Online Appendix for

Inference on Winners

Isaiah Andrews Toru Kitagawa Adam McCloskey

June 26, 2023

This online appendix contains proofs and additional results for the paper "Inference on Winners." Appendix A details the assumptions used to impute standard errors for the JOBSTART estimates reported in Cave et al. (1993), and reports additional results for our simulations calibrated to the JOBSTART data. Appendix B shows that our unconditional and conditional coverage requirements arise as necessary conditions for minimax decision rules in two-step decision problems. Appendix C generalizes the conditional inference results discussed in the main text, extending these results to allow additional conditioning variables and unbiased confidence intervals. Appendix D proves our results for the finite-sample normal model. Appendix E provides further details on how our conditional and unconditional inference procedures can be adapted to provide forecast intervals. Appendix F states and proves the uniform asymptotic results referenced in the main text. Appendix G provides additional results and discussion to complement the application in Section 7 of the main text. Finally, Appendix H provides empirical results for an additional empirical example, based on Karlan and List (2007).

Contents

A Details and Additional Results for JOBSTART Application

This appendix provides additional details and results for the JOBSTART application discussed in Section 2 of the main text.

A.1 Standard Error Imputation

This appendix shows how we obtain the standard errors for the average treatment effects reported in Table I of the main text based upon estimates and standard errors in Tables 5.13 and B.4 of Cave et al. (1993).

The results in Table 5.13 of Cave et al. (1993) are based on what we term "the long regression". Letting i denote the individual, $t=1$ denote the third year (months 25–36) and $t=2$ the fourth year (months 37–48) after individual assignment, we suppose that the long regression has the form

$$Y_{it}=\delta_0+\gamma_{s_r}D_i 1_{s_r}\{S_i\}+\sum_{s\in\mathscr{S}\setminus s_r}(\beta_s(1-D_i)+\gamma_s D_i)1_s\{S_i\}+X_i'\delta+\varepsilon_{it}, \tag{A.1}$$

where Y_{it} is the outcome of interest (annual earnings in year t), D_i is an indicator of treatment, $1_s\{S_i\}$ is an indicator for site s, X_i is a vector of time-invariant binary characteristics, and ε_{it} is the unobserved residual for individual i in period t. We base our specification of the long regression upon information in and around Table 5.13 of Cave et al. (1993) since the regression equation is not formally stated. We equate the reference site, which we denote by $s_r\in\mathscr{S}$ for $\mathscr{S}$ the set of sites, with SER/Jobs for Progress in Corpus Christi, TX in accordance with Table B.4 of Cave et al. (1993). An (implicit) maintained assumption in our analysis of this example is that all regression coefficients are time-invariant.

Table 5.13 in Cave et al. (1993) reports estimates for the average *cumulative* treatment effect on outcomes over months 25-48 after the treatment. Using the form of the long regression, we can write these cumulative outcomes as

$$Y_{i1}+Y_{i2}=2\delta_0+2\gamma_{s_r}D_i 1_{s_r}\{S_i\}+\sum_{s\in\mathscr{S}\setminus s_r}(2\beta_s(1-D_i)+2\gamma_s D_i)1_s\{S_i\}+2X_i'\delta+\varepsilon_{i1}+\varepsilon_{i2}. \tag{A.2}$$

Denoting estimates of the parameters of (A.1) in the usual way, Table 5.13 thus reports, for each binary d and $s \in \mathscr{S}$,

$$\hat{\mathrm{E}}[Y_{i1}+Y_{i2}|D_i=d,S_i=s,X_i=\bar{x}_{ds}]=\begin{cases}2\hat{\delta}_0+2\hat{\gamma}_{s_r}d+2\bar{x}'_{ds_r}\hat{\delta}, & \text{for } s=s_r,\\ 2\hat{\delta}_0+2\hat{\beta}_s(1-d)+2\hat{\gamma}_s d+2\bar{x}'_{ds}\hat{\delta}, & \text{for } s\neq s_r,\end{cases} \tag{A.3}$$

where $\bar{x}_{ds}$ is a group average.

To give a causal interpretation to the results from the long regression (A.1), we assume that potential outcomes are also linear, with the potential outcome for person i in period t under treatment status d taking the form

$$Y_{it}(d)=\delta_0+\gamma_{s_r}d1_{s_r}\{S_i\}+\sum_{s\in\mathscr{S}\setminus s_r}(\beta_s(1-d)+\gamma_s d)1_s\{S_i\}+X_i'\delta+\varepsilon_{it}(d),$$

where $E[\varepsilon_{it}(d)]=0$. Note that this model implies that conditional average treatment effects on earnings given (S_i,X_i) depend on S_i but not on X_i, and that the site-specific average treatment effect on cumulative earnings at site $s\in\mathscr{S}$ is then

$$\begin{aligned}\tau_s &\equiv \mathrm{E}[Y_{i1}(1)+Y_{i2}(1)|S_i=s]-\mathrm{E}[Y_{i1}(0)+Y_{i2}(0)|S_i=s]\\ &=\mathrm{E}[Y_{i1}+Y_{i2}|D_i=1,S_i=s,X_i=x]-\mathrm{E}[Y_{i1}+Y_{i2}|D_i=0,S_i=s,X_i=x],\\ &=\begin{cases}2\gamma_{s_r} & \text{for } s=s_r\\ 2\gamma_s-2\beta_s & \text{for } s\neq s_r,\end{cases}\end{aligned}$$

where the second line follows from randomization of D_i and homogeneity of the conditional average treatment effects in X_i. The estimates for τ_s reported inTable 5.13 of Cave et al. (1993) thus correspond to

$$\hat{\tau}_s=\begin{cases}2\hat{\gamma}_{s_r} & \text{for } s=s_r\\ 2\hat{\gamma}_s-2\hat{\beta}_s & \text{for } s\neq s_r,\end{cases} \tag{A.4}$$

which measures the difference between the average outcomes in the treated and control groups at a given site. Cave et al. (1993) report neither confidence intervals nor p-values for these estimates; we must, therefore, impute them.

We base this imputation on the results reported in Table B.4 of Cave et al. (1993).

These results are based on what we call "the short regression",

$$Y_{i2} = \eta_0 + \psi D_i + \sum_{s \in \mathscr{S} \setminus s_r} \lambda_s 1_s\{S_i\} + X_i'\eta + \nu_{i2}, \tag{A.5}$$

where the definition of the regressors is the same as in (A.1). Note that, in contrast with Table 5.13, the dependent variable of the regression reported in Table B.4 is earnings over months 37–48, rather than cumulative earnings over months 25–48. Moreover, since the functional form of the short regression differs from (A.1), ν_{i2} and ε_{i2} differ.

Table B.4 of Cave et al. (1993) reports both estimates and standard errors for the parameters in (A.5). Our goal is to infer standard errors for $\hat{\tau}_s$ based upon the standard error estimates obtained for the linear regression coefficients of (A.5). In order to do so, we rely on several assumptions.

Assumption 1 (Independence of binary characteristics)
X_i is independent of (D_i, S_i).

Although useful for simplifying our analysis, Assumption 1 can be rejected using the data reported in Cave et al. (1993). Specifically, the ratio of men to women varies across sites in Table 2.1 of Cave et al. (1993), and this variation is more than we would expect due solely to sampling variability. To accommodate this and other potential failures of our assumptions, we examine the sensitivity of our empirical results in this application to changes in the standard errors.

Assumption 2 (Uniform random assignment at sites)
$\Pr\{D_i = 1 | S_i = s\} = 1/2$.

Imposing Assumption 2 means we do not have to worry about the relative sizes of the treatment and controls groups at each site, or their influence on the variance of any obtained estimates.[1] Although there are slight differences in the ratio of treated to control individuals across sites, these differences are small.

We now rewrite (A.1) using Assumptions 1 and 2. Let $p_s = \mathrm{E}[1_s\{S_i\}]$ for all $s \in \mathscr{S}$, and rewrite (A.1) for $t = 2$ as

$$Y_{i2} = \delta_0 + \gamma_{sr} D_i + \sum_{s \in \mathscr{S} \setminus s_r} (\gamma_s - \beta_s - \gamma_{s_r}) D_i 1_s\{S_i\} + \sum_{s \in \mathscr{S} \setminus s_r} \beta_s 1_s\{S_i\} + X_i'\delta + \varepsilon_{i2}$$

[1] Differences in the number of individuals at each site are already captured by the standard errors in Table B.4.

$$=\eta_0+\psi D_i+\sum_{s\in\mathscr{S}\setminus s_r}\frac{1}{2}[\gamma_s+\beta_s-\gamma_{s_r}]1_s\{S_i\}+X_i'\delta+\nu_{i2}, \tag{A.6}$$

where

$$\nu_{i2}=\sum_{s\in\mathscr{S}\setminus s_r}(\gamma_s-\beta_s-\gamma_{sr})(D_i-1/2)(1_s\{S_i\}-p_s)+\varepsilon_{i2}. \tag{A.7}$$

Given that ν_{i2} is mean-zero, and is uncorrelated with any of the regressors in (A.6) under Assumptions 1 and 2, the projection coefficients in (A.6) coincide with the coefficients of (A.5). We can hence associate, for all $s\in\mathscr{S}\setminus s_r$,

$$\lambda_s=\frac{1}{2}[\gamma_s+\beta_s-\gamma_{s_r}].$$

Table B.4 in Cave et al. (1993) reports the standard error of $\tilde{\lambda}_s$ for $s\in\mathscr{S}\setminus s_r$, where $\tilde{\lambda}_s$ is the OLS estimator for the site dummy coefficient using observations collected over months 37–48. We then let $\tilde{\gamma}_s$ and $\tilde{\beta}_s$, $s\in\mathscr{S}$, be the OLS estimators that we would obtain via (A.1) using earning observations from $t=2$ only.

Under Assumption 1, $\tilde{\gamma}_s$ and $\tilde{\beta}_s$ are uncorrelated. In addition, (A.7) implies that conditional on the regressors (D_i,S_i,X_i), ε_{i2} in the long regression of $t=2$ is necessarily smaller (stochastically) than ν_{i2} in the short regression, in the sense that $E[\nu_{i2}^2|D_i,S_i,X_i]\geq E[\varepsilon_{i2}^2|D_i,S_i,X_i]$ almost surely. Accordingly, the asymptotic variance of the OLS estimator for $\tilde{\lambda}_s$ obtained from the short regression is larger than the asymptotic variance of the estimator for λ_s based on the coefficient estimates from the long regression, i.e.,

$$\mathrm{Var}(\tilde{\lambda}_s)\geq\mathrm{Var}\left(\frac{1}{2}(\tilde{\gamma}_s+\tilde{\beta}_s-\tilde{\gamma}_{s_r})\right)\geq\frac{1}{4}\left[\mathrm{Var}(\tilde{\gamma}_s)+\mathrm{Var}(\tilde{\beta}_s)+\mathrm{Var}(\tilde{\gamma}_{s_r})\right], \tag{A.8}$$

for each s in large samples.

Next, to relate the variances of $\tilde{\gamma}_s$ and $\tilde{\beta}_s$ to the variances of the OLS coefficients in the long regression for cumulative earnings (i.e., to relate these variances to the coefficients in (A.2)), we impose the following assumption.

Assumption 3 (Homoskedasticity of the earning shocks)
Conditional on regressors (D_i,S_i,X_i), the variance of ε_{i1} is identical to the variance of ε_{i2}, and ε_{i1} is uncorrelated with ε_{i2}.

