(\tilde\theta,\theta)}}{2A(\tilde\theta,\theta)}\right.$$

$$\forall\theta\in\Theta \text{ s.th. } A(\tilde\theta,\theta)<0 \text{ and } D_Z(\tilde\theta,\theta)\geq 0,$$

$$Y(\tilde\theta) \leq \frac{-B_Z(\tilde\theta,\theta)-\sqrt{D_Z(\tilde\theta,\theta)}}{2A(\tilde\theta,\theta)} \text{ or } Y(\tilde\theta) \geq \frac{-B_Z(\tilde\theta,\theta)+\sqrt{D_Z(\tilde\theta,\theta)}}{2A(\tilde\theta,\theta)}$$

$$
\begin{aligned}
&\forall\theta\in\Theta \text{ s.th. } A(\tilde{\theta},\theta)>0 \text{ and } D_Z(\tilde{\theta},\theta)\geq 0,\\
&Y(\tilde{\theta})\geq\frac{-C_Z(\tilde{\theta},\theta)}{B_Z(\tilde{\theta},\theta)}\ \forall\theta\in\Theta \text{ s.th. } A(\tilde{\theta},\theta)=0 \text{ and } B_Z(\tilde{\theta},\theta)>0,\\
&Y(\tilde{\theta})\leq\frac{-C_Z(\tilde{\theta},\theta)}{B_Z(\tilde{\theta},\theta)}\ \forall\theta\in\Theta \text{ s.th. } A(\tilde{\theta},\theta)=0 \text{ and } B_Z(\tilde{\theta},\theta)<0,\\
&C_Z(\tilde{\theta},\theta)\geq 0\ \forall\theta\in\Theta \text{ s.th. } A(\tilde{\theta},\theta)=B_Z(\tilde{\theta},\theta)=0,\\
&C_Z(\tilde{\theta},\theta)>0\ \forall\theta\in\Theta \text{ s.th. } D_Z(\tilde{\theta},\theta)<0\Big\}\\
=&\Bigg\{Y(\tilde{\theta})\in\bigcap_{\theta\in\Theta:A(\tilde{\theta},\theta)<0,D_Z(\tilde{\theta},\theta)\geq 0}\left[\frac{-B_Z(\tilde{\theta},\theta)-\sqrt{D_Z(\tilde{\theta},\theta)}}{2A(\tilde{\theta},\theta)},\frac{-B_Z(\tilde{\theta},\theta)+\sqrt{D_Z(\tilde{\theta},\theta)}}{2A(\tilde{\theta},\theta)}\right]\\
&\cap\bigcap_{\theta\in\Theta:A(\tilde{\theta},\theta)>0,D_Z(\tilde{\theta},\theta)\geq 0}\left(-\infty,\frac{-B_Z(\tilde{\theta},\theta)-\sqrt{D_Z(\tilde{\theta},\theta)}}{2A(\tilde{\theta},\theta)}\right]\cup\left[\frac{-B_Z(\tilde{\theta},\theta)+\sqrt{D_Z(\tilde{\theta},\theta)}}{2A(\tilde{\theta},\theta)},\infty\right)\\
&\cap\bigcap_{\theta\in\Theta:A(\tilde{\theta},\theta)=0,B_Z(\tilde{\theta},\theta)>0}\Big[H_Z(\tilde{\theta},\theta),\infty\Big)\cap\bigcap_{\theta\in\Theta:A(\tilde{\theta},\theta)=0,B_Z(\tilde{\theta},\theta)<0}\Big(-\infty,H_Z(\tilde{\theta},\theta)\Big]\Bigg\}\\
&\cap\Big\{\min_{\theta\in\Theta:A(\tilde{\theta},\theta)=B_Z(\tilde{\theta},\theta)=0 \text{ or } D_Z(\tilde{\theta},\theta)<0}C_Z(\tilde{\theta},\theta)\geq 0\Big\}\\
=&\Big\{Y(\tilde{\theta})\in\Big[\max_{\theta\in\Theta:A(\tilde{\theta},\theta)<0,D_Z(\tilde{\theta},\theta)\geq 0}G_Z(\tilde{\theta},\theta),\min_{\theta\in\Theta:A(\tilde{\theta},\theta)<0,D_Z(\tilde{\theta},\theta)\geq 0}K_Z(\tilde{\theta},\theta)\Big]\\
&\cap\Big[\max_{\theta\in\Theta:A(\tilde{\theta},\theta)=0,B_Z(\tilde{\theta},\theta)>0}H_Z(\tilde{\theta},\theta),\infty\Big)\cap\Big(-\infty,\min_{\theta\in\Theta:A(\tilde{\theta},\theta)=0,B_Z(\tilde{\theta},\theta)<0}H_Z(\tilde{\theta},\theta)\Big]\\
&\cap\bigcap_{\theta\in\Theta:A(\tilde{\theta},\theta)>0,D_Z(\tilde{\theta},\theta)\geq 0}\Big(-\infty,K_Z(\tilde{\theta},\theta)\Big]\cup\Big[G_Z(\tilde{\theta},\theta),\infty\Big)\Big\}\cap\Big\{\mathcal{V}(\tilde{\theta},Z_{\tilde{\theta}})\geq 0\Big\}\\
=&\Big\{Y(\tilde{\theta})\in\bigcap_{\theta\in\Theta:A(\tilde{\theta},\theta)>0,D_Z(\tilde{\theta},\theta)\geq 0}\Big[\ell_Z^1(\tilde{\theta},\theta),u_Z^1(\tilde{\theta},\theta)\Big]\cup\Big[\ell_Z^2(\tilde{\theta},\theta),u_Z^2(\tilde{\theta},\theta)\Big]\Big\}\cap\Big\{\mathcal{V}(\tilde{\theta},Z_{\tilde{\theta}})\geq 0\Big\}
\end{aligned}
$$

for $D_Z(\tilde{\theta},\theta)$, $G_Z(\tilde{\theta},\theta)$, $H_Z(\tilde{\theta},\theta)$, $K_Z(\tilde{\theta},\theta)$, $\ell_Z^1(\tilde{\theta})$, $\ell_Z^2(\tilde{\theta},\theta)$, $u_Z^1(\tilde{\theta},\theta)$, $u_Z^2(\tilde{\theta})$, and $\mathcal{V}(\tilde{\theta},Z_{\tilde{\theta}})$ again defined in the statement of the proposition. The result follows immediately. □

**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

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

Note, next, that both the conventional and conditional confidence sets are equivariant under shifts, in the sense that the conditional confidence set 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 set 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 $\mu_Y^*$ 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 sets 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 sets in the problem with $\mu_Y$ 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*

$Pr\Big\{Y(\tilde{\theta})\in\mathcal{Y}_m\Big\}\to 1$, *we have that for*

$$\phi_{ET}(\mu_{Y,0})=1\Big\{Y(\tilde{\theta})\notin[c_{l,ET}(\mu_{Y,0},\mathcal{Y}_m),c_{u,ET}(\mu_{Y,0},\mathcal{Y}_m)]\Big\} \tag{28}$$

*and*

$$\phi_{U}(\mu_{Y,0})=1\Big\{Y(\tilde{\theta})\notin[c_{l,U}(\mu_{Y,0},\mathcal{Y}_m),c_{u,U}(\mu_{Y,0},\mathcal{Y}_m)]\Big\}, \tag{29}$$

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

*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 set formed by inverting a family of tests

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

Hence, we want to show that

$$CS_m\to_p\Big[Y(\tilde{\theta})-c_{\frac{\alpha}{2},N},Y(\tilde{\theta})+c_{\frac{\alpha}{2},N}\Big], \tag{30}$$

as $m\to\infty$, for $CS_m$ formed by inverting either (28) or (29).

We assume that $CS_m$ is a finite interval for all $m$, which holds trivially for the equal-tailed confidence set $CS_{ET}$, 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

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

for $\phi_m$ equal to either (28) or (29). This convergence in probability holds jointly for all finite collections of values $\mu_{Y,0}$, however, which implies (30). 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. □

**Proof of Proposition 5** 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 sets. The same argument likewise implies the result for our median unbiased estimator. □

**Proof of Proposition 6** 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** Consider first the level-maximization case. Note that the assumption of the lemma implies that $X(\tilde{\theta})-X(\theta)$ has a non-degenerate normal distribution for all $\mu$. Since $\Theta$ is finite, almost-sure uniqueness of $\hat{\theta}$ follows immediately.

For norm-maximization, assume without loss of generality that $Var\left(X(\theta)|X(\tilde{\theta})\right)\neq 0$. Note that $\|X(\theta)\|$ is continuously distributed conditional on $X(\tilde{\theta})=x(\tilde{\theta})$ for all $x(\tilde{\theta})$ and all

$\mu$, so $Pr_\mu\Big\{\|X(\theta)\|=\|X(\tilde{\theta})\|\Big\}=0$. Almost-sure uniqueness of $\hat{\theta}$ again follows immediately from finiteness of $\Theta$. $\square$

**Proof of Proposition 7** 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 sets,

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

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

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

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

**Proof of Proposition 8** 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 $\Big\{\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma},Z_{\hat{\theta}}=z_{\tilde{\theta}},\mu_Y(\tilde{\theta})\in CS_P^\beta\Big\}$. Since $F_{TN}^H(Y(\hat{\theta});\mu_Y(\tilde{\theta}),\tilde{\theta},\tilde{\gamma},Z_{\tilde{\theta}})$ is decreasing in $\mu_Y(\tilde{\theta})$,

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

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

and thus $\hat{\mu}_\alpha^H$ is $\alpha$-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 $\alpha$-quantile-unbiasedness conditional on $\mu_Y(\tilde{\theta}) \in CS_P^\beta$ follows by the same argument as in the proof of Proposition 6.

**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

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

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 $c_{u,ET}(\mathcal{Y}_m) \to F_N^{-1}\big(1 - \frac{\alpha}{2}\big)$, 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 $\zeta_m$ 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

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

As in the argument for equal-tailed tests above, we see that both $F_N(c_u) - Pr\{\zeta_m \leq c_u\}$

and $F_N(c_l) - Pr\{\zeta_m \leq c_l\}$ 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

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

Hence, since

$$\begin{aligned} &\left| \int_{c_l}^{c_u} \left( \frac{1\{y \in \mathcal{Y}_m\}}{Pr\{\xi \in \mathcal{Y}_m\}} - 1\{y \in \mathcal{Y}_m\} \right) y f_N(y) dy \right| \\ &\leq \left| \left( \frac{1}{Pr\{\xi \in \mathcal{Y}_m\}} - 1 \right) \right| E[|\xi| 1\{\xi \notin \mathcal{Y}_m\}] \leq \left| \left( \frac{1}{Pr\{\xi \in \mathcal{Y}_m\}} - 1 \right) \right| \sqrt{P(\xi \notin \mathcal{Y}_m)} \end{aligned}$$

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. □

**Proof of Lemma 4** Note that the critical values solve

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

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

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

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,

$$\frac{\partial}{\partial c}g(c)=\begin{pmatrix} -f_N(c_l) & f_N(c_u) \\ -c_l f_N(c_l) & c_u f_N(c_u) \end{pmatrix},$$

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

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

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

## B.1 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

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

## B.2 Details for Threshold Regression Estimation Example

This section provides additional results to supplement our discussion of the threshold regression example in the text.