Under Assumption 3, the variance of the OLS coefficients of (A.1) with $t=2$ is half of the variance of the OLS estimators for the linear regression coefficients of (A.2); given that the regressor matrix is common to both equations, the variance of the OLS estimators is proportional to the variance of the regression residuals. Thus, $\text{Var}(\tilde{\gamma}_s) = \text{Var}(2\hat{\gamma}_s)/2$ and $\text{Var}(\tilde{\beta}_s) = \text{Var}(2\hat{\beta}_s)/2$. Accordingly, inequality (A.8) can be written as

$$\text{Var}(\tilde{\lambda}_s) \geq \frac{1}{8}\Big[\text{Var}(2\hat{\gamma}_s) + \text{Var}(2\hat{\beta}_s) + \text{Var}(2\hat{\gamma}_{s_r})\Big]. \tag{A.9}$$

Note that if Assumption 3 fails and ε_{i1} and ε_{i2} are correlated, this would tend to increase the standard errors for the long regression, and hence the importance of our winner's curse corrections.

Our goal is to obtain the variance of $\hat{\tau}_s$, $s \in \mathscr{S}$, which is

$$\text{Var}(\hat{\tau}_s) = \begin{cases} \text{Var}(2\hat{\gamma}_{s_r}) & \text{for } s = s_r, \\ \text{Var}(2\hat{\gamma}_s) + \text{Var}(2\hat{\beta}_s) & \text{for } s \neq s_r. \end{cases} \tag{A.10}$$

To pin this quantity down, we impose an additional assumption:

Assumption 4 (Variance)
(i.) The inequality in (A.9) holds with equality. (ii.) For all $s \in \mathscr{S} \setminus s_r$ *and* $t = 1, 2$, $Var(Y_{it}(0)|S_i = s) = Var(Y_{it}(1)|S_i = s)$*, which implies that* $Var(2\hat{\beta}_s) = Var(2\hat{\gamma}_s)$*. (iii.)* $Var(Y_{i1}(1) + Y_{i2}(1)|S_i = s_r) = Var(Y_{i1}(1) + Y_{i2}(1)|S_i = s_p)$*, where* s_p *indicates Connelley Skill Learning Center in Pittsburgh, PA.*

We justify our choice of Connelley Skill Learning Center as a reference for SER/Jobs for Progress by noting that the sample mean of $Y_{i1}(1) + Y_{i2}(1)$ in Pittsburgh is closest to that of Corpus Christi, TX.

Focusing on $s = s_p$ under Assumptions 1 and 4, and noting that the sample size of Connelley Skill Learning Center is 2/3 that of SER/Jobs for Progress, we obtain

$$\text{Var}(\tilde{\lambda}_{s_p}) = \frac{1}{8}\left(2\text{Var}(2\hat{\gamma}_{s_p}) + \frac{2}{3}\text{Var}(2\hat{\gamma}_{s_p})\right) = \frac{1}{3}\text{Var}(2\hat{\gamma}_{s_p}),$$

from (A.9). Hence, we can pin down the variance of $\hat{\tau}_{s_p}$ as

$$\text{Var}(\hat{\tau}_{s_p}) = 2\text{Var}(2\hat{\gamma}_{s_p}) = 6\text{Var}(\tilde{\lambda}_{s_p}).$$

Similarly, for $s=s_r$, (A.10) gives

$$\mathrm{Var}(\hat{\tau}_{s_r})=\mathrm{Var}(2\hat{\gamma}_{s_r})=\frac{2}{3}\mathrm{Var}(2\hat{\gamma}_{s_p})=2\mathrm{Var}(\tilde{\lambda}_{s_p}),$$

whilst for the remaining sites, $s\in\mathscr{S}\setminus\{s_p,s_r\}$, we obtain

$$\mathrm{Var}(\hat{\tau}_s)=\mathrm{Var}(\widehat{2\gamma_s})+\mathrm{Var}(\widehat{2\beta_s})=8\mathrm{Var}(\tilde{\lambda}_s)-\mathrm{Var}(2\hat{\gamma}_{s_r})=8\mathrm{Var}(\tilde{\lambda}_s)-2\mathrm{Var}(\tilde{\lambda}_{s_p}),$$

which allows us to impute standard errors for the remaining sites. Note that our imputed standard errors for all sites are consistent with the reported significance levels in Table 5.13 of Cave et al. (1993).

A.2 Conditional Coverage Results

To complement the simulation results reported in the main text, Figure 7 plots the conditional coverage given $\hat{\theta}=\theta^*$ for θ^* the site yielding the largest effect in the JOBSTART data (i.e. CET), where to illustrate coverage distortions we extend the horizonal axis to include negative scaling factors. As expected the conditional interval has correct conditional coverage, while coverage distortions appear for the hybrid and projection intervals for negative scaling factors. In this case $\hat{\theta}=\theta^*$ with low probability, but conditional on this event $X(\hat{\theta})$ tends to be far away from $\mu_X(\theta^*)$, since for $s<0$ site θ^* has the *smallest* ATE. Consequently, projection and hybrid confidence intervals under-cover.

A.3 Split-Sample Results

This appendix reports the results from applying split-sample methods in the simulations calibrated to the JOBSTART data. As in the main text, we report results corresponding to the case where we use half of the data to select a target site, and the other half is used for inference.

As discussed in the main text, sample splitting changes the site selected. Hence, a first important question when considering sample splitting is to what extent it reduces the quality of the treatment selected, relative to using the full data for targeting. Figure 8 provides one answer to this question, plotting the average difference in treatment effects between the best site (CET/San Jose in this simulation design) and the selected site, that is $\mu_X(\theta^*)-E[\mu_X(\hat{\theta})]$ for θ^* corresponding to CET/San Jose. As in the main text, the horizontal axis varies the scaling s on the site-specific average treatment effects. As these results make clear, there is a substantial loss from sample splitting in this context, with

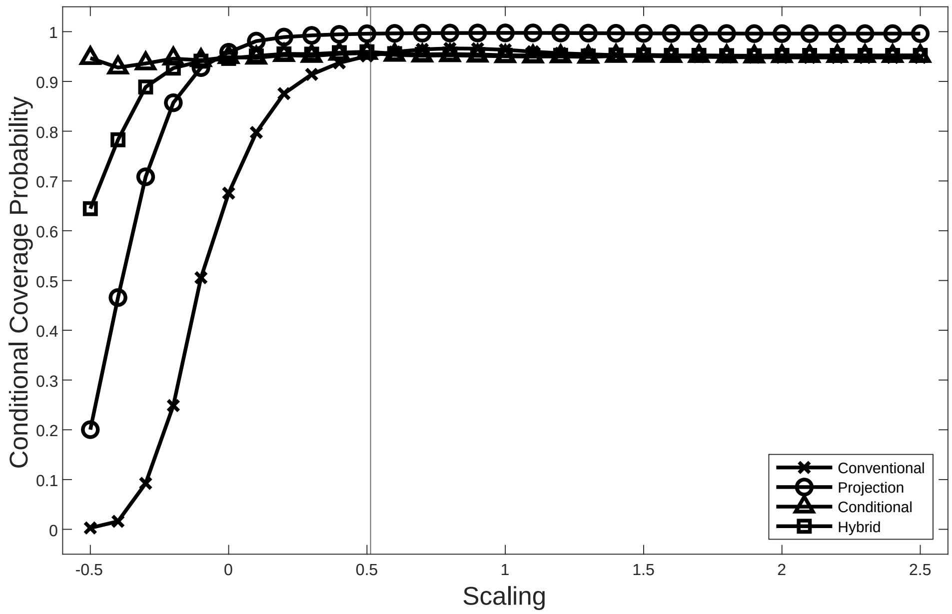

Figure 7: Coverage conditional on $\hat{\theta} = \theta^*$, for θ^* the site with the largest effect in the JOBSTART data (i.e. CET/San Jose), where $X \sim N(s \cdot \hat{\mu}_X, \Sigma)$ for $\hat{\mu}_X$ the JOBSTART point estimates and Σ the diagonal matrix with the squared JOBSTART standard errors on the diagonal. The horizontal axis varies the scaling factor s, and our preferred scaling s^* is marked with a vertical line.

the regret increasing by nearly 40% at our preferred scaling s^*.

B A Decision-Theoretic Model of Inference After Selection

This appendix shows that our unconditional and conditional coverage requirements arise naturally as necessary conditions for minimaxity in a two-step decision-theoretic model.

In particular, consider a decisionmaker who observes

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \Sigma\right),$$

with $\mu = (\mu_X', \mu_Y')'$ unknown and Σ known. They must make a two-part decision, first selecting an element $\theta \in \Theta$ and then reporting an interval I intended to cover $\mu_Y(\theta)$. Suppose that the decisionmaker has lexicographic preferences, prioritizing the selection problem first and the inference problem second.

For the first step selection problem, the decisionmaker must select a decision rule

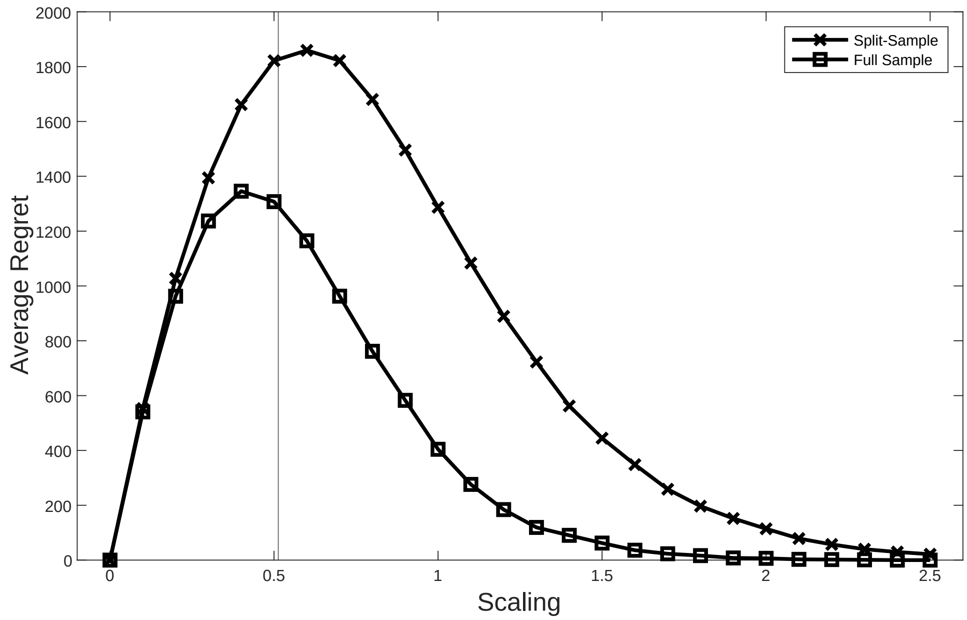

Figure 8: Average regret $\mu_X(\theta^*)-E[\mu_X(\hat{\theta})]$, for selection based on full vs. split sample, where 50% of the data are used for selection.

mapping data realizations to (possibly randomized) selections,

$$\delta_\theta : \mathcal{X}\times\mathcal{Y}\rightarrow\Delta(\Theta)$$

for $\Delta(\Theta)$ the set of probability distributions on Θ. In a slight abuse of notation we use $\delta_\theta(X,Y)\in\Theta$ to denote the realized choice. We assume that the decisionmaker aims to minimize an expected loss that depends on μ,

$$E_\mu[L_\theta(\delta_\theta(X,Y),\mu)].$$

Since μ is unknown the decisionmaker must aggregate across μ values in some way, for example putting a prior on μ and selecting Bayes decision rules or focusing on the worst case and selecting a minimax rule.