We begin by establishing the weak convergence (14). To do so, we show uniform convergence over any compact set $\tilde{\Theta}$ in the interior of the support of $Q_i$, which implies uniform convergence over $\Theta$. Note, in particular, that under (12) and (13), the continuous mapping theorem implies that

$$X_n(\theta)\Rightarrow X(\theta)$$
$$=\begin{pmatrix}\Sigma_C(\theta)^{-1/2}\Sigma_{Cg}(\theta)\\ (\Sigma_C(\infty)-\Sigma_C(\theta))^{-1/2}(\Sigma_{Cg}(\infty)-\Sigma_{Cg}(\theta))\end{pmatrix}+\begin{pmatrix}\Sigma_C(\theta)^{-1/2}G(\theta)\\ (\Sigma_C(\infty)-\Sigma_C(\theta))^{-1/2}(G(\infty)-G(\theta))\end{pmatrix}$$

uniformly on $\tilde{\Theta}$, where we use the following slight abuse of notation:

$$\frac{1}{n}\sum_{i=1}^n C_iC_i'\to_p\Sigma_C(\infty),\quad \frac{1}{n}\sum_{i=1}^n C_iC_i'g(Q_i)\to_p\Sigma_{Cg}(\infty),\quad\text{and}\quad \frac{1}{\sqrt{n}}\sum_{i=1}^n C_iU_i\Rightarrow G(\infty).$$

Hence, if we define $\mu_X(\theta)$ to equal the first term, we obtain the convergence (14) for $X_n$.

Likewise, standard regression algebra (e.g. the FWL theorem) shows that

$$\sqrt{n}\hat{\delta}(\theta) \equiv \mathcal{A}_n(\theta)^{-1}[\mathcal{B}_n(\theta)+\mathcal{C}_n(\theta)],$$

for

$$\mathcal{A}_n(\theta) \equiv n^{-1}\sum_{i=1}^n C_i C_i' 1(Q_i > \theta) - \left(n^{-1}\sum_{i=1}^n C_i C_i' 1(Q_i > \theta)\right)\left(n^{-1}\sum_{i=1}^n C_i C_i'\right)^{-1}\left(n^{-1}\sum_{i=1}^n C_i C_i' 1(Q_i > \theta)\right),$$
$$\mathcal{B}_n(\theta) \equiv n^{-1}\sum_{i=1}^n C_i C_i' 1(Q_i > \theta) g(Q_i) - \left(n^{-1}\sum_{i=1}^n C_i C_i' 1(Q_i > \theta)\right)\left(n^{-1}\sum_{i=1}^n C_i C_i'\right)^{-1}\left(n^{-1}\sum_{i=1}^n C_i C_i' g(Q_i)\right),$$
$$\mathcal{C}_n(\theta) \equiv n^{-1/2}\sum_{i=1}^n C_i U_i 1(Q_i > \theta) - \left(n^{-1}\sum_{i=1}^n C_i C_i' 1(Q_i > \theta)\right)\left(n^{-1}\sum_{i=1}^n C_i C_i'\right)^{-1}\left(n^{-1/2}\sum_{i=1}^n C_i U_i\right).$$

Under (12) and (13), however, the continuous mapping theorem implies that

$$\mathcal{A}_n(\theta) \to_p \Sigma_C(\infty) - \Sigma_C(\theta) - (\Sigma_C(\infty) - \Sigma_C(\theta))\Sigma_C(\infty)^{-1}(\Sigma_C(\infty) - \Sigma_C(\theta)) \equiv \mathcal{A}(\theta),$$
$$\mathcal{B}_n(\theta) \to_p \Sigma_{Cg}(\infty) - \Sigma_{Cg}(\theta) - (\Sigma_C(\infty) - \Sigma_C(\theta))\Sigma_C(\infty)^{-1}\Sigma_{Cg}(\infty) \equiv \mathcal{B}(\theta),$$
$$\mathcal{C}_n(\theta) \Rightarrow G(\infty) - G(\theta) - (\Sigma_C(\infty) - \Sigma_C(\theta))\Sigma_C(\infty)^{-1}G(\infty) \equiv \mathcal{C}(\theta)$$

all uniformly on $\tilde{\Theta}$, where this convergence holds jointly with that for $X_n$. By another application of the continuous mapping theorem,

$$Y_n(\theta) = e_j'\sqrt{n}\hat{\delta}(\theta) \Rightarrow Y(\theta) = e_j'\mathcal{A}(\theta)^{-1}[\mathcal{B}(\theta)+\mathcal{C}(\theta)].$$

Hence, if we define $\mu_Y(\theta) = e_j'\mathcal{A}(\theta)^{-1}\mathcal{B}(\theta)$, then $\mu_{Y,n}(\theta) \to \mu_Y(\theta)$ uniformly in $\theta \in \tilde{\Theta}$ and we obtain the convergence (14), as desired.

**Additional Conditioning Events** Arguments as in the proof of Proposition 4 show that if we define

$$\bar{A}(\tilde{\theta}) \equiv \Sigma_Y(\tilde{\theta})^{-2}\sum_{i=1}^{d_X}\Sigma_{XY,i}(\tilde{\theta})^2,$$

$$\bar{B}_Z(\tilde{\theta}) \equiv 2\Sigma_Y(\tilde{\theta})^{-1}\sum_{i=1}^{d_X}\Sigma_{XY,i}(\tilde{\theta})Z_{\tilde{\theta},i}(\tilde{\theta}),$$

$$\bar{C}_Z(\tilde{\theta}) \equiv \sum_{i=1}^{d_X} Z_{\tilde{\theta},i}(\tilde{\theta})^2 - c, \quad \bar{D}_Z(\tilde{\theta}) \equiv \bar{B}_Z(\tilde{\theta})^2 - 4\bar{A}(\tilde{\theta})\bar{C}_Z(\tilde{\theta}),$$

$$\{\|X(\tilde{\theta})\|^2 \geq c\} = \left\{ Y(\tilde{\theta}) \leq \frac{-\bar{B}_Z(\tilde{\theta}) - \sqrt{D_Z(\tilde{\theta})}}{2\bar{A}(\tilde{\theta})} \text{ or } Y(\tilde{\theta}) \geq \frac{-\bar{B}_Z(\tilde{\theta}) + \sqrt{D_Z(\tilde{\theta})}}{2\bar{A}(\tilde{\theta})}, D_Z(\tilde{\theta}) \geq 0 \right\}$$
$$\cap \{\bar{C}_Z(\tilde{\theta}) \geq 0, D_Z(\tilde{\theta}) < 0\}$$

if $\bar{A}(\tilde{\theta}) > 0$ and $\{\|X(\tilde{\theta})\|^2 \geq c\} = \{\bar{C}_Z(\tilde{\theta}) \geq 0\}$ if $\bar{A}(\tilde{\theta}) = 0$, since $\bar{A}(\tilde{\theta}) \geq 0$ by definition. Then for

$$\bar{\mathcal{L}}(Z_{\tilde{\theta}}) \equiv \frac{-\bar{B}_Z(\tilde{\theta}) - \sqrt{D_Z(\tilde{\theta})}}{2\bar{A}(\tilde{\theta})},$$
$$\bar{\mathcal{U}}(Z_{\tilde{\theta}}) \equiv \frac{-\bar{B}_Z(\tilde{\theta}) + \sqrt{D_Z(\tilde{\theta})}}{2\bar{A}(\tilde{\theta})},$$
$$\bar{\mathcal{V}}(Z_{\tilde{\theta}}) \equiv [1\{\bar{A}(\tilde{\theta}) = 0\} + 1\{\bar{A}(\tilde{\theta}) > 0, D_Z(\tilde{\theta}) < 0\}]\bar{C}_Z(\tilde{\theta}),$$

we see that if $\bar{\mathcal{V}}(Z_{\tilde{\theta}}) \geq 0$ then $\mathcal{Y}_\gamma(1, Z_{\tilde{\theta}}) = \big(\bar{\mathcal{L}}(Z_{\tilde{\theta}}), \bar{\mathcal{U}}(Z_{\tilde{\theta}})\big)^c$, while $\mathcal{Y}_\gamma(1, Z_{\tilde{\theta}}) = \emptyset$ otherwise.

# 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 sets and estimators as well as easy-to-implement confidence sets and estimators that dominate their conventional split-sample counterparts in the asymptotic version of the split-sample problem.

**The Split-Sample Limit Experiment** Let $\tau$ 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

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

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

with $[a]$ denoting the nearest integer to $a \in \mathbb{R}$. Finally, let $\hat{\theta}_n^1 = \arg\max_{\theta\in\Theta} X_n^1(\theta)$ or $\hat{\theta}_n^1 = \operatorname{argmax}_{\theta\in\Theta}\|X_n^1(\theta)\|$ 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

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

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

and

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

or

$$\hat{\theta}^1 = \operatorname{argmax}_{\theta\in\Theta} \big\|X^1(\theta)\big\|,$$

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.[29]

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 sets

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

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 $\mu$ 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

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

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

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 $\begin{pmatrix}X^*\\Y^*\end{pmatrix}=\left(\begin{pmatrix}X^1\\Y^1\end{pmatrix}+c\begin{pmatrix}X^2\\Y^2\end{pmatrix}\right)$ is sufficient for $\mu$. 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 $\mu$. Thus, to study questions of optimality it is without loss to limit attention to confidence sets and estimators that depend only on $(X^*,Y^*)$.

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

Note that the unconditional density of $(X^*,Y^*)$ is proportional to

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

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

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

where we have used sufficiency to drop dependence of the numerator on $\mu$.

This joint distribution has the same exponential family structure used to derive the optimal estimators and confidence sets 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

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

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

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

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

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

This expression shows that when we are interested in inference on $\mu_Y(\tilde{\theta})$ conditional on $\left\{\hat{\theta}^1=\tilde{\theta}\right\}$, $\mu_{Z^*}$ 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 $\alpha$ 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 $\zeta$ 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 $\mu_X$. 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 set $CS_{SS}$ 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 $\mu_X$ 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 computational method for computing this object.

In analogy with full sample inference, let

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

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 $k=1,...,K$ 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 $\alpha$ 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 set 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 $\zeta$ distributed according to $F^A_{SS}(\cdot;\mu_{Y,0},\tilde{\theta},z)$. These dominating procedures condition on $Z^1_{\tilde{\theta}}$ rather than $(X^1,Y^1)$, and so unlike conventional sample splitting continue to treat $(X^1,Y^1)$ as random for inference.

# 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 large classes of data generating processes. 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 sets $\hat{\mu}_{\alpha,n}$, $\hat{\mu}^H_{\alpha,n}$, $CS_{ET,n}$, $CS^H_{ET,n}$, $CS_{U,n}$, $CS^H_{U,n}$ and $CS_{P,n}$ are equal to their asymptotic counterparts $\hat{\mu}_{\alpha}$, $\hat{\mu}^H_{\alpha}$, $CS_{ET}$, $CS^H_{ET}$, $CS_{U}$, $CS^H_{U}$ and $CS_{P}$ after replacing $X$, $Y$, $\Sigma$ with $X_n$, $Y_n$, $\widehat{\Sigma}_n$.