As a concrete example of the first-stage preference, consider a decisionmaker who selects δ_θ to minimize the maximum regret based on μ_X

$$\min_{\delta_\theta}\sup_{\mu}\left\{\max_{\theta\in\Theta}\mu_X(\theta)-E_\mu[\mu_X(\delta_\theta(X,Y))]\right\}.$$

Results from Lehmann (1966) and Eaton (1967) imply that picking $\delta_\theta(X,Y) = \text{argmax} X(\theta)$ solves this problem when μ is unconstrained and Σ_X is proportional to the identity matrix. More broadly there is a substantial statistical decision theory literature on optimal selection. For our purposes, we will take the solution to the first-step decision problem as given and focus on the second-stage decision problem.

In the second stage the decisionmaker solves an inference problem. Define the loss

$$L_2(\theta,I,\mu) = f(|I|) + 1\{\mu_Y(\theta) \notin I\} = f(|I|) + \sum_{\tilde{\theta}\in\Theta} 1\left\{\theta = \tilde{\theta}\right\} 1\left\{\mu_Y\left(\tilde{\theta}\right) \notin I\right\}$$

where $|I|$ is the Lebesgue measure of the interval I while $f(\cdot)$ is a non-negative and weakly increasing function with $f(\infty) = \alpha^*$. We consider two versions of expected loss in the second-step problem. The first simply averages the loss across (X,Y) realizations. For second-step decision rule $\delta_I : \mathcal{X} \times \mathcal{Y} \to \Delta(\mathcal{I})$ (for $\mathcal{I}$ the set of intervals on $\mathbb{R}$ and $\Delta(\mathcal{I})$ the set of probability distributions on $\mathcal{I}$) this yields expected loss

$$E_\mu[f(|\delta_I(X,Y)|) + 1\{\mu_Y(\delta_\theta(X,Y)) \notin \delta_I(X,Y)\}]. \tag{A.11}$$

Alternatively, for settings where we are concerned about selection we introduce a selection dummy S where as in the main text we assume that S is independent of (X,Y) conditional on $\delta_\theta(X,Y)$. Denote the conditional distribution of $S|\delta_\theta(X,Y)$ by $F_{S|\delta_\theta}$. If we care about the second-step loss only in those cases where $S = 1$, this yields expected loss

$$E_{\mu,F_{S|\delta_\theta}}[f(|\delta_I(X,Y)|) + 1\{\mu_Y(\delta_\theta(X,Y)) \notin \delta_I(X,Y)\}|S = 1]. \tag{A.12}$$

We next show that minimax decision rules in the second stage imply correct coverage, with criterion (A.11) yielding unconditional coverage and criterion (A.12) yielding conditional coverage. For (A.11), note that always choosing $I = \mathbb{R}$ yields a loss of $f(\infty) = \alpha^*$ regardless of the value of μ. Hence, for a minimax decision rule δ_I^* we must have that the worst-case expected loss is weakly less than α^*,

$$\sup_\mu E_\mu[f(|\delta_\theta(X,Y)|) + 1\{\mu_Y(\delta_\theta(X,Y)) \notin \delta_I^*(X,Y)\}] \leq \alpha^*.$$

Since $f(\cdot)$ is non-negative this implies that

$$\sup_{\mu} Pr_{\mu}\{\mu_Y(\delta_\theta(X,Y)) \notin \delta_I^*(X,Y)\} \leq \alpha^*,$$

which is precisely our unconditional coverage requirement. Hence, minimax second-stage rules under criterion (A.11) always have unconditional coverage at least $1-\alpha^*$.

If we instead consider (A.12), note that the unknown parameters now include both μ and the conditional distribution $F_{S|\delta_\theta(X,Y)}$. By the same argument as in the unconditional case the minimax expected loss must be bounded above by α^*,

$$\sup_{\mu} \sup_{F_{S|\delta_\theta}} E_{\mu,F_{S|\delta_\theta}}[f(|\delta_\theta(X,Y)|)+1\{\mu_Y(\delta_\theta(X,Y)) \notin \delta_I^*(X,Y)\}|S=1] \leq \alpha^*,$$

which implies that

$$\sup_{\mu} \sup_{F_{S|\delta_\theta}} Pr_{\mu,F_{S|\delta_\theta}}\{\mu_Y(\delta_\theta(X,Y)) \notin \delta_I^*(X,Y)|S=1\} \leq \alpha^*.$$

As noted in the main text, however, to ensure that

$$\sup_{F_{S|\delta_\theta}} Pr_{\mu,F_{S|\delta_\theta}}\{\mu_Y(\delta_\theta(X,Y)) \notin \delta_I^*(X,Y)|S=1\} \leq \alpha^*$$

we must have that

$$Pr_{\mu}\{\mu_Y(\delta_\theta(X,Y)) \notin \delta_I^*(X,Y)|\delta_\theta(X,Y)\} \leq \alpha^*,$$

so minimax rules in this setting must ensure conditional coverage at least $1-\alpha^*$.

C Conditional Inference

This section extends the conditional inference results developed in Section 4 of the main text in two directions, first allowing dependence on additional conditioning variables, and then introducing uniformly most accurate unbiased confidence intervals.

C.1 Additional Conditioning Events

Suppose that in addition to conditioning on $\{\hat{\theta}=\tilde{\theta}\}$, we also want to condition on an additional event $\{\hat{\gamma}=\tilde{\gamma}\}$, for $\hat{\gamma}=\gamma(X)$ some function of X. We thus seek estimators that are quantile-unbiased conditional on $(\hat{\theta},\hat{\gamma})$,

$$Pr_\mu\left\{\hat{\mu}_\alpha \geq \mu_Y(\hat{\theta}) | \hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma}\right\}=\alpha \text{ for all } \tilde{\theta}\in\Theta,\ \tilde{\gamma}\in\Gamma, \text{ and all } \mu, \tag{A.13}$$

and confidence sets with correct conditional coverage

$$Pr_\mu\left\{\mu_Y(\hat{\theta})\in CI | \hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma}\right\}\geq 1-\alpha \text{ for all } \tilde{\theta}\in\Theta,\ \tilde{\gamma}\in\Gamma, \text{ and all } \mu. \tag{A.14}$$

One reason we might want to allow such additional conditioning is that we are interested in performance conditional on $S=1$ for an unobserved variable $S\in\{0,1\}$ as discussed in Section 3 of the main text, where we generalize the assumption in the main text and assume that S is conditionally independent of (X,Y) given the pair $(\hat{\theta},\hat{\gamma})$. If we have no other restrictions on the distribution of S, then in order to guarantee conditional coverage given $S=1$,

$$\inf_{\mu} \inf_{F_{S|\hat{\theta},\hat{\gamma}}} Pr_{\mu,F_{S|\hat{\theta},\hat{\gamma}}}\left\{\mu_Y(\hat{\theta})\in CS | S=1\right\}\geq 1-\alpha,$$

it is both necessary and sufficient that we have conditional coverage (A.14).

As in the main text, we re-write the conditioning event in terms of the sample space of X as $\left\{X:\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma}\right\}=\mathcal{X}(\tilde{\theta},\tilde{\gamma})$, and study the conditional distribution of $(X,Y(\tilde{\theta}))$ given $X\in\mathcal{X}(\tilde{\theta},\tilde{\gamma})$. For $Z_{\tilde{\theta}}$ as defined in (10) of the main text, let

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

Conditional on $\hat{\theta}=\tilde{\theta}$, $\hat{\gamma}=\tilde{\gamma}$, and $Z_{\tilde{\theta}}=z$, $Y(\hat{\theta})$ again follows a one-dimensional normal distribution $N(\mu_Y(\tilde{\theta}),\Sigma_Y(\tilde{\theta}))$ truncated to $\mathcal{Y}(\tilde{\theta},\tilde{\gamma},z)$.

To characterize $\mathcal{Y}(\tilde{\theta},\tilde{\gamma},z)$, note that for $\mathcal{X}(\tilde{\theta})$ as derived in the main text, we can write $\mathcal{X}(\tilde{\theta},\tilde{\gamma})=\mathcal{X}(\tilde{\theta})\cap\mathcal{X}_\gamma(\tilde{\gamma})$ for $\mathcal{X}_\gamma(\tilde{\gamma})=\{X\in\mathcal{X}:\gamma(X)=\tilde{\gamma}\}$. Likewise, for $\mathcal{Y}_\gamma(\tilde{\gamma},z)$ defined analogously to (11) in the main text, $\mathcal{Y}(\tilde{\theta},\tilde{\gamma},z)=\mathcal{Y}(\tilde{\theta},z)\cap\mathcal{Y}_\gamma(\tilde{\gamma},z)$. The form of $\mathcal{X}_\gamma(\tilde{\gamma})$ and $\mathcal{Y}_\gamma(\tilde{\gamma},z)$ depends on the conditioning variables $\hat{\gamma}$ considered.

To construct quantile-unbiased estimators, let $F_{TN}(y;\mu_Y(\tilde{\theta}),\tilde{\theta},\tilde{\gamma},z)$ denote the distribution function for a $N(\mu_Y(\tilde{\theta},\Sigma_Y(\tilde{\theta})))$ distribution truncated to $\mathcal{Y}(\tilde{\theta},\tilde{\gamma},z)$. This function is strictly decreasing in $\mu_Y(\tilde{\theta})$, so define $\hat{\mu}_\alpha$ as the unique solution to

$$F_{TN}(Y(\hat{\theta});\hat{\mu}_\alpha,\tilde{\theta},\tilde{\gamma},Z_{\tilde{\theta}})=1-\alpha. \tag{A.15}$$

To establish optimality, we impose one additional assumption.

Assumption 5
If $\Sigma = Cov((X', Y')')$ has full rank, then the parameter space for μ is $\mathbb{R}^{2|\Theta|}$. Otherwise, there exists some μ^* such that the parameter space for μ is $\left\{\mu^* + \Sigma^{\frac{1}{2}} v : v \in \mathbb{R}^{2|\Theta|}\right\}$, where $\Sigma^{\frac{1}{2}}$ is the symmetric square root of Σ.

This assumption requires that the parameter space for μ be sufficiently rich. When Σ is degenerate (for example when $X = Y$, as in Section 2 of the main text), this assumption further implies that (X,Y) have the same support for all values of μ. This rules out cases in which a pair of parameter values μ_1, μ_2 can be perfectly distinguished based on the data. Under this assumption, $\hat{\mu}_\alpha$ is an optimal quantile-unbiased estimator.

Proposition 7
Let $\hat{\mu}_\alpha$ solve (A.15). $\hat{\mu}_\alpha$ is conditionally α-quantile-unbiased in the sense of (A.13). If Assumption 5 holds, then $\hat{\mu}_\alpha$ is the uniformly most concentrated α-quantile-unbiased estimator in that for any other conditionally α-quantile-unbiased estimator $\hat{\mu}_\alpha^*$ and any loss function $L\left(d, \mu_Y(\tilde{\theta})\right)$ that attains its minimum at $d = \mu_Y(\tilde{\theta})$ and is quasiconvex in d for all $\mu_Y(\tilde{\theta})$,

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

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

Proposition 7 shows that $\hat{\mu}_\alpha$ is optimal in the strong sense that it has lower expected loss than any other quantile-unbiased estimator for a large class of loss functions. Hence, $\hat{\mu}_{\frac{1}{2}}$ is an optimal median-unbiased estimator, while $CI_{ET} = [\hat{\mu}_{\frac{\alpha}{2}}, \hat{\mu}_{1-\frac{\alpha}{2}}]$ is an optimal equal-tailed confidence interval.

C.2 Unbiased Confidence Intervals

Rather than considering equal-tailed intervals, we can alternatively consider unbiased confidence intervals. Following Lehmann and Romano (2005), we say that a level $1-\alpha$ two-sided confidence interval CI 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 interval 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 intervals, a level $1-\alpha$ confidence interval CI is unbiased if and only if $\phi(\mu_{Y,0}) = 1\{\mu_{Y,0} \notin CI\}$

is an unbiased test for the corresponding family of hypotheses.[2] 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 α test of the two-sided hypothesis $H_0:\mu_Y(\tilde{\theta})=\mu_{Y,0}$ as

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

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 ζ 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 α test of the one-sided hypothesis $H_0:\mu_Y(\tilde{\theta})\geq\mu_{Y,0}$ as

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

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

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

Proposition 8

If Assumption 5 holds, $\phi_{TS,\alpha}$, $\phi_{OS-,\alpha}$, and $\phi_{OS+,\alpha}$ are uniformly most powerful unbiased size α tests of their respective null hypotheses conditional on $\hat{\theta}=\tilde{\theta}$ and $\hat{\gamma}=\tilde{\gamma}$.

To form uniformly most accurate unbiased confidence intervals we collect the values not rejected by these tests. The two-sided uniformly most accurate unbiased confidence interval is $CI_U=\{\mu_{Y,0}:\phi_{TS,\alpha}(\mu_{Y,0})=0\}$. CI_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 $CI_{U,-}=\{\mu_{Y,0}:\phi_{OS-,\alpha}(\mu_{Y,0})=0\}=(-\infty,\hat{\mu}_{1-\alpha}]$, and $CI_{U,+}=\{\mu_{Y,0}:\phi_{OS+,\alpha}(\mu_{Y,0})=0\}=[\hat{\mu}_{\alpha},\infty)$, respectively. Hence, we can view CI_{ET} as

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

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 CI_U, though Lemma 5.5.1 of Lehmann and Romano (2005) guarantees that this set is an interval.

C.3 Behavior When $Pr_\mu\left\{\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma}\right\}$ is Large

In Proposition 3 of the main text, we showed that our median-unbiased estimators and equal-tailed confidence intervals converge to conventional ones when $Pr_\mu\left\{\hat{\theta}=\tilde{\theta}\right\}\to 1$. The same result holds for general conditioning events and unbiased confidence intervals.

Lemma 2
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 $CI_U\to_p CI_N$, $CI_{ET}\to_p CI_N$, and $\hat{\mu}_{\frac{1}{2}}\to_p Y(\tilde{\theta})$.

Proposition 9
Consider any sequence of values μ_m such that $Pr_{\mu_m}\left\{\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma}\right\}\to 1$. Then under μ_m, we have $CI_U\to_p CI_N$, $CI_{ET}\to_p CI_N$, 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.

D Proofs

We first prove the results stated in Section C, and then build on these to prove the results for the finite-sample normal model discussed in the main text.

D.1 Proofs for Results in Section C

Proof of Proposition 7 For ease of reference, let us abbreviate $(Y(\tilde{\theta}),\mu_Y(\tilde{\theta}),Z_{\tilde{\theta}})$ by $(\tilde{Y},\tilde{\mu}_Y,\tilde{Z})$. Let $Y(-\tilde{\theta})$ collect the elements of Y other than $Y(\tilde{\theta})$ and define $\mu_Y(-\theta)$ analagously. Let

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

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

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

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

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

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

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

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

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

Thus,

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

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

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

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

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

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

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

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

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

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

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

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

$$\frac{1\left\{\left(\tilde{z} + \Sigma_{XY}(\cdot,\tilde{\theta})\Sigma_Y(\tilde{\theta})^{-1}\tilde{y}\right) \in \mathcal{X}(\tilde{\theta},\tilde{\gamma})\right\}}{Pr_{\tilde{\mu}_Z,\tilde{\mu}_Y}\left\{X \in \mathcal{X}(\tilde{\theta},\tilde{\gamma})\right\}} f_{N,\tilde{Z}}(\tilde{z};\tilde{\mu}_Z) f_{N,\tilde{Y}}(\tilde{y};\tilde{\mu}_Y) f_{N,Y^*}(\tilde{y}^*;\mu_Y^*). \tag{A.16}$$

The density (A.16) 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 8 In the proof of Proposition 7, 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 5 implies that the parameter space for (μ_X,μ_Y) is convex and is not contained in any proper linear subspace. Thus, the parameter space for $(\tilde{\mu}_Z,\tilde{\mu}_Y,\mu_Y^*)$ inherits the same property, and satisfies the conditions of Theorem 4.4.1 of Lehmann and Romano (2005). The result follows immediately. □

Proof of Lemma 2 Recall that conditional on $Z_{\tilde{\theta}}=z_{\tilde{\theta}}$, $\hat{\theta}=\tilde{\theta}$ and $\hat{\gamma}=\tilde{\gamma}$ if and only if $Y(\tilde{\theta})\in\mathcal{Y}(\tilde{\theta},\tilde{\gamma},z_{\tilde{\theta}})$. Hence, the assumption of the lemma implies that

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

Note, next, that both the conventional and conditional confidence intervals are equivariant under shifts, in the sense that the conditional confidence interval for $\mu_Y(\tilde{\theta})$ based on observing $Y(\tilde{\theta})$ conditional on $Y(\tilde{\theta})\in\mathcal{Y}(\tilde{\theta},\tilde{\gamma},Z_{\tilde{\theta}})$ is equal to the conditional confidence interval for $\mu_Y(\tilde{\theta})$ based on observing $Y(\tilde{\theta})-\mu_Y^*(\tilde{\theta})$ conditional on $Y(\tilde{\theta})-\mu_Y^*(\tilde{\theta})\in\mathcal{Y}(\tilde{\theta},\tilde{\gamma},Z_{\tilde{\theta}})-\mu_Y^*(\tilde{\theta})$ for any constant $\mu_Y^*(\tilde{\theta})$. Hence, rather than considering a sequence of values $\mu_{Y,m}$, we can fix some μ_Y^* and note that $Pr_{\mu_Y^*}\left\{Y(\tilde{\theta})\in\mathcal{Y}_m^*|Z_{\tilde{\theta}}=z_{\tilde{\theta},m}\right\}\to 1$, where $\mathcal{Y}_m^*=\mathcal{Y}(\tilde{\theta},\tilde{\gamma},Z_{\tilde{\theta}})-\mu_{Y,m}(\tilde{\theta})+\mu_Y^*(\tilde{\theta})$. Confidence intervals for $\mu_{Y,m}(\tilde{\theta})$ in the original problem are equal to those for $\mu_Y^*(\tilde{\theta})$ in the new problem, shifted by $\mu_{Y,m}(\tilde{\theta})-\mu_Y^*(\tilde{\theta})$. Hence, to prove the result it suffices to prove the equivalence of conditional and conventional confidence intervals in the problem with μ_Y fixed (and likewise for estimators).

To prove the result, we make use of the following lemma, which is proved below. First, we must introduce the following notation. Let $(c_{l,ET}(\mu_{Y,0},\mathcal{Y}),c_{u,ET}(\mu_{Y,0},\mathcal{Y}))$ denote the critical values for an equal-tailed test of $H_0:\mu_Y(\tilde{\theta})=\mu_{Y,0}$ for $Y(\tilde{\theta})\sim N\left(\mu_Y(\tilde{\theta}),\Sigma_Y(\tilde{\theta})\right)$ conditional on $Y(\tilde{\theta})\in\mathcal{Y}$. That is, $(c_{l,ET}(\mu_{Y,0},\mathcal{Y}),c_{u,ET}(\mu_{Y,0},\mathcal{Y}))$ solve $F_{TN}(c_{l,ET}(\mu_{Y,0},\mathcal{Y});\mu_{Y,0},\mathcal{Y})=\frac{\alpha}{2}$ and $F_{TN}(c_{u,ET}(\mu_{Y,0},\mathcal{Y});\mu_{Y,0},\mathcal{Y})=1-\frac{\alpha}{2}$, 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 $Pr\{\zeta\in[c_{l,U}(\mu_{Y,0},\mathcal{Y}),c_{u,U}(\mu_{Y,0},\mathcal{Y})]\}=1-\alpha$ and $E[\zeta 1\{\zeta\in[c_{l,U}(\mu_{Y,0},\mathcal{Y}),c_{u,U}(\mu_{Y,0},\mathcal{Y})]\}]=(1-\alpha)E[\zeta]$ for $\zeta\sim\xi|\xi\in\mathcal{Y}$ where $\xi\sim N\left(\mu_{Y,0},\Sigma_Y(\tilde{\theta})\right)$.

Lemma 3
Suppose that we observe $Y(\tilde{\theta})\sim N\left(\mu_Y(\tilde{\theta}),\Sigma_Y(\tilde{\theta})\right)$ conditional on $Y(\tilde{\theta})$ falling in a set $\mathcal{Y}$. If we hold $\left(\Sigma_Y(\tilde{\theta}),\mu_{Y,0}\right)$ fixed and consider a sequence of sets $\mathcal{Y}_m$ such that $Pr\left\{Y(\tilde{\theta})\in\mathcal{Y}_m\right\}\to 1$, we have that for

$$\phi_{ET}(\mu_{Y,0})=1\left\{Y(\tilde{\theta})\notin[c_{l,ET}(\mu_{Y,0},\mathcal{Y}_m),c_{u,ET}(\mu_{Y,0},\mathcal{Y}_m)]\right\} \tag{A.17}$$

and

$$\phi_U(\mu_{Y,0})=1\left\{Y(\tilde{\theta})\notin[c_{l,U}(\mu_{Y,0},\mathcal{Y}_m),c_{u,U}(\mu_{Y,0},\mathcal{Y}_m)]\right\}, \tag{A.18}$$

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

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 CI_{ET} and CI_U are formed by inverting (families of) equal-tailed and unbiased tests, respectively. Let CI_m denote a generic conditional confidence interval formed by inverting a family of tests

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

Hence, we want to show that

$$CI_m\to_p\left[Y(\tilde{\theta})-c_{\frac{\alpha}{2},N}\sqrt{\Sigma_Y(\tilde{\theta})},Y(\tilde{\theta})+c_{\frac{\alpha}{2},N}\sqrt{\Sigma_Y(\tilde{\theta})}\right], \tag{A.19}$$

as $m\to\infty$, for CI_m formed by inverting either (A.17) or (A.18).