With this notation, we aim to prove, for example, that for $\hat{\mu}_{\alpha,n}$ our $\alpha$-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 $\alpha$-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 not our norm maximization problem, and impose an assumption that implies bounded asymptotic means (analogous to our Assumption 5 below). Since we do not impose this assumption in our analysis of level-maximization, neither our norm nor level maximization results are 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 An-

drews 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 sets 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 $A$,*

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

*and*

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

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 $\xi_P \sim N(0,\Sigma(P))$,*

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

*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 9**

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

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}_n} \left| Pr_P\left\{ \hat{\mu}_{\alpha,n} \geq \mu_{Y,n}\left(\hat{\theta}_n; P\right) | \hat{\theta}_n = \tilde{\theta} \right\} - \alpha \right| Pr_P\left\{ \hat{\theta}_n = \tilde{\theta} \right\} = 0, \tag{31}$$

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

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}_n} \left| Pr_P\left\{ \hat{\mu}_{\alpha,n} \geq \mu_{Y,n}\left(\hat{\theta}_n; P\right) \right\} - \alpha \right| = 0. \tag{32}$$

**Corollary 1**

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

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

*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 10**

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

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}_n} \left| Pr_P\left\{ \mu_{Y,n}\left(\hat{\theta}_n;P\right) \in CS_{U,n} | \hat{\theta}_n = \tilde{\theta} \right\} - (1-\alpha) \right| Pr_P\left\{ \hat{\theta}_n = \tilde{\theta} \right\} = 0, \tag{33}$$

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

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}_n} \left| Pr_P\left\{ \mu_{Y,n}\left(\hat{\theta}_n;P\right) \in CS_{U,n} \right\} - (1-\alpha) \right| = 0. \tag{34}$$

**Proposition 11**

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

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}_n} Pr_P\left\{ \mu_{Y,n}\left(\hat{\theta}_n;P\right) \in CS_{P,n} \right\} \geq 1-\alpha. \tag{35}$$

**Proposition 12**

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

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}_n} \left| Pr_P\left\{ \hat{\mu}^H_{\alpha,n} \geq \mu_{Y,n}\left(\hat{\theta}_n;P\right) | C^H_n\left(\tilde{\theta};P\right) = 1 \right\} - \alpha \right| E_P\left\{ C^H_n\left(\tilde{\theta};P\right) \right\} = 0, \tag{36}$$

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

$$\operatorname*{limsup}_{n\to\infty} \sup_{P\in\mathcal{P}_n} \left| Pr_P\left\{ \hat{\mu}^H_{\alpha,n} \geq \mu_{Y,n}\left(\hat{\theta}_n;P\right) \right\} - \alpha \right| \leq \max\{\alpha, 1-\alpha\}\beta. \tag{37}$$

**Corollary 2**

*Under Assumptions 2-4, for* $\hat{\theta}_n = \arg\max_\theta X_n(\theta)$ *and* $CS^H_{ET,n}$ *the level* $1-\alpha$ *equal-tailed*

*hybrid confidence set based on initial confidence set* $CS_{P,n}^{\beta}$,

$$\lim_{n\to\infty}\sup_{P\in\mathcal{P}_n}\left|Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS_{ET,n}^H|C_n^H\left(\tilde{\theta};P\right)=1\right\}-\frac{1-\alpha}{1-\beta}\right|E_P\left\{C_n^H\left(\tilde{\theta};P\right)\right\}=0, \tag{38}$$

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

$$\liminf_{n\to\infty}\inf_{P\in\mathcal{P}_n}Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS_{ET,n}^H\right\}\geq 1-\alpha, \tag{39}$$

*and*

$$\limsup_{n\to\infty}\sup_{P\in\mathcal{P}_n}Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS_{ET,n}^H\right\}\leq\frac{1-\alpha}{1-\beta}\leq 1-\alpha+\beta. \tag{40}$$

**Proposition 13**

*Under Assumptions 2-4, for* $\hat{\theta}_n=\text{argmax}_\theta X_n(\theta)$ *and* $CS_{U,n}^H$ *the level* $1-\alpha$ *unbiased hybrid confidence set based on initial confidence set* $CS_{P,n}^{\beta}$,

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

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

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

*and*

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

### 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 sets* $CS_n$, *any sequence of sets* $\mathcal{C}_n(P)$ *indexed by* $P$, $C_n(P)=1\left\{\left(X_n,Y_n,\widehat{\Sigma}_n\right)\in\mathcal{C}_n(P)\right\}$, *and any constant* $\alpha$, *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*

$$\mathcal{S}=\left\{\Sigma:1/\bar{\lambda}\leq\lambda_{\min}(\Sigma_X)\leq\lambda_{\max}(\Sigma_X)\leq\bar{\lambda},1/\bar{\lambda}\leq\Sigma_Y(\theta)\leq\bar{\lambda}\right\}, \tag{41}$$

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*

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

**Lemma 6**

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

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

*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

$$\mathcal{L}\left(\tilde{\theta},Z,\Sigma\right)=\max_{\theta\in\Theta:\Sigma_{XY}(\tilde{\theta})>\Sigma_{XY}(\tilde{\theta},\theta)}\frac{\Sigma_Y\left(\tilde{\theta}\right)\left(Z(\theta)-Z\left(\tilde{\theta}\right)\right)}{\Sigma_{XY}\left(\tilde{\theta}\right)-\Sigma_{XY}\left(\tilde{\theta},\theta\right)} \tag{43}$$

$$\mathcal{U}\left(\tilde{\theta},Z,\Sigma\right)=\min_{\theta\in\Theta:\Sigma_{XY}\left(\tilde{\theta}\right)<\Sigma_{XY}\left(\tilde{\theta},\theta\right)}\frac{\Sigma_Y\left(\tilde{\theta}\right)\left(Z(\theta)-Z\left(\tilde{\theta}\right)\right)}{\Sigma_{XY}\left(\tilde{\theta}\right)-\Sigma_{XY}\left(\tilde{\theta},\theta\right)}, \tag{44}$$

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

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

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 (43) and (44), 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 $(Y^*,X^*,\Sigma^*)$ for*

$$\left(\begin{array}{c} X^* \\ Y^* \end{array}\right)\sim N(\mu^*,\Sigma^*)$$

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

**Lemma 9**

*For* $F_N$ *again the standard normal distribution function, the function*

$$F_{TN}(Y(\theta);\mu,\Sigma_Y(\theta),\mathcal{L},\mathcal{U}) = \frac{F_N\left(\frac{Y(\theta)\wedge\mathcal{U}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right) - F_N\left(\frac{\mathcal{L}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)}{F_N\left(\frac{\mathcal{U}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right) - F_N\left(\frac{\mathcal{L}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)} 1(Y(\theta) \geq \mathcal{L}) \tag{45}$$

*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*

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

*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

$$\liminf_{n\to\infty}\inf_{P\in\mathcal{P}_n}\left(Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS_n|C_n(P)=1\right\}-\alpha\right)Pr_P\{C_n(P)=1\}\geq 0 \tag{46}$$

and

$$\limsup_{n\to\infty}\sup_{P\in\mathcal{P}_n}\left(Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CS_n|C_n(P)=1\right\}-\alpha\right)Pr_P\{C_n(P)=1\}\leq 0. \tag{47}$$

We prove that to show (46), it suffices to show that for all $\{n_s\}$, $\{P_{n_s}\}$ satisfying conditions (1)-(3) of the lemma,

$$\liminf_{s\to\infty}Pr_{P_{n_s}}\left\{\mu_{Y,n_s}\left(\hat{\theta}_{n_s};P_{n_s}\right)\in CS_{n_s}|C_{n_s}(P_{n_s})=1\right\}\geq\alpha. \tag{48}$$

An argument along the same lines implies that to prove (47) it suffices to show that

$$\limsup_{s\to\infty}Pr_{P_{n_s}}\left\{\mu_{Y,n_s}\left(\hat{\theta}_{n_s};P_{n_s}\right)\in CS_{n_s}|C_{n_s}(P_{n_s})=1\right\}\leq\alpha. \tag{49}$$

Note, however, that (48) and (49) together are equivalent to (42).

Towards contradiction, suppose that (46) 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 (48) 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

$$\limsup_{q\to\infty}\left(Pr_{P_{n_q}}\left\{\mu_{Y,n_q}\left(\hat{\theta}_{n_q};P_{n_q}\right)\in CS_{n_q}|C_{n_q}\left(P_{n_q}\right)=1\right\}-\alpha\right)Pr_{P_{n_q}}\left\{C_{n_q}\left(P_{n_q}\right)=1\right\}<-\varepsilon. \tag{50}$$

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 (41) 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, (50) 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, (48) 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 (50). Hence, we have established a contradiction and so proved that (48) for all subsequences satisfying conditions (1)-(3) of the lemma implies (46). An argument along the same lines shows that (49) along all subsequences satisfying conditions (1)-(3) of lemma implies (47). □

**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

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

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

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

and

$$\liminf_{n\to\infty}\inf_{P\in\mathcal{P}} Pr_P\Big\{\mu_{Y,n}\Big(\hat{\theta}_n;P\Big)\in CS_n\Big\}-(1-\alpha)$$

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

$$=\liminf_{n\to\infty}\inf_{P\in\mathcal{P}_n}\Big(Pr_P\Big\{\mu_{Y,n}\Big(\hat{\theta}_n;P\Big)\in CS_n|C_{n,J+1}(P)=1\Big\}-(1-\alpha)\Big)Pr_P\{C_{n,J+1}(P)=1\}$$

$$\geq-(1-\alpha)\limsup_{n\to\infty}\sup_{P\in\mathcal{P}_n} Pr_P\{C_{n,J+1}(P)=1\}$$

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

which immediately implies that

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

Likewise,

$$\limsup_{n\to\infty}\sup_{P\in\mathcal{P}_n} Pr_P\Big\{\mu_{Y,n}\Big(\hat{\theta}_n;P\Big)\in CS_n\Big\}-(1-\alpha)$$

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

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

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

This immediately implies that

$$\limsup_{n\to\infty}\sup_{P\in\mathcal{P}_n} Pr_P\Big\{\mu_{Y,n}\Big(\hat{\theta}_n;P\Big)\in CS_n\Big\}\leq1-\alpha\cdot\liminf_{n\to\infty}\inf_{P\in\mathcal{P}_n}\sum_{j=1}^{J}Pr_P\{C_{n,j}(P)=1\},$$

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 $(X^*,Y^*,\Sigma^*)$. 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 $\Sigma^*_X$ 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

$$\left|\frac{\widehat{\Sigma}_{Y,n_s}\left(\tilde{\theta}\right)\left(Z^*_{n_s,\tilde{\theta}}(\theta)-Z^*_{n_s,\tilde{\theta}}\left(\tilde{\theta}\right)\right)}{\widehat{\Sigma}_{XY,n_s}\left(\tilde{\theta}\right)-\widehat{\Sigma}_{XY,n_s}\left(\tilde{\theta},\theta\right)}\right|\to\infty. \tag{51}$$

Let us define

$$\Theta^*\left(\tilde{\theta}\right)=\left\{\theta\in\Theta\backslash\tilde{\theta}:\Sigma^*_{XY}\left(\tilde{\theta}\right)-\Sigma^*_{XY}\left(\tilde{\theta},\theta\right)\neq 0\right\}.$$

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 (51), the same convergence holds when we minimize and maximize over $\Theta$ 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. □

**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. □

**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:

$$F_N\left(\frac{\mathcal{U}\wedge c_u-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-F_N\left(\frac{\mathcal{L}\vee c_l-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-(1-\alpha)\left(F_N\left(\frac{\mathcal{U}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-F_N\left(\frac{\mathcal{L}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)\right)=0 \qquad (52)$$

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 (52), 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

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

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

This implies that

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

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

$$(c_l(0,\Sigma_Y(\theta),\mathcal{L},\mathcal{U}),c_u(0,\Sigma_Y(\theta),\mathcal{L},\mathcal{U})). \tag{53}$$

Recall that (53) solves

$$Pr\{\zeta\in[c_l,c_u]\}=(1-\alpha) \tag{54}$$

and

$$E[\zeta 1\{\zeta\in[c_l,c_u]\}]=(1-\alpha)E[\zeta] \tag{55}$$

for $\zeta\sim\xi|\xi\in[\mathcal{L},\mathcal{U}]$ where $\xi\sim N(0,\Sigma_Y(\theta))$. Note, however, that since $\mathcal{L}<\mathcal{U}$, (54) 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 (53), $\mathcal{L}<c_l<c_u<\mathcal{U}$.