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

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

for ϕ_m equal to either (A.17) or (A.18). This convergence in probability holds jointly for all finite collections of values $\mu_{Y,0}$, however, which implies (A.19). 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 9 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 $Pr_{\mu_m}\left\{\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma}\right\}\to 1$ implies $Pr_{\mu_{Y,m}}\left\{\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma}|Z_{\tilde{\theta}}\right\}\to_p 1$. Hence, for $g(\mu_Y,z)=Pr_{\mu_Y}\left\{\hat{\theta}=\tilde{\theta},\hat{\gamma}=\tilde{\gamma}|Z_{\tilde{\theta}}=z\right\}$, 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 $h_\varepsilon(\mu_Y, z) = Pr_{\mu_Y}\{d(CI_U,CI_N)>\varepsilon|Z_{\tilde{\theta}}=z\}$, Lemma 2 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 $\mathcal{H}(\varepsilon)=\{(\mu_Y,z):h_\varepsilon(\mu_Y,z)<\varepsilon\}$, for all $\varepsilon>0$ there exists $\delta(\varepsilon)>0$ such that $\mathcal{G}(\delta(\varepsilon))\subseteq\mathcal{H}(\varepsilon)$.

Hence, since our argument above implies that for all $\delta>0$, $Pr_{\mu_m}\{(\mu_{Y,m},Z_{\tilde{\theta}})\in\mathcal{G}(\delta)\}\to 1$, we see that for all $\varepsilon>0$, $Pr_{\mu_m}\{(\mu_{Y,m},Z_{\tilde{\theta}})\in\mathcal{H}(\varepsilon)\}\to 1$ as well, which suffices to prove the desired claim for confidence intervals. The same argument likewise implies the result for our median unbiased estimator. □

Proof of 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 $F_{TN}(c_{l,ET}(\mathcal{Y}),\mathcal{Y})=\frac{\alpha}{2}$ and $F_{TN}(c_{u,ET}(\mathcal{Y}),\mathcal{Y})=1-\frac{\alpha}{2}$, 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 next that for all y we can write $F_{TN}(y,\mathcal{Y}_m)=a_m(y)+F_N(y)$, where F_N is the standard normal distribution function and $a_m(y)=\frac{\int_{-\infty}^{y}1\{\tilde{y}\in\mathcal{Y}_m\}f_N(\tilde{y})d\tilde{y}}{Pr\{\xi\in\mathcal{Y}_m\}}-F_N(y)$. Recall, however, that $Pr\{\xi\in\mathcal{Y}_m\}\to 1$ and

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

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

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

Unbiased Test We next consider the unbiased test. Recall that critical values $c_{l,U}(\mathcal{Y})$, $c_{u,U}(\mathcal{Y})$ for the unbiased test solve $Pr\left\{\zeta\in[c_{l,U}(\mathcal{Y}),c_{u,U}(\mathcal{Y})]\right\}=1-\alpha$ and $E[\zeta 1\{\zeta\in[c_{l,U}(\mathcal{Y}),c_{u,U}(\mathcal{Y})]\}]=(1-\alpha)E[\zeta]$ for $\zeta\sim\xi|\xi\in\mathcal{Y}$ where $\xi\sim N(0,1)$.

Note that for ζ_m the truncated normal random variable corresponding to $\mathcal{Y}_m$, we can write $Pr\{\zeta_m\in[c_l,c_u]\}=a_m(c_l,c_u)+(F_N(c_u)-F_N(c_l))$ with

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

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

Note, next, that we can write $E[\zeta_m 1\{\zeta_m\in[c_l,c_u]\}]=[\xi 1\{\xi\in[c_l,c_u]\}]+b_m(c_l,c_u)$ for

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

Note, however, that $\int_{c_l}^{c_u}(1\{y\in\mathcal{Y}_m\}-1)yf_N(y)dy\le E[|\xi|1\{\xi\notin\mathcal{Y}_m\}]$. Hence, since

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

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 $Pr\left\{\zeta_m\in[c_l,c_u]\right\}=1-\alpha$ and $E\left[\zeta_m 1\{\zeta_m\in[c_l,c_u]\}\right]=(1-\alpha)E\left[\zeta_m\right]$. From our results above, we can re-write the problem solved by $(c_{l,m},c_{u,m})$ as $F_N(c_u)-F_N(c_l)=1-\alpha-a_m(c_l,c_u)$, $E\left[\xi 1\{\xi\in[c_l,c_u]\}\right]=(1-\alpha)E\left[\zeta_m\right]-b_m(c_l,c_u)$. Letting $\bar{a}_m=\sup_{c_l,c_u}|a_m(c_l,c_u)|$, and $\bar{b}_m=\sup_{c_l,c_u}|b_m(c_l,c_u)|$ we thus see that $(c_{l,m},c_{u,m})$ solves $F_N(c_u)-F_N(c_l)=1-\alpha-a_m^*$ and $E\left[\xi 1\{\xi\in[c_l,c_u]\}\right]=(1-\alpha)E\left[\zeta_m\right]-b_m^*$ 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 $Pr\{\xi \in [c_l, c_u]\} = 1-\alpha+a_m$ and $E[\xi 1\{\xi \in [c_l, c_u]\}] = d_m$ for a_m, $d_m \to 0$. Then $(c_{l,m}, c_{u,m}) \to \left(-c_{\frac{\alpha}{2},N}, c_{\frac{\alpha}{2},N}\right)$.

Using this lemma, since $E[\zeta_m] \to 0$ as $m \to \infty$ we see that for any sequence of values $(a^*_m, b^*_m) \to 0$, $(c_{l,m}(a^*_m, b^*_m), c_{u,m}(a^*_m, b^*_m)) \to \left(-c_{\frac{\alpha}{2},N}, c_{\frac{\alpha}{2},N}\right)$. However, since $\bar{a}_m, \bar{b}_m \to 0$ we know that the values a^*_m and b^*_m corresponding to the true $c_{l,m}$, $c_{u,m}$ must converge to zero. Hence $(c_{l,m}, c_{u,m}) \to \left(-c_{\frac{\alpha}{2},N}, c_{\frac{\alpha}{2},N}\right)$ as we wanted to show. $\square$

Proof of Lemma 4 Note that the critical values solve

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

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 $g(c) - v_m = 0$, 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. □

D.2 Proofs for Results in Main Text

Proof of Proposition 1 Let us assume without loss of generality that $\tilde{\theta} = \theta_1$. Note that the conditioning event $\{\max_{\theta \in \Theta} X(\theta) = X(\theta_1)\}$ is equivalent to $\{MX \geq 0\}$, where

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

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

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

for $c = Cov(W,Y(\tilde{\theta}))/\Sigma_Y(\tilde{\theta})$, noting that the definition of $Z_{\tilde{\theta}}$ in (10) of the main text corresponds to extracting the elements of $\tilde{Z}^*_{\tilde{\theta}}$ corresponding to X. By Lemma 5.1 of Lee et al. (2016),

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

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

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

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

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

Note, however, that

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

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

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

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

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

and

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

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

Proof of Proposition 2 Follows as a special case of Proposition 7. □

Proof of Proposition 3 Follows as a special case of Proposition 9. □

Proof of Proposition 4 We prove the result for coverage, while the result for median unbiasedness is analagous. Provided $\hat{\theta}$ is unique with probability one, we can write

$$Pr_\mu\left\{\mu(\hat{\theta}) \in CI\right\} = \sum_{\tilde{\theta}\in\Theta} Pr_\mu\left\{\hat{\theta} = \tilde{\theta}\right\} Pr_\mu\left\{\mu(\tilde{\theta}) \in CI | \hat{\theta} = \tilde{\theta}\right\}.$$

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

Proof of Lemma 1 The assumption of the lemma implies that $X(\tilde{\theta}) - X(\theta)$ has a non-degenerate normal distribution for all μ. Since Θ is finite, almost-sure uniqueness of $\hat{\theta}$ follows immediately. □

Proof of Proposition 5 We first establish uniqueness of $\hat{\mu}^H_\alpha$. To do so, it suffices to show that $F^H_{TN}(Y(\tilde{\theta});\mu_Y(\tilde{\theta}),\tilde{\theta},Z_{\tilde{\theta}})$ 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 $F^H_{TN}(Y(\tilde{\theta});\mu_Y(\tilde{\theta}),\tilde{\theta},Z_{\tilde{\theta}})$, we further truncate 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 $\mu_Y(\tilde{\theta})$, $F^H_{TN}(Y(\hat{\theta});\mu_Y(\tilde{\theta}),\tilde{\theta},Z_{\tilde{\theta}})$ is a-fortiori decreasing in $\mu_Y(\tilde{\theta})$. Uniqueness of $\hat{\mu}^H_\alpha$ for $\alpha\in(0,1)$ follows. Note, next, that $F^H_{TN}(Y(\tilde{\theta});\mu_Y(\tilde{\theta}),\tilde{\theta},Z_{\tilde{\theta}})\in\{0,1\}$ for $\mu_Y(\tilde{\theta})\notin CI^\beta_P$ from which we immediately see that $\hat{\mu}^H_\alpha\in CI^\beta_P$.

Finally, note that for $\mu_Y(\tilde{\theta})$ the true value, $F^H_{TN}(Y(\hat{\theta});\mu_Y(\tilde{\theta}),\tilde{\theta},Z_{\tilde{\theta}})\sim U[0,1]$ conditional on $\left\{\hat{\theta}=\tilde{\theta},Z_{\hat{\theta}}=z_{\tilde{\theta}},\mu_Y(\tilde{\theta})\in CI^\beta_P\right\}$. Since $F^H_{TN}(Y(\hat{\theta});\mu_Y(\tilde{\theta}),\tilde{\theta},Z_{\tilde{\theta}})$ is decreasing in $\mu_Y(\tilde{\theta})$,

$$Pr_\mu\left\{\hat{\mu}^H_\alpha\geq\mu_Y(\tilde{\theta})|\hat{\theta}=\tilde{\theta},Z_{\hat{\theta}}=z_{\tilde{\theta}},\mu_Y(\tilde{\theta})\in CI^\beta_P\right\}$$

$$=Pr_\mu\left\{F^H_{TN}(Y(\hat{\theta});\mu_Y(\tilde{\theta}),\tilde{\theta},\tilde{\gamma},Z_{\tilde{\theta}})\geq1-\alpha|\hat{\theta}=\tilde{\theta},Z_{\hat{\theta}}=z_{\tilde{\theta}},\mu_Y(\tilde{\theta})\in CI^\beta_P\right\}=\alpha,$$

and thus $\hat{\mu}^H_\alpha$ is α-quantile-unbiased conditional on $\left\{\hat{\theta}=\tilde{\theta},Z_{\hat{\theta}}=z_{\tilde{\theta}},\mu_Y(\tilde{\theta})\in CI^\beta_P\right\}$. We can drop the conditioning on $Z_{\tilde{\theta}}$ by the law of iterated expectations, and α-quantile unbiasedness conditional on $\mu_Y(\tilde{\theta})\in CI^\beta_P$ follows by the same argument as in the proof of Proposition 4.