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

$$\frac{\partial}{\partial c'}g\Big(c;0,\sqrt{\Sigma_Y(\theta)},\mathcal{L},\mathcal{U}\Big)=$$

$$\begin{pmatrix} -1(c_l>\mathcal{L})f_N\Big(c_l/\sqrt{\Sigma_Y(\theta)}\Big)/\sqrt{\Sigma_Y(\theta)} & 1(c_u<\mathcal{U})f_N\Big(c_u/\sqrt{\Sigma_Y(\theta)}\Big)/\sqrt{\Sigma_Y(\theta)} \\ -1(c_l>\mathcal{L})c_l f_N\Big(c_l/\sqrt{\Sigma_Y(\theta)}\Big)/\Sigma_Y(\theta) & 1(c_u<\mathcal{U})c_u f_N\Big(c_u/\sqrt{\Sigma_Y(\theta)}\Big)/\Sigma_Y(\theta) \end{pmatrix}.$$

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

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

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

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

as we wanted to show. □

### D.1.5 Proofs for Uniformity Results

**Proof of Proposition 9** Note that

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

for $CS_{U,-,n}=(-\infty,\hat{\mu}_{\alpha,n}]$. Hence, by Lemma 5, to prove that (31) holds it suffices to show that for all $\{n_s\}$ and $\{P_{n_s}\}$ such that conditions (1)-(3) of the lemma hold with $C_n(P)=1\Big\{\hat{\theta}_n=\tilde{\theta}\Big\}$, we have

$$\lim_{s\to\infty}Pr_{P_{n_s}}\Big\{\hat{\mu}_{Y,n_s}\Big(\hat{\theta}_{n_s};P_{n_s}\Big)\in CS_{U,-,n_s}|\hat{\theta}_{n_s}=\tilde{\theta}\Big\}=\alpha. \tag{56}$$

To this end, recall that for $F_{TN}(Y(\theta);\mu,\Sigma_Y(\theta),\mathcal{L},\mathcal{U})$ as defined in (45), the estimator $\hat{\mu}_{\alpha,n}$ solves

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

where $(\mathcal{L}_n,\mathcal{U}_n)$ are defined following Lemma 7. This cdf is strictly decreasing in $\mu$ as argued in the proof of Proposition 8, and is increasing in $Y_n\Big(\hat{\theta}\Big)$. Hence, $\hat{\mu}_{\alpha,n}\geq\mu_{Y,n}\Big(\hat{\theta}_n;P\Big)$ if and only if

$$F_{TN}\Big(Y_n\Big(\hat{\theta}_n\Big);\mu_{Y,n}\Big(\hat{\theta}_n;P\Big),\widehat{\Sigma}_{Y,n}\Big(\hat{\theta}_n\Big),\mathcal{L}_n,\mathcal{U}_n\Big)\geq1-\alpha.$$

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

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

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 (56).

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 (32). □

**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 9 and Lemma 6. □

**Proof of Proposition 10** Note that by the definition of $CS_{U,n}$

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

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,

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

By Lemma 5, to prove that (33) 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

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

To this end, note that by Lemma 8,

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

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

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

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,

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

as we wanted to show, so (33) follows by Lemma 5.

Since this result again holds for all $\tilde{\theta}\in\Theta$, (34) follows immediately by the same argument as in the proof of Proposition 9. □

**Proof of Proposition 11** By the same argument as in the proof of Lemma 5, to show that (35) holds it suffices to show that for all $\{n_s\}$, $\{P_{n_s}\}$ satisfying conditions (1)-(3) of Lemma 5,

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

To this end, note that

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

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 $\Sigma$ on $\mathcal{S}$ as defined in (41). Hence, for all $\theta$, $c_\alpha(\Sigma_Y)\sqrt{\Sigma_Y(\theta)}$ 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 $Pr\left\{\left|Y^*\left(\hat{\theta}\right)\right|-c_\alpha(\Sigma^*_Y)\sqrt{\Sigma^*_Y\left(\hat{\theta}\right)}=0\right\}=0$,

$$Pr_{P_{n_s}}\left\{\mu_{Y,n_s}\left(\hat{\theta}_{n_s};P_{n_s}\right)\in CS_{P,n_s}\right\}\to Pr\left\{Y^*\left(\hat{\theta}\right)\in\left[-c_\alpha(\Sigma^*_Y)\sqrt{\Sigma^*_Y\left(\hat{\theta}\right)},c_\alpha(\Sigma^*_Y)\sqrt{\Sigma^*_Y\left(\hat{\theta}\right)}\right]\right\}\qquad(57)$$

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

**Proof of Proposition 12** Note that

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

for $CS^H_{U,-,n}=(-\infty,\hat{\mu}^H_{\alpha,n}]$. Hence, by Lemma 5, to prove that (36) 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 (45), $\hat{\mu}^H_{\alpha,n}$ 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

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

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

The proof of Proposition 8 shows that $F_{TN}\left(Y_n\left(\hat{\theta}_n\right); \mu, \widehat{\Sigma}_{Y,n}\left(\hat{\theta}_n\right), \mathcal{L}_n^H(\mu), \mathcal{U}_n^H(\mu)\right)$ is strictly decreasing in $\mu$, so for a given value $\mu_{Y,0}$,

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

As in the proof of Proposition 9

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

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

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

Lemma 8 implies that

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

where $\mathcal{L}^{H*}$ and $\mathcal{U}^{H*}$ are equal to $\mathcal{L}_n^{H*}$ and $\mathcal{U}_n^{H*}$ after replacing $(X_n, Y_n, \widehat{\Sigma}_n)$ with $(X, Y, \Sigma^*)$. Then by the continuous mapping theorem and (57),

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

Hence, by the same argument as in the proof of Proposition 9,

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

as we aimed to show.

To prove (37), note that for $\widetilde{CS}^H_{U,+,n}=(\hat{\mu}^H_{\alpha,n},\infty)$,

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

and thus that the argument above proves that

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

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

$$\sum_{\tilde{\theta}\in\Theta}Pr_P\Big\{\hat{\theta}_{n_s}=\tilde{\theta},\mu_{Y,n_s}\Big(\hat{\theta}_{n_s};P_{n_s}\Big)\in CS^\beta_{P,n_s}\Big\}=Pr_P\Big\{\mu_{Y,n_s}\Big(\hat{\theta}_{n_s};P_{n_s}\Big)\in CS^\beta_{P,n_s}\Big\}+o(1),\qquad(58)$$

and Proposition 11 shows that

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

Lemma 6 together with (36) implies that

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

and

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

from which the second result of the proposition follows immediately. □

**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,

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

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

so Proposition 12 immediately implies (38).

Equation (58) in the proof of Proposition 12 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 (39) 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,

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

By the first part of the proposition, this implies that

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

so (40) holds as well. □

**Proof of Proposition 13** The first part of the result follows by the same argument as in the proof of Proposition 10, where as in the proof of Proposition 12 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. □

### D.2 Asymptotic Validity of Norm-Maximization

We next turn to the asymptotic validity of our results in norm-maximization settings. As discussed in the main text and Appendix B.2, the norm-maximization problem arises when we follow Elliott and Müller (2007, 2014) and Wang (2018) and model the degree of parameter instability as shrinking with the sample size. If we instead take the degree of parameter instability to be fixed, one can show that the threshold regression and structural break models reduce to level maximization asymptotically.

The issue here is similar to the difference in the asymptotic distribution of the Vuong

(1989) test between the nested and non-nested cases. As this analogy suggests, it may be possible to develop asymptotic results for threshold regression and structural break models that, analogous to the results of Shi (2015) and Schennach and Wilhelm (2017) for the Vuong test, cover cases with both fixed and local parameter instability. We are unaware of such results for existing procedures inthreshold regression and structural break literatures, however, and this point is far afield from our primary focus in this project. Hence, in this section we follow Elliott and Müller (2007, 2014) and Wang (2018) and limit attention to cases with local parameter instability and, refer readers interested in fixed parameter instability to the level-maximization results discussed above.

Section D.2.1 states the bounded asymptotic means assumption. Section D.2.2 then states our uniformity results for norm-maximization settings. Section D.2.3 collects additional technical lemmas for this setting. Finally, Sections D.2.4 and D.2.5 collect proofs for the lemmas and the uniformity results, respectively.

### D.2.1 Assumptions

To prove uniform asymptotic validity for norm maximization, we will continue to impose Assumptions 2-4 of the last section. To limit attention to the case with local parameter instability, we further impose the following assumption.

**Assumption 5**
*There exists a finite constant $C>0$ such that*

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_n} (\|\mu_{X,n}(P)\|+\|\mu_{Y,n}(P)\|)\leq C.$$

This assumption requires that $\|\mu_{X,n}(P)\|$ and $\|\mu_{Y,n}(P)\|$ be uniformly bounded over $\mathcal{P}_n$ by a constant that does not depend on the sample size. Given the scaling of $(X_n,Y_n)$ in our threshold regression and structural break examples, this corresponds to the case with local parameter instability. It may be possible to relax this assumption, but it holds in all settings we have encountered that give rise to the norm-maximization problem asymptotically. Specifically, note that Assumption 5 holds if we take $\mathcal{P}_n$ to correspond to any finite collection of local sequences of the sort studied by Elliott and Müller (2007, 2014) and Wang (2018). If we instead consider nonlocal sequences, then as discussed above we instead obtain a level-maximization problem asymptotically.

### D.2.2 Norm Maximization Uniformity Results

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

**Proposition 14**

*Under Assumptions 2-5, for* $\hat{\theta}_n = \arg\max_\theta \|X_n(\theta)\|$ *and* $\hat{\mu}_{\alpha,n}$ *the* $\alpha$*-quantile unbiased estimator,*

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}_n} \left| Pr_P\left\{\hat{\mu}_{\alpha,n} \geq \mu_{Y,n}\left(\hat{\theta}_n;P\right) | \hat{\theta}_n = \tilde{\theta}\right\} - \alpha \right| Pr_P\left\{\hat{\theta}_n = \tilde{\theta}\right\} = 0, \tag{59}$$

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

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}_n} \left| Pr_P\left\{\hat{\mu}_{\alpha,n} \geq \mu_{Y,n}\left(\hat{\theta}_n;P\right)\right\} - \alpha \right| = 0. \tag{60}$$

**Corollary 3**

*Under Assumptions 2-5, for* $\hat{\theta}_n = \text{argmax}_\theta \|X_n(\theta)\|$ *and* $CS_{ET,n}$ *the level* $1-\alpha$ *equal-tailed confidence set,*

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

*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 15**

*Under Assumptions 2-5, for* $\hat{\theta}_n = \text{argmax}_\theta \|X_n(\theta)\|$ *and* $CS_{U,n}$ *the level* $1-\alpha$ *unbiased confidence set,*

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}_n} \left| Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right) \in CS_{U,n} | \hat{\theta}_n = \tilde{\theta}\right\} - (1-\alpha) \right| Pr_P\left\{\hat{\theta}_n = \tilde{\theta}\right\} = 0, \tag{61}$$

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

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}_n} \left| Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right) \in CS_{U,n}\right\} - (1-\alpha) \right| = 0. \tag{62}$$

**Proposition 16**

*Under Assumptions 2-5, for* $\hat{\theta}_n = \text{argmax}_\theta \|X_n(\theta)\|$ *and* $CS_{P,n}$ *the level* $1-\alpha$ *projection confidence set,*

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

**Proposition 17**

*Under Assumptions 2-5, for* $\hat{\theta}_n = \text{argmax}_\theta \|X_n(\theta)\|$*,* $\hat{\mu}^H_{\alpha,n}$ *the* $\alpha$*-quantile unbiased hybrid*

*estimator based on initial confidence set* $CS^{\beta}_{P,n}$, *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*

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

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

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

**Corollary 4**

*Under Assumptions 2-5, for* $\hat{\theta}_n=\operatorname{argmax}_\theta\|X_n(\theta)\|$ *and* $CS^H_{ET,n}$ *the level* $1-\alpha$ *equal-tailed hybrid confidence set based on initial confidence set* $CS^{\beta}_{P,n}$,

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

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

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

*and*

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

**Proposition 18**

*Under Assumptions 2-5, for* $\hat{\theta}_n=\operatorname{argmax}_\theta\|X_n(\theta)\|$ *and* $CS^H_{U,n}$ *the level* $1-\alpha$ *unbiased hybrid confidence set based on initial confidence set* $CS^{\beta}_{P,n}$,

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

*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*

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

### D.2.3 Auxiliary Lemmas

To prove uniformity in norm-maximization settings, we rely on some of the lemmas in Section D.1.3 along with a few additional results.