Proof of Proposition 6 The first part of the proposition follows immediately from Proposition 5. For the second part of the proposition, note that

$$Pr_\mu\left\{\mu_Y(\hat{\theta})\in CI^H_{ET}\right\}=Pr_\mu\left\{\mu_Y(\hat{\theta})\in CI^\beta_P\right\}\times$$

$$\sum_{\tilde{\theta}\in\Theta}Pr_\mu\left\{\hat{\theta}=\tilde{\theta}|\mu_Y(\hat{\theta})\in CI^\beta_P\right\}Pr_\mu\left\{\mu_Y(\tilde{\theta})\in CI^H_{ET}|\hat{\theta}=\tilde{\theta},\mu_Y(\tilde{\theta})\in CI^\beta_P\right\}$$

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

where the second equality follows from the first part of the proposition. The upper bound follows by the same argument and the fact that $Pr_\mu\left\{\mu_Y(\hat{\theta})\in CI^\beta_P\right\}\leq1$. □

E Forecast Intervals

This appendix provides additional details on the forecast intervals discussed in the main text. Let $Z_{1-2,\tilde{\theta}}=X-\frac{\Sigma_{XY}(\cdot,\tilde{\theta})}{\Sigma_{Y_{1-2}}(\tilde{\theta})}Y_{1-2}(\tilde{\theta})$ denote the analog of $Z_{\tilde{\theta}}$ in the main text which uses Y_{1-2} in place of Y. The same argument as in Section 4 of the main text implies that conditional on $\hat{\theta}=\tilde{\theta}$ and $Z_{1-2,\tilde{\theta}}=z$, for $\Sigma_{Y_{1-2}}=\Sigma_Y+\Sigma_{Y_2}$, $Y_{1-2}(\tilde{\theta})$ is distributed as a $N(0,\Sigma_{Y_{1-2}}(\tilde{\theta}))$ variable truncated to $\mathcal{Y}_{1-2}(\tilde{\theta},z)=[\mathcal{L}_{1-2}(\tilde{\theta},z),\mathcal{U}_{1-2}(\tilde{\theta},z)]$, where

$$\mathcal{L}_{1-2}(\tilde{\theta},z)=\max_{\theta\in\Theta:\Sigma_{XY}(\tilde{\theta})>\Sigma_{XY}(\tilde{\theta},\theta)}\frac{\Sigma_{Y_{1-2}}(\tilde{\theta})\left(z(\theta)-z(\tilde{\theta})\right)}{\Sigma_{XY}(\tilde{\theta})-\Sigma_{XY}(\tilde{\theta},\theta)},$$

$$\mathcal{U}_{1-2}(\tilde{\theta},z)=\min_{\theta\in\Theta:\Sigma_{XY}(\tilde{\theta})<\Sigma_{XY}(\tilde{\theta},\theta)}\frac{\Sigma_{Y_{1-2}}(\tilde{\theta})\left(z(\theta)-z(\tilde{\theta})\right)}{\Sigma_{XY}(\tilde{\theta})-\Sigma_{XY}(\tilde{\theta},\theta)}.$$

The interval CI_{ET}^{1-2} fails to cover zero only when $Y_{1-2}(\hat{\theta})$ lies in the tails of this conditional distribution. Letting $q_\alpha(\tilde{\theta},z)$ denote the α quantile of the truncated normal distribution, the resulting *conditional* forecast interval for $Y_2(\hat{\theta})$ is thus

$$FI=\{y_2:Y(\hat{\theta})-q_{1-\alpha/2}(\hat{\theta},Z_{1-2,\hat{\theta}}(y_2))\leq y_2\leq Y(\hat{\theta})-q_{\alpha/2}(\hat{\theta},Z_{1-2,\hat{\theta}}(y_2))\},$$

where with a slight abuse of notation $Z_{1-2,\tilde{\theta}}(y_2)=X-\frac{\Sigma_{XY}(\cdot,\tilde{\theta})}{\Sigma_{Y_{1-2}}(\tilde{\theta})}(Y(\tilde{\theta})-y_2)$. To see why this forecast interval has correct coverage conditional on any realization of $\hat{\theta}$, note that

$$Pr_\mu\left\{Y_2(\hat{\theta})\in FI|\hat{\theta}=\tilde{\theta},Z_{1-2,\hat{\theta}}=z\right\}$$
$$=Pr_\mu\left\{Y(\hat{\theta})-q_{1-\alpha/2}(\hat{\theta},Z_{1-2,\hat{\theta}}(Y_2(\hat{\theta})))\leq Y_2(\hat{\theta})\leq Y(\hat{\theta})-q_{\alpha/2}(\hat{\theta},Z_{1-2,\hat{\theta}}(Y_2(\hat{\theta})))|\hat{\theta}=\tilde{\theta},Z_{1-2,\hat{\theta}}=z\right\}$$
$$=Pr_\mu\left\{q_{\alpha/2}(\tilde{\theta},z)\leq Y_{1-2}(\tilde{\theta})\leq q_{1-\alpha/2}(\tilde{\theta},z)|\hat{\theta}=\tilde{\theta},Z_{1-2,\hat{\theta}}=z\right\}=1-\alpha$$

since $Z_{1-2,\tilde{\theta}}(Y_2(\tilde{\theta}))=Z_{1-2,\tilde{\theta}}$. We can construct conditional upper and lower one-sided forecast intervals analogously.

We can similarly construct *unconditional* forecast intervals for $Y_2(\hat{\theta})$ based on our hybrid approach. Let $q_\alpha^H(\tilde{\theta},\beta,z)$ denote the α-quantile of $Y_{1-2}(\tilde{\theta})$ truncated to $\mathcal{Y}^H(\tilde{\theta},\beta,z)=[\mathcal{L}_{1-2}^H(\tilde{\theta},\beta,z),\mathcal{U}_{1-2}^H(\tilde{\theta},\beta,z)]$, where

$$\mathcal{L}_{1-2}^H(\tilde{\theta},\beta,z)=\max\left\{-c_\beta^{1-2}\sqrt{\Sigma_{Y_{1-2}}(\tilde{\theta})},\mathcal{L}_{1-2}(\tilde{\theta},z)\right\}$$

$$\mathcal{U}_{1-2}^H(\tilde{\theta},\beta,z)=\min\left\{c_\beta^{1-2}\sqrt{\Sigma_{Y_{1-2}}(\tilde{\theta})},\mathcal{U}_{1-2}(\tilde{\theta},z)\right\}$$

for c_β^{1-2} the $1-\beta$ quantile of $\max_\theta|Y_{1-2}(\theta)|/\sqrt{\Sigma_{Y_{1-2}}(\theta)}$ (i.e. the projection critical value).

The unconditional forecast interval is then

$$FI^H = \left\{ y_2 : Y(\hat\theta) - q^H_{1-\frac{\alpha-\beta}{2(1-\beta)}}(\hat\theta,\beta,Z_{1-2,\hat\theta}(y_2)) \le y_2 \le Y(\hat\theta) - q^H_{\frac{\alpha-\beta}{2(1-\beta)}}(\hat\theta,\beta,Z_{1-2,\hat\theta}(y_2)) \right\}.$$

To see why this forecast interval has correct unconditional coverage, note that

$$\begin{aligned}
& Pr_\mu\left\{ Y_2(\hat\theta) \in FI^H \right\} \\
&= Pr_\mu\left\{ Y_2(\hat\theta) \in FI^H \middle| -c_\beta^{1-2}\sqrt{\Sigma_{Y_{1-2}}(\hat\theta)} \le Y_{1-2}(\hat\theta) \le c_\beta^{1-2}\sqrt{\Sigma_{Y_{1-2}}(\hat\theta)} \right\} \\
&\quad \times Pr_\mu\left\{ -c_\beta^{1-2}\sqrt{\Sigma_{Y_{1-2}}(\hat\theta)} \le Y_{1-2}(\hat\theta) \le c_\beta^{1-2}\sqrt{\Sigma_{Y_{1-2}}(\hat\theta)} \right\} \\
&= Pr_\mu\Big\{ Y(\hat\theta) - q^H_{1-\frac{\alpha-\beta}{2(1-\beta)}}(\hat\theta,\beta,Z_{1-2,\hat\theta}(Y_2(\hat\theta))) \le Y_2(\hat\theta) \le Y(\hat\theta) - q^H_{\frac{\alpha-\beta}{2(1-\beta)}}(\hat\theta,\beta,Z_{1-2,\hat\theta}(Y_2(\hat\theta))) \\
&\qquad \Big| -c_\beta^{1-2}\sqrt{\Sigma_{Y_{1-2}}(\hat\theta)} \le Y_{1-2}(\hat\theta) \le c_\beta^{1-2}\sqrt{\Sigma_{Y_{1-2}}(\hat\theta)} \Big\} \\
&\quad \times Pr_\mu\left\{ -c_\beta^{1-2}\sqrt{\Sigma_{Y_{1-2}}(\hat\theta)} \le Y_{1-2}(\hat\theta) \le c_\beta^{1-2}\sqrt{\Sigma_{Y_{1-2}}(\hat\theta)} \right\} \\
&\ge Pr_\mu\Big\{ q^H_{\frac{\alpha-\beta}{2(1-\beta)}}(\hat\theta,\beta,Z_{1-2,\hat\theta}) \le Y_{1-2}(\hat\theta) \le q^H_{1-\frac{\alpha-\beta}{2(1-\beta)}}(\hat\theta,\beta,Z_{1-2,\hat\theta}) \\
&\qquad \Big| -c_\beta^{1-2}\sqrt{\Sigma_{Y_{1-2}}(\hat\theta)} \le Y_{1-2}(\hat\theta) \le c_\beta^{1-2}\sqrt{\Sigma_{Y_{1-2}}(\hat\theta)} \Big\} \times (1-\beta) \\
&= \left(1 - \frac{\alpha-\beta}{1-\beta}\right)(1-\beta) = 1-\alpha,
\end{aligned}$$

where the inequality follows from the fact that

$$Pr_\mu\left\{ -c_\beta^{1-2}\sqrt{\Sigma_{Y_{1-2}}(\hat\theta)} \le Y_{1-2}(\hat\theta) \le c_\beta^{1-2}\sqrt{\Sigma_{Y_{1-2}}(\hat\theta)} \right\} \ge 1-\beta.$$

F Uniform Asymptotic Validity

This section establishes uniform asymptotic validity for plug-in versions of the procedures discussed in the main text. 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, we do not pursue such an extension here.

Feasible finite-sample estimators and confidence intervals are denoted as their counterparts in Sections 4–5 of the main text, with the addition of an n subscript. We suppose that the sample of size n is drawn from some (unknown) distribution $P \in \mathcal{P}_n$. To simplify repetitive notation we work with scaled estimates (X_n, Y_n) which correspond to $\sqrt{n} \cdot (\tilde{X}_n, \tilde{Y}_n)$ for $(\tilde{X}_n, \tilde{Y}_n)$ as discussed in Section 6 of the main text. Similarly, we work with a variance estimator $\widehat{\Sigma}_n = n \cdot \widetilde{\Sigma}_n$ for $\widetilde{\Sigma}_n$ as discussed in the main text.

We first impose that (X_n, Y_n) are uniformly asymptotically normal under $P \in \mathcal{P}_n$, where the centering vectors $(\mu_{X,n}, \mu_{Y,n})$ and the limiting variance Σ may depend on P.

Assumption 6

For the class of Lipschitz functions that are bounded in absolute value by one and have Lipschitz constant bounded by one, BL_1, there exist sequences of functions $\mu_{X,n}(P)$ and $\mu_{Y,n}(P)$ and a function $\Sigma(P)$ such that for $\xi_P \sim N(0, \Sigma(P))$,

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

Uniform convergence in bounded Lipschitz metric is one formalization for uniform convergence in distribution. When X_n and Y_n are scaled sample averages based on independent data, as in Section 2 of the main text, Assumption 6 will follow from moment bounds, while for dependent data it will follow from moment and dependence bounds.

We next assume that the asymptotic variance is uniformly consistently estimable.