**Lemma 11**

*Under Assumptions 3 and 5, for any sequence of confidence sets $CS_n$, any sequence of sets $\mathcal{C}_n(P)$ indexed by $P$, $C_n(P)=1\left\{\left(X_n,Y_n,\widehat{\Sigma}_n\right)\in\mathcal{C}_n(P)\right\}$, and any constant $\alpha$, 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 $\mathcal{S}$ as defined in (41)*
2. *$(\mu_{X,n_s}(P_{n_s}),\mu_{Y,n_s}(P_{n_s}))\to(\mu_X^*,\mu_Y^*)$ for $(\mu_X^*,\mu_Y^*)$ finite*

*we have*

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

To state the next result, for $Z_{\tilde{\theta},n,j}$ the $j$th element of $Z_{\tilde{\theta},n}$ as defined in Lemma 7, let us define

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

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

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

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

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

and

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

Based on these objects, let us further define

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

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

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

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

**Lemma 12**

*Under Assumptions 2 and 4, for any $\{n_s\}$ and $\{P_{n_s}\}$ satisfying conditions (1) and (2) of Lemma 11,*

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

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

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

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

**Lemma 13**

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

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

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

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

To state the next lemma, let

$$\left(c_l\left(\mu,\Sigma_Y(\theta),\mathcal{Y}^K\right),c_u\left(\mu,\Sigma_Y(\theta),\mathcal{Y}^K\right)\right) \tag{63}$$

solve

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

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

for $\zeta$ as in Lemma 13.

**Lemma 14**
*The function (63) is continuous in* $\left(\mu,\Sigma_Y(\theta),\mathcal{Y}^K\right)$ *for Lebesgue almost-every* $\left\{\ell^k,u^k\right\}_{k=1}^K$ *on the set*

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

*Moreover, if we fix any* $(\mu,\Sigma_Y(\theta))$ *in this set, and fix all but one element of* $\left\{\ell^k,u^k\right\}_{k=1}^K$, *(63) is almost-everywhere continuous in the remaining element.*

### D.2.4 Proofs of Auxiliary Lemmas

**Proof of Lemma 11** Follows by the same argument as in the proof of Lemma 5.

**Proof of Lemma 12** Note that Assumption 4 along with condition (2) of Lemma 11 imply that

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

while Assumption 2 implies that $\widehat{\Sigma}_{n_s}\to_p\Sigma^*$.

If we define

$$\left(A^*\left(\tilde{\theta},\theta\right),B_Z^*\left(\tilde{\theta},\theta\right),C_Z^*\left(\tilde{\theta},\theta\right),D_Z^*\left(\tilde{\theta},\theta\right),G_Z^*\left(\tilde{\theta},\theta\right),K_Z^*\left(\tilde{\theta},\theta\right),H_Z^*\left(\tilde{\theta},\theta\right)\right)$$

as the analog of

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

based on $(X^*,Y^*,\Sigma^*)$, the continuous mapping theorem implies that

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

where this convergence holds jointly over all $\left(\theta,\tilde{\theta}\right)\in\Theta^2$. If $A^*\left(\tilde{\theta},\theta\right)\neq 0$, another application of the continuous mapping theorem implies that[30]

$$\left(D_{Z,n_s}\left(\tilde{\theta},\theta\right),G_{Z,n_s}\left(\tilde{\theta},\theta\right),K_{Z,n_s}\left(\tilde{\theta},\theta\right)\right)\rightarrow_d\left(D_Z^*\left(\tilde{\theta},\theta\right),G_Z^*\left(\tilde{\theta},\theta\right),K_Z^*\left(\tilde{\theta},\theta\right)\right).$$

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

$$Z_{\tilde{\theta},j}^*(\theta)=X_j^*(\theta)-\frac{\Sigma_{XY,j}^*\left(\theta,\tilde{\theta}\right)}{\Sigma_Y^*\left(\tilde{\theta}\right)}Y^*\left(\tilde{\theta}\right)=X_j^*(\theta)-\frac{\Sigma_{XY,j}^*\left(\tilde{\theta}\right)}{\Sigma_Y^*\left(\tilde{\theta}\right)}Y^*\left(\tilde{\theta}\right).$$

Hence, in this setting

$$B_Z^*\left(\tilde{\theta},\theta\right)=2\Sigma_Y\left(\tilde{\theta}\right)^{-2}\sum_{j=1}^{d_X}\left[X_j^*\left(\tilde{\theta}\right)-X_j^*(\theta)\right]$$

and condition (1) of Lemma 11 implies that $Pr\left\{B_Z^*\left(\tilde{\theta},\theta\right)=0\right\}=0$ for all $\theta\neq\tilde{\theta}$. Hence, $Pr\left\{D_Z^*\left(\tilde{\theta},\theta\right)>0\right\}=1$. Moreover, note that for $b\neq 0$ and all $c$

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

while

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

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

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

[30]Note that we allow the possibility that $\left(D_{Z,n}\left(\tilde{\theta},\theta\right),D_Z^*\left(\tilde{\theta},\theta\right)\right)$ may be negative, so $\left(G_{Z,n}\left(\tilde{\theta},\theta\right),K_{Z,n}\left(\tilde{\theta},\theta\right)\right)$ and $\left(G_Z^*\left(\tilde{\theta},\theta\right),K_Z^*\left(\tilde{\theta},\theta\right)\right)$ may be complex-valued.

and

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

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

$$H_{Z,n_s}\left(\tilde{\theta},\theta\right)\rightarrow_d H_Z^*\left(\tilde{\theta},\theta\right).$$

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

$$\left(\ell_{Z,n_s}^1\left(\tilde{\theta}\right),\ell_{Z,n_s}^2\left(\tilde{\theta},\theta\right),u_{Z,n_s}^1\left(\tilde{\theta},\theta\right),u_{Z,n_s}^2\left(\tilde{\theta}\right)\right)\rightarrow_d\left(\ell_Z^{1*}\left(\tilde{\theta}\right),\ell_Z^{2*}\left(\tilde{\theta},\theta\right),u_Z^{1*}\left(\tilde{\theta},\theta\right),u_Z^{2*}\left(\tilde{\theta}\right)\right).$$

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

**Proof of Lemma 13** Note that we can write

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

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

$$\left\{\ell_m^k,u_m^k\right\}_{k=1}^K\rightarrow\left\{\ell_\infty^k,u_\infty^k\right\}_{k=1}^K$$

with $\ell_\infty^1=-\infty$ and/or $u_\infty^K=\infty$ and the other elements finite,

$$F_{TN}\left(Y(\theta);\mu,\Sigma_Y(\theta),\mathcal{Y}_m^K\right)\rightarrow F_{TN}\left(Y(\theta);\mu,\Sigma_Y(\theta),\mathcal{Y}_\infty^K\right).$$

□

**Proof of Lemma 14** Note that

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

while

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

where

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

Thus,

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

and

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

Using analogous reasoning to that in the proof of Lemma 10, we can write (63) as the solution to

$$g\left(c;\mu,\sqrt{\Sigma_Y(\theta)},\mathcal{Y}^K\right)=0 \tag{64}$$

for

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

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

Note that by construction

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

which implies that

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

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

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

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

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

Note, however, that if we fix all but one element of $\big\{\ell^k,u^k\big\}_{k=1}^K$ and change the remaining element, the set of values for which there exists a solution $c$ to (64) with $c_u \in (\ell^j,u^j)$ and $c_l \in \big(\ell^k,u^k\big)$ for some $j,k$ has Lebesgue measure one by arguments along the same lines as in the proof of Lemma 10. Likewise, the set of values such that there exists a solution $c$ to (64) with $c_l = c_u$ has Lebesgue measure zero as well. The implicit function theorem thus implies that (63) is almost-everywhere continuously differentiable in the element we have selected. Since we can repeat this argument for each element of $\big\{\ell^k,u^k\big\}_{k=1}^K$, we obtain that (63) is continuously differentiable in $\big\{\ell^k,u^k\big\}_{k=1}^K$ Lebesgue almost-everywhere. Moreover, as in the proof of Lemma 10 the form of (63) implies that the same remains true if we take $\ell^1 \to -\infty$ or $u^K \to \infty$.

### D.2.5 Proofs of Uniformity Results

**Proof of Proposition 14** As in the proof of Proposition 9, note that

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

for $CS_{U,-,n} = (-\infty, \hat{\mu}_{\alpha,n}]$. Hence, by Lemma 11, to prove that (59) holds it suffices to show that for all $\{n_s\}$ and $\{P_{n_s}\}$ such that conditions (1) and (2) of the lemma hold with $C_n = 1\left\{\hat{\theta}_n = \tilde{\theta}\right\}$, we have

$$\lim_{s\to\infty} Pr_{P_{n_s}}\left\{\hat{\mu}_{Y,n_s}\left(\hat{\theta}_{n_s}; P_{n_s}\right) \in CS_{U,-,n_s} | \hat{\theta}_{n_s} = \tilde{\theta}\right\} = \alpha. \tag{65}$$

To this end, note that for $F_{TN}\left(Y(\theta); \mu, \Sigma_Y(\theta), \mathcal{Y}^K\right)$ as defined in the statement of Lemma 13, the estimator $\hat{\mu}_{\alpha,n}$ solves

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

for

$$\mathcal{Y}_n = \bigcap_{\theta\in\Theta: A_n\left(\tilde{\theta},\theta\right)>0, D_{Z,n}\left(\tilde{\theta},\theta\right)\geq 0} \left[\ell^1_{Z,n}\left(\tilde{\theta}\right), u^1_{Z,n}\left(\tilde{\theta},\theta\right)\right] \cap \left[\ell^2_{Z,n}\left(\tilde{\theta},\theta\right), u^2_{Z,n}\left(\tilde{\theta}\right)\right] \tag{66}$$

(see Proposition 4 in the main text). The set $\mathcal{Y}_n$ can be written as a finite union of disjoint intervals by DeMorgan's Laws.

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

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

Lemma 12 shows that $\left(Y_n\left(\hat{\theta}_{n_s}\right), \widehat{\Sigma}_{Y,n_s}\left(\hat{\theta}_{n_s}\right), \mathcal{Y}_{n_s}, \hat{\theta}_{n_s}\right)$ converges in distribution as $s \to \infty$,[31] so since $F_{TN}$ is continuous by Lemma 13 while $\text{argmax}_\theta \|X^*(\theta)\|$ is almost everywhere continuous for $X^*$, the continuous mapping theorem implies that

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

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

Since we can write

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

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

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

and by construction

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

and $Pr\left\{\hat{\theta}=\tilde{\theta}\right\}=p^*>0$ by Assumption 5, we thus have that

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

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

which verifies (65).

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 $\theta\neq\tilde{\theta}$,

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

Lemma 6 implies (60). □

**Proof of Corollary 3** Follows from Proposition 14 by the same argument used to prove Corollary 1. □

**Proof of Proposition 15** Note that by the definition of $CS_{U,n}$

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

where $\mathcal{Y}_n$ is as defined in (66) while $(c_l(\mu,\Sigma_Y(\theta),\mathcal{Y}_n),c_u(\mu,\Sigma_Y(\theta),\mathcal{Y}_n))$ are as defined immediately before Lemma 14, after replacing $\mathcal{Y}^K$ with $\mathcal{Y}_n$.

By Lemma 11, to prove that (61) holds it suffices to show that for all $\{n_s\}$ and $\{P_{n_s}\}$ satisfying conditions (1) and (2) of Lemma 11,

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

Thus, it suffices to show that

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

To this end, note that by Lemma 12,

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

and thus, by Lemma 14 and the continuous mapping theorem, that[32]

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

By construction,

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

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

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

Hence,

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

as we wanted to show, so (61) follows by Lemma 5.

Since this result again holds for all $\tilde{\theta}\in\Theta$, (62) follows immediately by the same argument as in the proof of Proposition 14. □

[32]Note that when $\hat{\theta}=\tilde{\theta}$, $\mathcal{Y}^*$ is either equal to the real line, or contains at least one interval with a continuously distributed endpoint. Hence, the almost-everywhere continuity established in Lemma 14 is sufficient for us to apply the continuous mapping theorem.

**Proof of Proposition 16** Follows by the same argument as in the proof of Proposition 11. □

**Proof of Proposition 17** Follows by an argument along the same lines as in the proof of Proposition 12, using Lemmas 11, 12, and 13 in place of 5, 8, and 9, and using the conditioning event $\{Y_n(\hat{\theta}_n) \in \mathcal{Y}_n^H\} = \{Y_n(\hat{\theta}_n) \in \mathcal{Y}_n\} \cap \left\{\mu_{Y,n}\left(\hat{\theta}_n, P_n\right) \in CS_{P,n}^{\beta}\right\}$. □

**Proof of Corollary 4** Follows by the same argument as in the proof of Corollary 2. □

**Proof of Proposition 18** Follows by the same argument as the proof of Proposition 17, using Lemma 14 rather than Lemma 13. □

# 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 sets and the median absolute error of estimators. In this section, we report results for other quantiles, in particular that $\tau$-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 sets $CS_U$ 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 $\tau$-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 sets $CS_U^H$ 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 sets, which ensures that $CS_U^H$ and $CS_{ET}^H$ are contained in $CS_P^{\beta}$. 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 8 and 9 provide the ratios of the $5^{th}$, $25^{th}$, $50^{th}$, $75^{th}$ and $95^{th}$ quantiles of the lengths of $CS_{ET}$, $CS_U$, $CS^H_{ET}$ and $CS^H_U$ relative to the corresponding length quantiles of $CS_P$ for the EWM data-calibrated designs described in Section 6 of the main text. Looking at the upper quantiles in Table 8, we can see that the conditional confidence sets $CS_{ET}$ and $CS_U$ can become very wide when the maximal element of $\mu_X$ is not well-separated from the others. On the other hand, Table 9 shows that the hybrid approach is very successful at mitigating this problem. Indeed, $CS^H_{ET}$ and $CS^H_U$ dominate $CS_P$ across nearly all quantiles and simulation designs considered. Table 10 reports the same quantiles of the studentized absolute errors of $\hat{\mu}_{\frac{1}{2}}$, $\hat{\mu}^H_{\frac{1}{2}}$ and $Y(\hat{\theta})$. Here we can see that, although the hybrid estimator $\hat{\mu}^H_{\frac{1}{2}}$ 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 sets: when minimal bias is desired, $\hat{\mu}^H_{\frac{1}{2}}$ is the preferred estimator.