Assumption 7

The estimator $\widehat{\Sigma}_n$ is uniformly consistent in the sense that for all $\varepsilon > 0$

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

Provided we use a variance estimator appropriate to the setting (e.g. the sample variance for iid data, long-run variance estimators for time series, and so on) Assumption 7 will follow from the same sorts of sufficient conditions as for Assumption 6.

Finally, we restrict the asymptotic variance.

Assumption 8

There exists a finite $\bar{\lambda} > 0$ such that

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

$$1/\bar{\lambda} \le \sqrt{\Sigma_X(\theta;P)\Sigma_X(\tilde{\theta};P)} - \Sigma_X(\theta,\tilde{\theta};P) \text{ for all } \theta, \tilde{\theta} \in \Theta \text{ with } \theta \ne \tilde{\theta} \text{ and all } P \in \mathcal{P}_n.$$

The upper bounds on $\Sigma_X(\theta;P)$ and $\Sigma_Y(\theta;P)$ ensure that the random variables ξ_P in Assumption 6 are stochastically bounded, while the lower bounds ensure that each entry (X_n,Y_n) has a nonzero asymptotic variance. The assumption of nonzero variance rules out the case where one element of X_n is a non-random threshold (as discussed in Section 3 of main text), but our asymptotic results can be extended to cover this case at the cost of additional notation. The second condition ensures that no two elements of X_n are perfectly (positively) correlated asymptotically, and hence, by Lemma 1, guarantees that $\hat{\theta}_n$ is unique with probability tending to one. Note that this condition is weaker than a standard assumption bounding the eigenvalues of $\Sigma_X(P)$ away from zero.

High-Dimensional Settings Our asymptotic analysis considers settings where $|\Theta|$, and hence the dimension of X_n and Y_n, are fixed as $n \to \infty$. One might also be interested in settings where $|\Theta|$ grows with n, but this will raise complications for both the normal approximation and estimation of the asymptotic variance. Such an extension is interesting, but beyond the scope of this paper.

Variance Estimation Practically, even for fixed $|\Theta|$ one might still worry about the difficulty of estimating Σ in finite samples, since this matrix has $|\Theta|(|\Theta|+1)/2$ entries. Fortunately, in many cases Σ has additional structure which renders variance estimation more tractable than in the fully general case. Suppose, for instance, that we want to conduct inference on the best-performing treatment from a randomized trial, as in Section 2 of the main text and Section H below. In this case, provided trial participants are drawn independently, elements of $X_n(\theta)$ corresponding to distinct treatments are uncorrelated and Σ is diagonal. In other cases, such as Section 7 of the main text, $|\Theta|$ may be large, but the elements of X_n are formed by taking combinations of a much lower-dimensional set of random variables. In this case, Σ_X can be written as a known linear transformation of a much lower-dimensional variance matrix.

F.1 Uniform Asymptotic Validity

In the finite-sample normal model, we study both conditional and unconditional properties of our methods. We would like to do the same in our asymptotic analysis, but may have $Pr\left\{\hat{\theta}_n = \tilde{\theta}\right\} \to 0$ for some $\tilde{\theta}$, in which case conditioning on $\hat{\theta}_n = \tilde{\theta}$ is problematic. To address this, we multiply conditional statements by the probability of the conditioning event.

Asymptotic uniformity results for conditional inference procedures were established by Tibshirani et al. (2018) and Andrews, Kitagawa, and McCloskey (2021) for settings where the target parameter is chosen in other ways. Their results, however, limit attention to classes of data generating processes with asymptotically bounded means $(\mu_{X,n},\mu_{Y,n})$. This rules out e.g. the conventional pointwise asymptotic case that fixes P and takes $n\to\infty$. We do not require such boundedness. Moreover, the results of Tibshirani et al. (2018) do not cover quantile-unbiased estimation, and also do not cover hybrid procedures, which are new to the literature.[3]

Our proofs are based on subsequencing arguments as in D. Andrews, Cheng, and Guggenberger (2020), 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. We first establish the asymptotic validity of our quantile-unbiased estimators.

Proposition 10

Under Assumptions 6-8, for $\hat{\mu}_{\alpha,n}$ the α-quantile unbiased estimator,

$$\lim_{n\to\infty}\sup_{P\in\mathcal{P}_n}\left|Pr_P\left\{\hat{\mu}_{\alpha,n}\geq\mu_{Y,n}\left(\hat{\theta}_n;P\right)|\hat{\theta}_n=\tilde{\theta}\right\}-\alpha\right|Pr_P\left\{\hat{\theta}_n=\tilde{\theta}\right\}=0, \tag{A.20}$$

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

$$\lim_{n\to\infty}\sup_{P\in\mathcal{P}_n}\left|Pr_P\left\{\hat{\mu}_{\alpha,n}\geq\mu_{Y,n}\left(\hat{\theta}_n;P\right)\right\}-\alpha\right|=0. \tag{A.21}$$

This immediately implies asymptotic validity of equal-tailed confidence intervals.

Corollary 1

Under Assumptions 6-8, for $CI_{ET,n}$ the level $1-\alpha$ equal-tailed confidence interval

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

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

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

We can likewise establish uniform asymptotic validity of projection confidence intervals.

Proposition 11

[3] In a follow-up paper, Andrews, Kitagawa, and McCloskey (2021), we apply the conditional and hybrid approaches developed here to settings where $\hat{\theta}=\operatorname{argmax}\|X(\theta)\|$.

Under Assumptions 6-8, for $CI_{P,n}$ the level $1-\alpha$ projection confidence interval,

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}_n} Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CI_{P,n}\right\}\geq 1-\alpha. \tag{A.22}$$

To state results for hybrid estimators and confidence intervals, let $C_n^H\left(\tilde{\theta};P\right) = 1\left\{\hat{\theta}_n=\tilde{\theta},\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CI_{P,n}^{\beta}\right\}$ be an indicator for the hybrid conditioning event that $\hat{\theta}_n$ is equal to $\tilde{\theta}$ and the parameter of interest $\mu_Y(\tilde{\theta})$ falls in the level β projection confidence interval $CI_{P,n}^{\beta}$. We can establish quantile unbiasedness of hybrid estimators given this event, along with bounded unconditional bias.

Proposition 12

Under Assumptions 6-8, for $\hat{\mu}_{\alpha,n}^H$ the α-quantile unbiased hybrid estimator based on $CI_{P,n}^{\beta}$,

$$\lim_{n\to\infty}\sup_{P\in\mathcal{P}_n}\left|Pr_P\left\{\hat{\mu}_{\alpha,n}^H\geq\mu_{Y,n}\left(\hat{\theta}_n;P\right)|C_n^H\left(\tilde{\theta};P\right)=1\right\}-\alpha\right|E_P\left\{C_n^H\left(\tilde{\theta};P\right)\right\}=0, \tag{A.23}$$

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

$$\limsup_{n\to\infty}\sup_{P\in\mathcal{P}_n}\left|Pr_P\left\{\hat{\mu}_{\alpha,n}^H\geq\mu_{Y,n}\left(\hat{\theta}_n;P\right)\right\}-\alpha\right|\leq\max\{\alpha,1-\alpha\}\beta. \tag{A.24}$$

Validity of hybrid estimators again implies validity of hybrid confidence intervals.

Corollary 2

Under Assumptions 6-8, for $CI_{ET,n}^H$ the level $1-\alpha$ equal-tailed hybrid confidence interval based on $CI_{P,n}^{\beta}$,

$$\lim_{n\to\infty}\sup_{P\in\mathcal{P}_n}\left|Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CI_{ET,n}^H|C_n^H\left(\tilde{\theta};P\right)=1\right\}-\frac{1-\alpha}{1-\beta}\right|E_P\left\{C_n^H\left(\tilde{\theta};P\right)\right\}=0, \tag{A.25}$$

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

$$\liminf_{n\to\infty}\inf_{P\in\mathcal{P}_n}Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CI_{ET,n}^H\right\}\geq 1-\alpha, \tag{A.26}$$

and

$$\limsup_{n\to\infty}\sup_{P\in\mathcal{P}_n}Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CI_{ET,n}^H\right\}\leq\frac{1-\alpha}{1-\beta}\leq 1-\alpha+\beta. \tag{A.27}$$

Hence, our procedures are uniformly asymptotically valid, unlike conventional inference.[4]

[4]The bootstrap also fails to deliver uniform validity, as it implicitly tries to estimate the difference be-

F.2 Auxiliary Lemmas

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

Lemma 5

Under Assumption 8, for any sequence of confidence intervals CI_n, any sequence of sets $\mathcal{C}_n(P)$ indexed by P, $C_n(P)=1\left\{\left(X_n,Y_n,\widehat{\Sigma}_n\right)\in\mathcal{C}_n(P)\right\}$, and any constant α, to show that

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

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*

$$\mathcal{S}=\left\{\Sigma:1/\bar{\lambda}\leq(\Sigma_X(\theta),\Sigma_Y(\theta))\leq\bar{\lambda},1/\bar{\lambda}\leq\sqrt{\Sigma_X(\theta;P)\Sigma_X(\tilde{\theta};P)}-\Sigma_X(\theta,\tilde{\theta};P)\right\},\tag{A.28}$$

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

$$\lim_{s\to\infty}Pr_{P_{n_s}}\left\{\mu_{Y,n_s}\left(\hat{\theta}_{n_s};P_{n_s}\right)\in CI_{n_s}|C_{n_s}(P_{n_s})=1\right\}=\alpha.\tag{A.29}$$

Lemma 6

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

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

for all j, then

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

---

tween the "winning" policy and the others, which cannot be done with sufficient precision. We are unaware of results for subsampling, m-out-of-n bootstrap, or other resampling-based approaches for this setting.

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

To state the next lemma, define

$$\mathcal{L}\left(\tilde{\theta},Z,\Sigma\right)=\max_{\theta\in\Theta:\Sigma_{XY}\left(\tilde{\theta}\right)>\Sigma_{XY}\left(\tilde{\theta},\theta\right)}\frac{\Sigma_Y\left(\tilde{\theta}\right)\left(Z(\theta)-Z\left(\tilde{\theta}\right)\right)}{\Sigma_{XY}\left(\tilde{\theta}\right)-\Sigma_{XY}\left(\tilde{\theta},\theta\right)} \tag{A.30}$$

$$\mathcal{U}\left(\tilde{\theta},Z,\Sigma\right)=\min_{\theta\in\Theta:\Sigma_{XY}\left(\tilde{\theta}\right)<\Sigma_{XY}\left(\tilde{\theta},\theta\right)}\frac{\Sigma_Y\left(\tilde{\theta}\right)\left(Z(\theta)-Z\left(\tilde{\theta}\right)\right)}{\Sigma_{XY}\left(\tilde{\theta}\right)-\Sigma_{XY}\left(\tilde{\theta},\theta\right)}, \tag{A.31}$$

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

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

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

Lemma 7

For $\mathcal{L}\left(\tilde{\theta},Z,\Sigma\right)$ and $\mathcal{U}\left(\tilde{\theta},Z,\Sigma\right)$ as defined in (A.30) and (A.31), and

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

we have

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

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 6 and 7, for any $\{n_s\}$ and $\{P_{n_s}\}$ satisfying conditions (1)-(3) of Lemma 5 and any $\tilde{\theta}$ with $\mu_X^*\left(\tilde{\theta}\right)>-\infty$, $\left(Y_{n_s}^*,\mathcal{L}_{n_s}^*,\mathcal{U}_{n_s}^*,\widehat{\Sigma}_{n_s},\hat{\theta}_{n_s}\right)\to_d\left(Y^*,\mathcal{L}^*,\mathcal{U}^*,\Sigma^*,\hat{\theta}\right)$, where the *objects on the right hand side are calculated based on* (Y^*,X^*,Σ^*) *for* $\left(X^{*\prime},Y^{*\prime}\right)'\sim N(\mu^*,\Sigma^*)$ *with* $\mu^*=(\mu_X^{*\prime},0')'$.