**Table 8:** 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 |

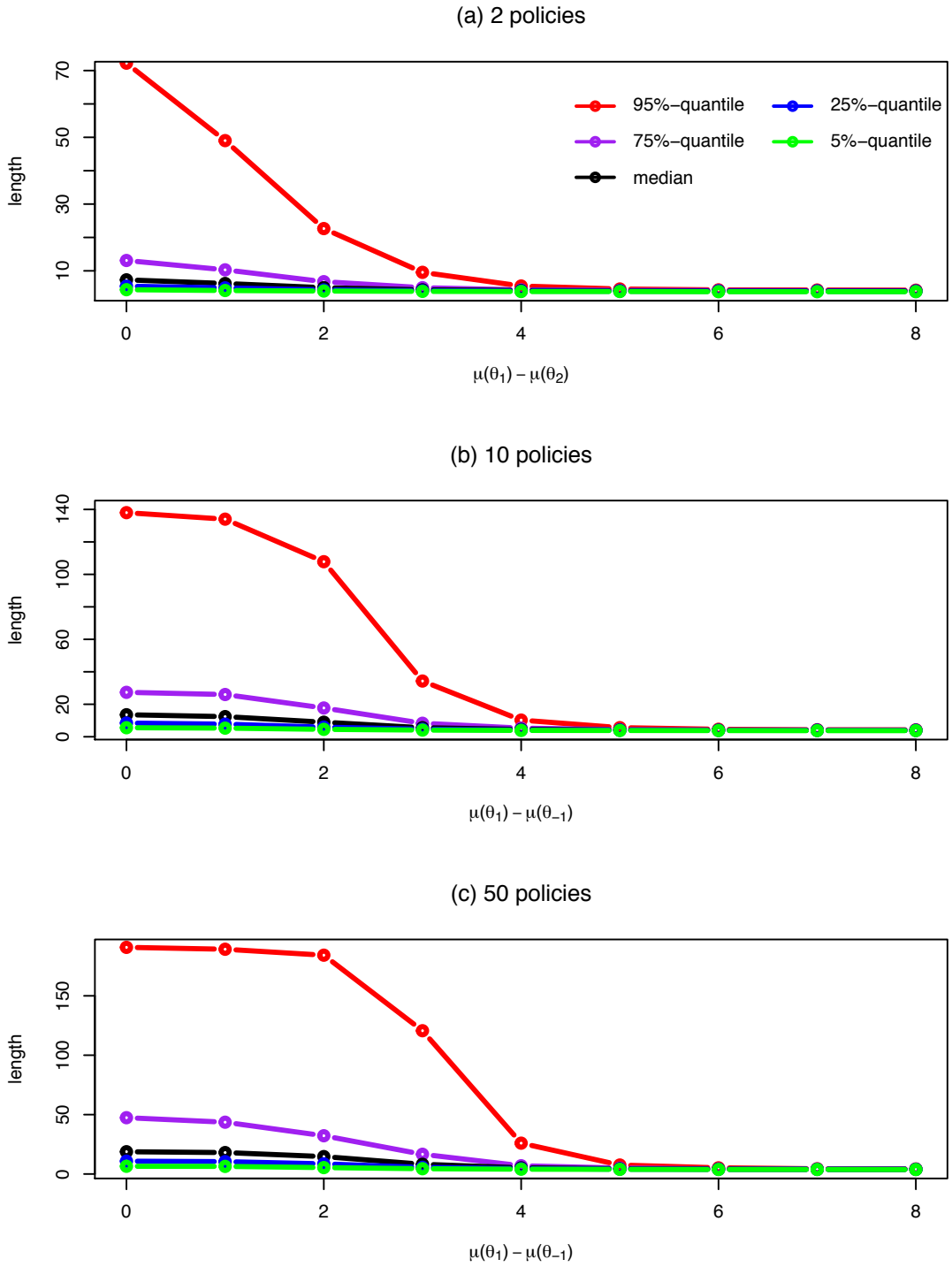


**Figure 6:** Quantiles of the length of 95% conditionally UMAU confidences sets $CS_U$.

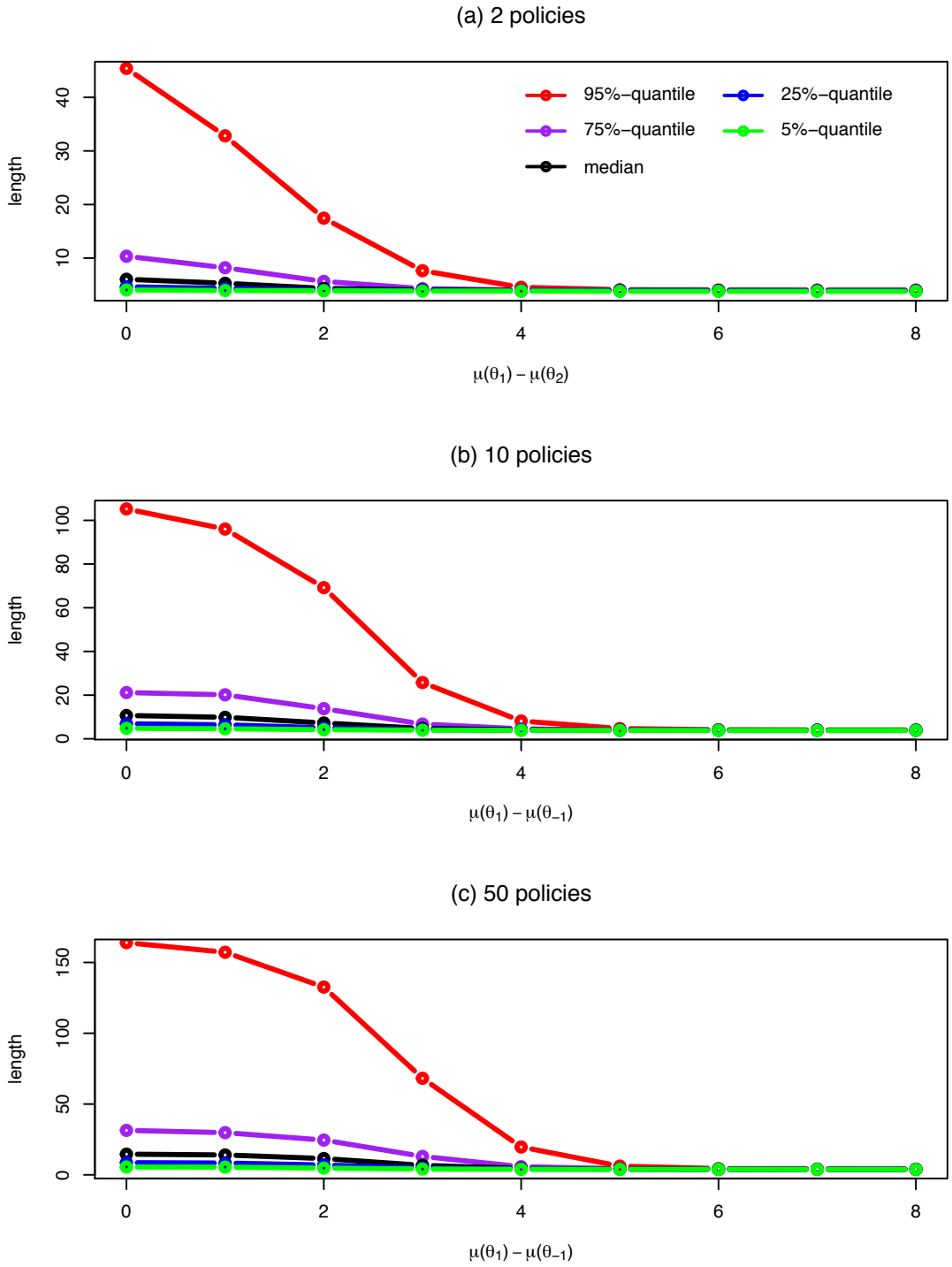


**Figure 7:** Quantiles of the length of 95% conditionally equal-tailed confidences sets $CS_{ET}$.

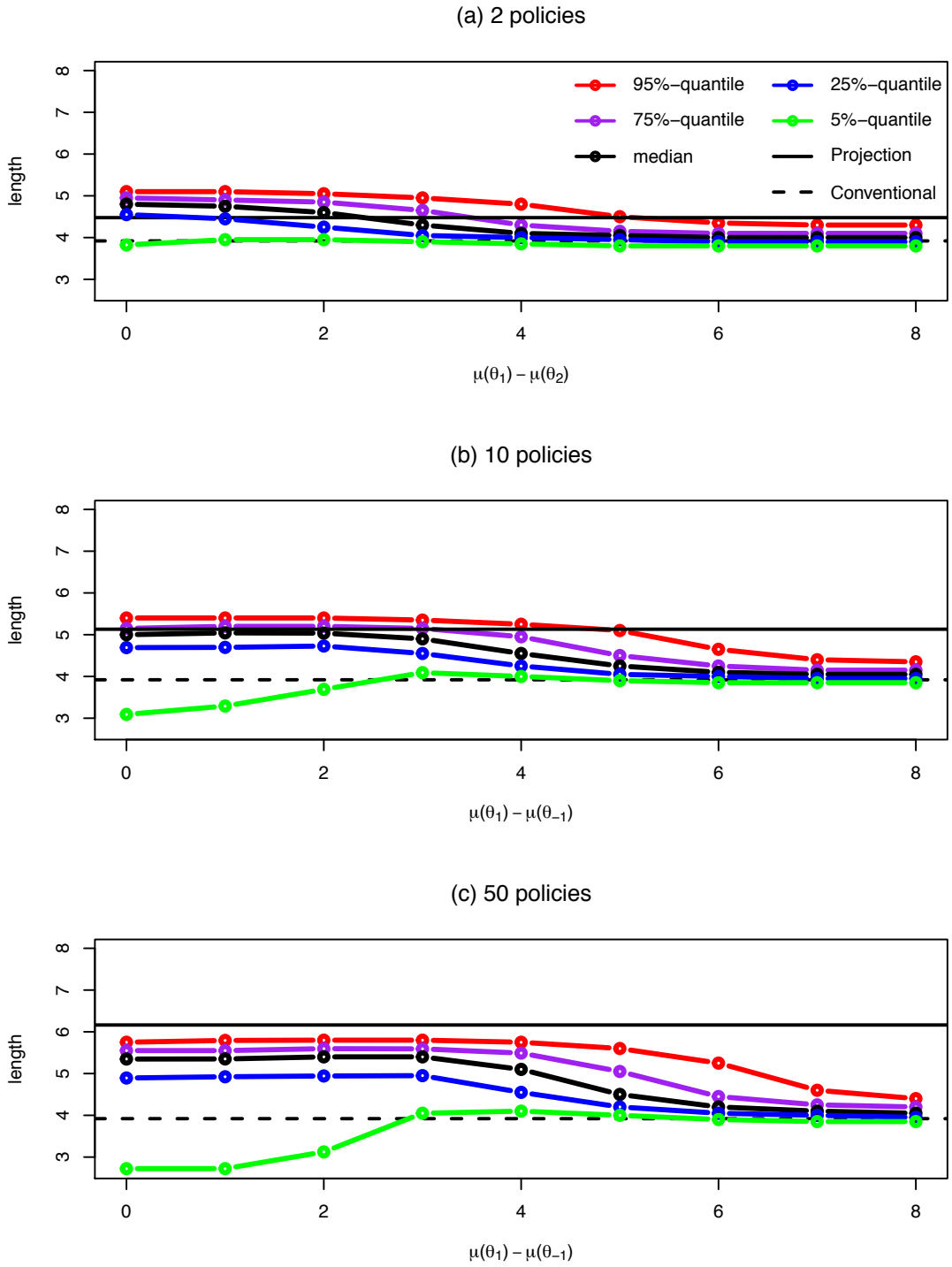


**Figure 8:** Quantiles of the length of 95% hybrid confidence sets $CS_U^H$, with $\beta = 0.005$.

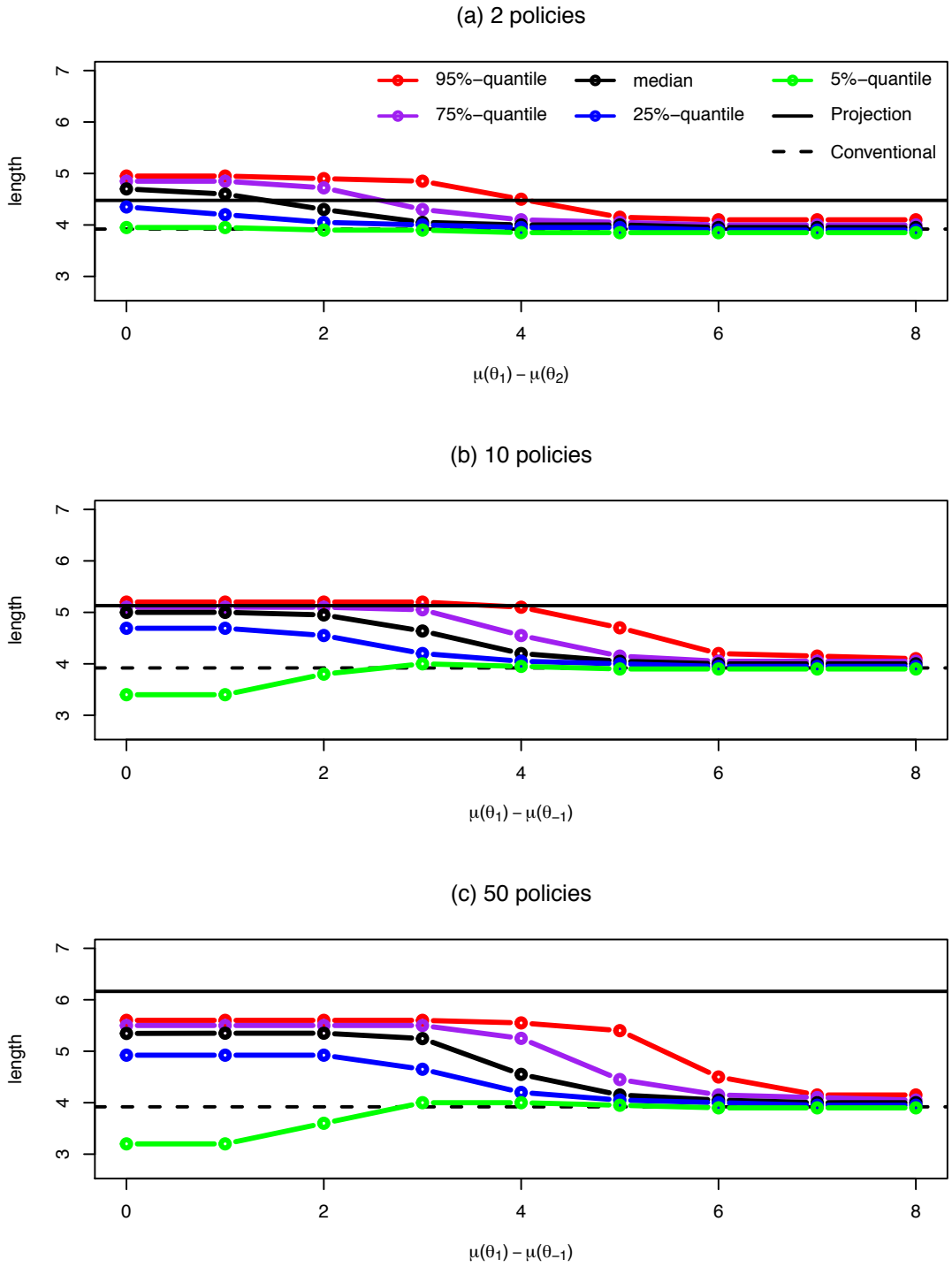


**Figure 9:** Quantiles of the length of 95% hybrid confidence sets $CS^H_{ET}$, with $\beta = 0.005$.

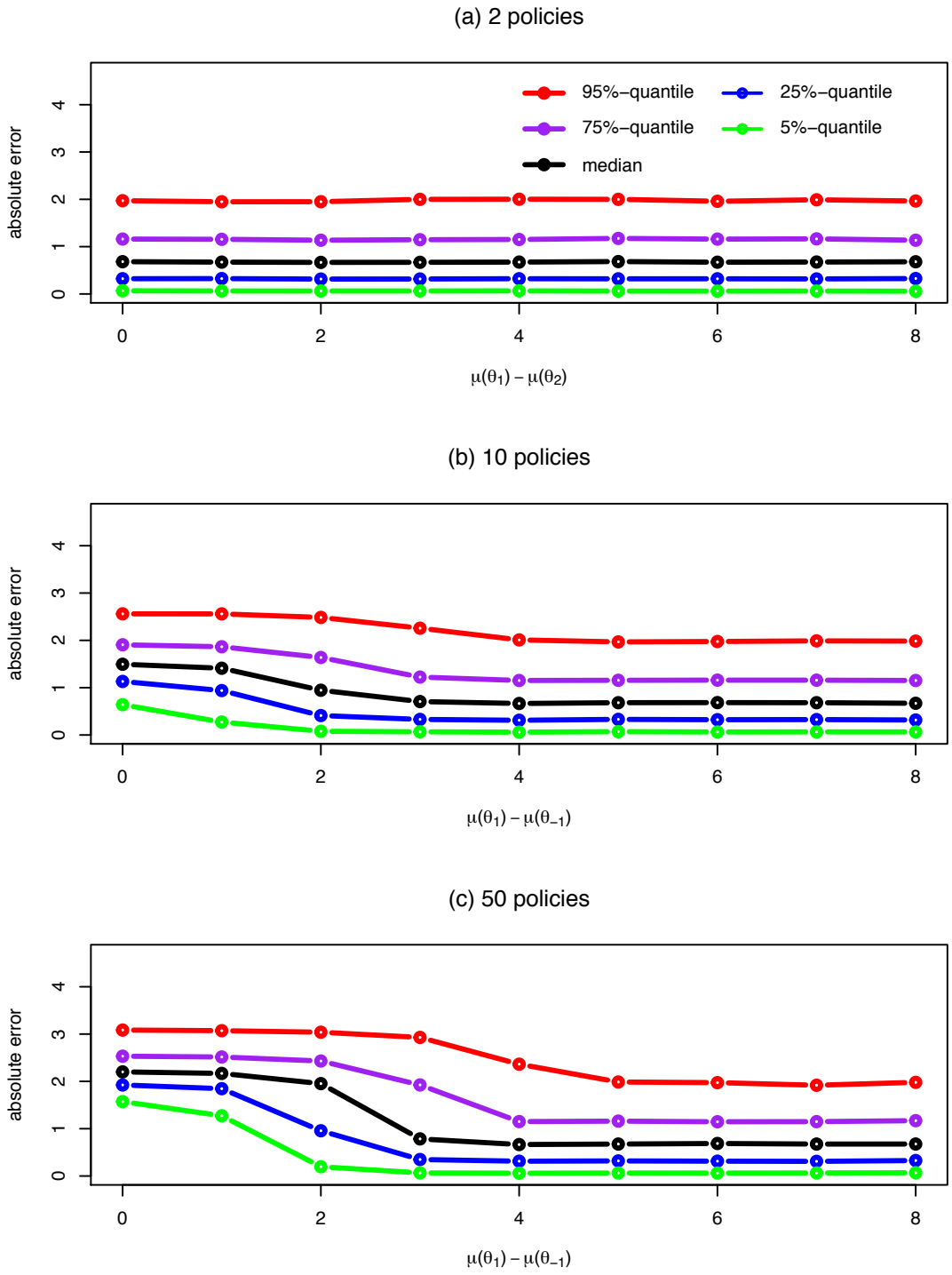


**Figure 10:** Quantiles of the absolute error of the conventional estimator (i.e. of $|X(\hat{\theta}) - \mu(\hat{\theta})|$).

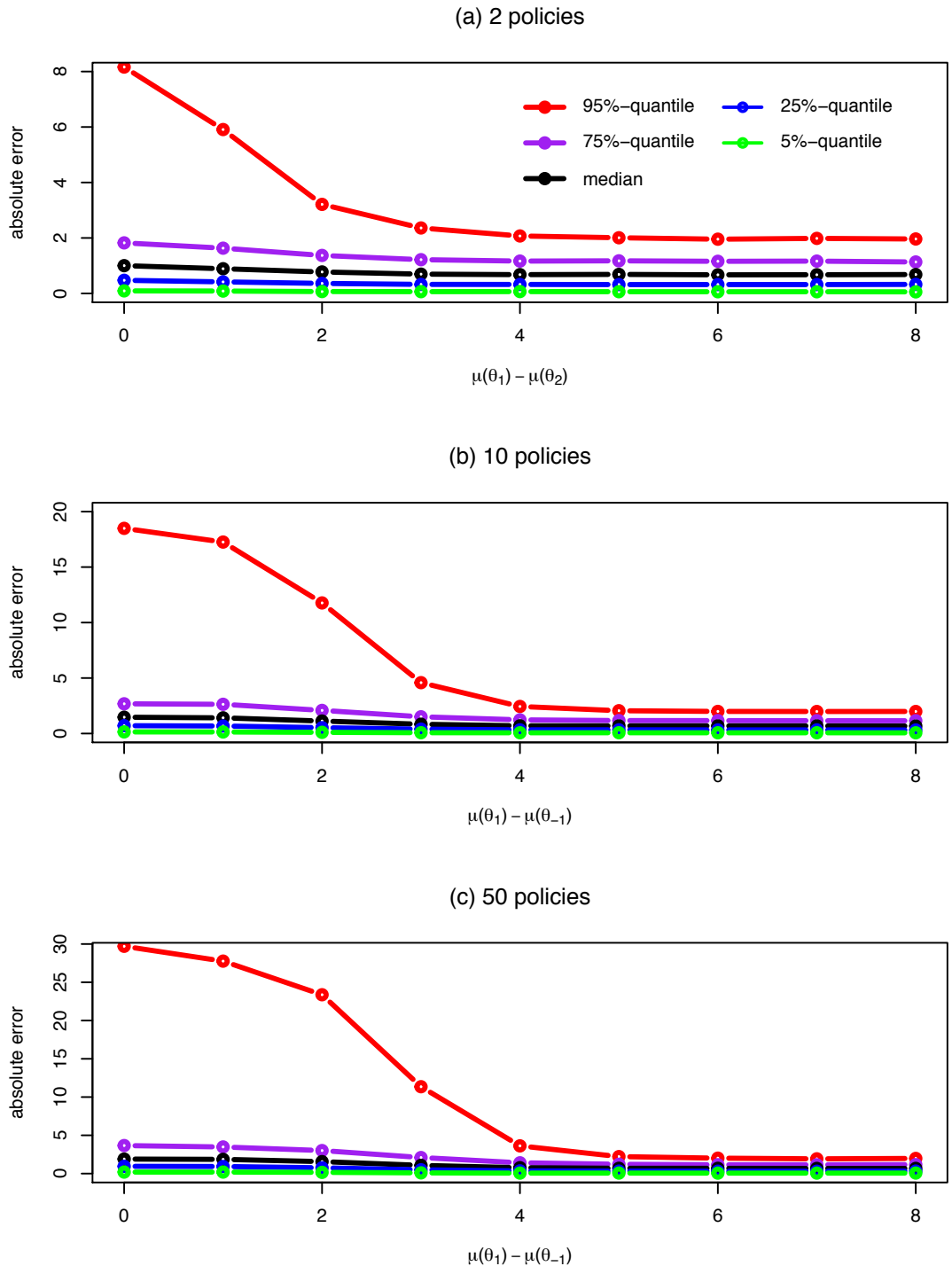


**Figure 11:** Quantiles of the absolute error of the conditionally optimal median unbiased estimator (i.e. of $|\hat{\mu}_{1/2} - \mu(\hat{\theta})|$).

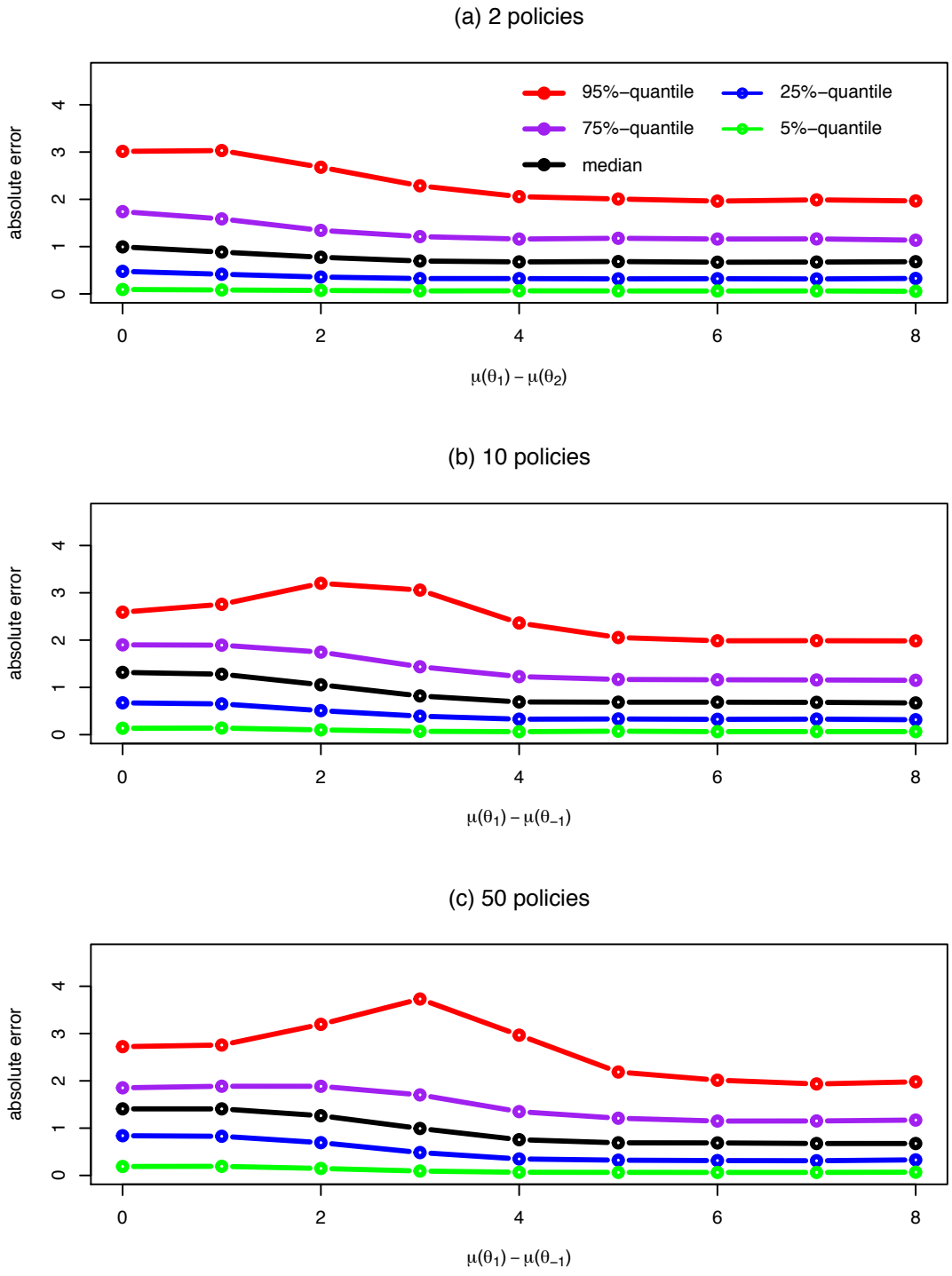


**Figure 12:** Quantiles of the absolute error of the hybrid estimator (i.e. of $|\hat{\mu}^H_{1/2} - \mu(\hat{\theta})|$) with $\beta = 0.005$.

**Table 9:** Ratios of Length Quantiles Relative to $CS_P$

| | $CS^H_{ET}$ Quantile | | | | | $CS^H_U$ 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) | 0.76 | 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 10:** 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 |

# G Additional Results for Tipping Point Simulations

Tables 11 and 12 provide the ratios of the $5^{th}$, $25^{th}$, $50^{th}$, $75^{th}$ and $95^{th}$ quantiles of the lengths of $CS_{ET}$, $CS_U$, $CS^H_{ET}$ and $CS^H_U$ relative to the corresponding length quantiles of $CS_P$ for the tipping point data-calibrated designs described in Section 7 of the main text. The main takeaways from these tables are analogous to those that apply to tables 8 and 9 for the EWM data-calibrated designs. Table 13 reports the same quantiles of the studentized absolute errors of $\hat{\mu}_{\frac{1}{2}}$, $\hat{\mu}^H_{\frac{1}{2}}$ and $Y(\hat{\theta})$. Again, the main features of this table are similar to those of Table 10. However, note that in this application, the hybrid estimator $\hat{\mu}^H_{\frac{1}{2}}$ not only exhibits minimal bias, in contrast to the standard estimator $Y(\hat{\theta})$, but also exhibits lower studentized absolute errors across most quantiles and designs considered.

**Table 11:** Ratios of Length Quantiles Relative to $CS_P$

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

**Table 12:** Ratios of Length Quantiles Relative to $CS_P$

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

**Table 13:** Quantiles of $\left|\hat{\mu}-\mu_Y(\hat{\theta})\right|/\sqrt{\Sigma_Y(\hat{\theta})}$

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

## G.1 Additional Results for Split-Sample Approaches

Table 14 provides the ratios of the $5^{th}$, $25^{th}$, $50^{th}$, $75^{th}$ and $95^{th}$ quantiles of the length of our newly proposed equal-tailed split-sample confidence set $CS^A_{SS}$ relative to the corresponding length quantiles of the conventional split-sample confidence set $CS_{SS}$ for each of the tipping point data-calibrated designs described in Section 7 of the main text. Since every entry in this table is less than one, we can see that the dominance result illustrated in Table 7 of the main text is further reinforced: the length quantiles of $CS^A_{SS}$ are shorter than those of $CS_{SS}$ across all quantiles and simulation designs considered. Table 15 reports the same quantiles of the studentized absolute errors of our newly proposed split-sample estimator $\hat{\mu}^A_{SS,\frac{1}{2}}$ and those of the conventional split-sample estimator $Y^2(\hat{\theta}^1)$. Though both of these estimators are median unbiased for $\mu_Y(\hat{\theta}^1)$, $\hat{\mu}^A_{SS,\frac{1}{2}}$ dominates $Y^2(\hat{\theta}^1)$ in terms of studentized absolute errors across all quantiles and simulation designs considered.

**Table 14:** Ratios of Length Quantiles of $CS^A_{SS}$ Relative to $CS_{SS}$

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

**Table 15:** Quantiles of $\left|\hat{\mu}-\mu_Y(\hat{\theta}^1)\right|/\sqrt{\Sigma_Y(\hat{\theta})^1}$

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

## Supplement References

Andrews, D. W. K., Cheng, X., and Guggenberger, P. (2018). Generic results for establishing the asymptotic size of confidence sets and tests. Unpublished Manuscript.

Schennach, S. M. and Wilhelm, D. (2017). A simple parametric model selection test. *Journal of the American Statistical Association*, 112(520):1663–1674.

Shi, X. (2015). A nondegenerate vuong test. *Quantitative Economics*, 6:85–121.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57(2):307–333.