Lemma 9

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

$$F_{TN}(Y(\theta);\mu,\Sigma_Y(\theta),\mathcal{L},\mathcal{U})=\frac{F_N\left(\frac{Y(\theta)\wedge\mathcal{U}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-F_N\left(\frac{\mathcal{L}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)}{F_N\left(\frac{\mathcal{U}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)-F_N\left(\frac{\mathcal{L}-\mu}{\sqrt{\Sigma_Y(\theta)}}\right)}1\{Y(\theta)\geq\mathcal{L}\} \tag{A.32}$$

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

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

F.3 Proofs for Auxiliary Lemmas

Proof of Lemma 5 To prove that

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

it suffices to show that

$$\liminf_{n\to\infty}\inf_{P\in\mathcal{P}_n}\left(Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CI_n|C_n(P)=1\right\}-\alpha\right)Pr_P\{C_n(P)=1\}\geq 0 \tag{A.33}$$

and

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

We prove that to show (A.33), it suffices to show that for all $\{n_s\}$, $\{P_{n_s}\}$ satisfying conditions (1)-(3) of the lemma,

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

An argument along the same lines implies that to prove (A.34) it suffices to show that

$$\limsup_{s\to\infty}Pr_{P_{n_s}}\left\{\mu_{Y,n_s}\left(\hat{\theta}_{n_s};P_{n_s}\right)\in CI_{n_s}|C_{n_s}(P_{n_s})=1\right\}\leq\alpha. \tag{A.36}$$

Note, however, that (A.35) and (A.36) together are equivalent to (A.29).

Towards contradiction, suppose that (A.33) fails, so

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

for some $\varepsilon>0$ but that (A.35) 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 $\{P_{n_q}\}$ with $P_{n_q}\in\mathcal{P}_{n_q}$ for all q such that

$$\limsup_{q\to\infty}\left(Pr_{P_{n_q}}\left\{\mu_{Y,n_q}\left(\hat{\theta}_{n_q};P_{n_q}\right)\in CI_{n_q}|C_{n_q}(P_{n_q})=1\right\}-\alpha\right)Pr_{P_{n_q}}\{C_{n_q}(P_{n_q})=1\}<-\varepsilon. \quad \text{(A.37)}$$

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 (A.28) is compact (e.g. in the Frobenius norm), and Assumption 8 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, (A.37) 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, (A.35) must hold along this subsequence. Thus,

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

which contradicts (A.37). Hence, we have established a contradiction and so proved that (A.35) for all subsequences satisfying conditions (1)-(3) of the lemma implies (A.33).

An argument along the same lines shows that (A.36) along all subsequences satisfying conditions (1)-(3) of the lemma implies (A.34). $\square$

Proof of Lemma 6 Define $C_{n,J+1}(P)=1\{C_{n,j}(P)=0 \text{ for all } j\in\{1,...,J\}\}$. Note that

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

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

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

and

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

which immediately implies that

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

Likewise,

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

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

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

This immediately implies that

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

as we wanted to show. □

Proof of Lemma 7 Note that

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

so

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

The result follows immediately. □

Proof of Lemma 8 By Assumption 6

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

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

7, if $\Sigma^*_{XY}\left(\tilde{\theta}\right)-\Sigma^*_{XY}\left(\tilde{\theta},\theta\right)\neq 0$ then

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

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

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

If instead $\Sigma^*_{XY}\left(\tilde{\theta}\right)-\Sigma^*_{XY}\left(\tilde{\theta},\theta\right)=0$, then since $\Sigma^*_X(\tilde{\theta},\theta)<\sqrt{\Sigma^*_X(\tilde{\theta})\Sigma^*_X(\theta)}$,

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

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

$$\left|\frac{\widehat{\Sigma}_{Y,n_s}\left(\tilde{\theta}\right)\left(Z^*_{n_s,\tilde{\theta}}(\theta)-Z^*_{n_s,\tilde{\theta}}\left(\tilde{\theta}\right)\right)}{\widehat{\Sigma}_{XY,n_s}\left(\tilde{\theta}\right)-\widehat{\Sigma}_{XY,n_s}\left(\tilde{\theta},\theta\right)}\right|\to\infty. \tag{A.38}$$

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

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

$$\to_d \mathcal{L}^*=\max_{\theta\in\Theta:\Sigma^*_{XY}\left(\tilde{\theta}\right)>\Sigma^*_{XY}\left(\tilde{\theta},\theta\right)}\frac{\Sigma^*_{Y}\left(\tilde{\theta}\right)\left(Z^*_{\tilde{\theta}}(\theta)-Z^*_{\tilde{\theta}}\left(\tilde{\theta}\right)\right)}{\Sigma^*_{XY}\left(\tilde{\theta}\right)-\Sigma^*_{XY}\left(\tilde{\theta},\theta\right)}, \tag{A.39}$$

and

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

$$\to_d \mathcal{U}^* = \min_{\theta\in\Theta:\Sigma^*_{XY}\left(\tilde{\theta}\right)<\Sigma^*_{XY}\left(\tilde{\theta},\theta\right)} \frac{\Sigma^*_Y\left(\tilde{\theta}\right)\left(Z^*_{\tilde{\theta}}(\theta)-Z^*_{\tilde{\theta}}\left(\tilde{\theta}\right)\right)}{\Sigma^*_{XY}\left(\tilde{\theta}\right)-\Sigma^*_{XY}\left(\tilde{\theta},\theta\right)}. \tag{A.40}$$

Since

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

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

with probability one for all n_s and $Y_{n_s} \xrightarrow{d} Y^*$, (A.38) implies

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

when $\Sigma^*_{XY}\left(\tilde{\theta}\right)=\Sigma^*_{XY}\left(\tilde{\theta},\theta\right)$ for all $\theta,\tilde{\theta}\in\Theta$ such that $\widehat{\Sigma}_{XY,n_s}\left(\tilde{\theta}\right)>\widehat{\Sigma}_{XY,n_s}\left(\tilde{\theta},\theta\right)$. Similarly,

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

when $\Sigma^*_{XY}\left(\tilde{\theta}\right)=\Sigma^*_{XY}\left(\tilde{\theta},\theta\right)$ for all $\theta,\tilde{\theta}\in\Theta$ such that $\widehat{\Sigma}_{XY,n_s}\left(\tilde{\theta}\right)<\widehat{\Sigma}_{XY,n_s}\left(\tilde{\theta},\theta\right)$. Thus, the same convergence results as (A.39)–(A.40) continue to hold when we minimize and maximize over Θ rather than $\Theta^*(\tilde{\theta})$. Hence, $(\mathcal{L}^*_{n_s},\mathcal{U}^*_{n_s})\to_d(\mathcal{L}^*,\mathcal{U}^*)$. 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. □

F.4 Proofs for Uniformity Results

Proof of Proposition 10 Note that

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

for $CI_{U,-,n}=(-\infty,\hat{\mu}_{\alpha,n}]$. Hence, by Lemma 5, to prove that (A.20) holds it suffices to show that for all $\{n_s\}$ and $\{P_{n_s}\}$ such that conditions (1)-(3) of the lemma hold with $C_n(P)=1\left\{\hat{\theta}_n=\tilde{\theta}\right\}$, we have

$$\lim_{s\to\infty}Pr_{P_{n_s}}\left\{\hat{\mu}_{Y,n_s}\left(\hat{\theta}_{n_s};P_{n_s}\right)\in CI_{U,-,n_s}|\hat{\theta}_{n_s}=\tilde{\theta}\right\}=\alpha. \tag{A.41}$$

To this end, recall that for $F_{TN}(Y(\theta);\mu,\Sigma_Y(\theta),\mathcal{L},\mathcal{U})$ as defined in (A.32), the estimator $\hat{\mu}_{\alpha,n}$ solves $F_{TN}\left(Y_n\left(\hat{\theta}_n\right);\mu,\widehat{\Sigma}_{Y,n}\left(\hat{\theta}_n\right),\mathcal{L}\left(\hat{\theta}_n,Z_{\hat{\theta}_n,n},\widehat{\Sigma}_n\right),\mathcal{U}\left(\hat{\theta}_n,Z_{\hat{\theta}_n,n},\widehat{\Sigma}_n\right)\right)=1-\alpha$. This cdf is strictly decreasing in μ as argued in the proof of Proposition 5, and is increasing in $Y_n\left(\hat{\theta}\right)$. Hence, $\hat{\mu}_{\alpha,n}\geq\mu_{Y,n}\left(\hat{\theta}_n;P\right)$ if and only if

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

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

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

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

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

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

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 $\operatorname{argmax}_\theta X^*(\theta)$ is almost surely unique and continuous for X^* as in Lemma 8, the continuous mapping theorem implies that

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

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 7 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 (A.41).

Since this argument holds for all $\tilde{\theta} \in \Theta$, and Assumptions 6 and 8 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 (A.21). $\square$

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

$$Pr_P\left\{\mu_{Y,n}\left(\hat{\theta}_n;P\right) \in CI_{ET,n} | \hat{\theta}_n=\tilde{\theta}\right\}$$

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

so the result is immediate from Proposition 10 and Lemma 6. □

Proof of Proposition 11 By the same argument as in the proof of Lemma 5, to show that (A.22) holds it suffices to show that for all $\{n_s\}$, $\{P_{n_s}\}$ satisfying conditions (1)-(3) of Lemma 5, $\liminf_{n\to\infty}Pr_{P_{n_s}}\left\{\mu_{Y,n_s}\left(\hat{\theta}_{n_s};P_{n_s}\right)\in CI_{P,n_s}\right\}\geq 1-\alpha$.

To this end, note that

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

for $c_\alpha(\Sigma_Y)$ the $1-\alpha$ quantile of $\max_\theta|\xi(\theta)|/\sqrt{\Sigma_Y(\theta)}$ where $\xi\sim N(0,\Sigma_Y)$. Next, note that $c_\alpha(\Sigma_Y)$ is continuous in Σ on $\mathcal{S}$ as defined in (A.28). Hence, for all θ, $c_\alpha(\Sigma_Y)\sqrt{\Sigma_Y(\theta)}$ is continuous as well. Assumptions 6 and 7 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 $Pr\left\{\left|Y^*\left(\hat{\theta}\right)\right|-c_\alpha(\Sigma^*_Y)\sqrt{\Sigma^*_Y\left(\hat{\theta}\right)}=0\right\}=0$,

$$Pr_{P_{n_s}}\left\{\mu_{Y,n_s}\left(\hat{\theta}_{n_s};P_{n_s}\right)\in CI_{P,n_s}\right\}\to Pr\left\{Y^*\left(\hat{\theta}\right)\in\left[-c_\alpha(\Sigma^*_Y)\sqrt{\Sigma^*_Y\left(\hat{\theta}\right)},c_\alpha(\Sigma^*_Y)\sqrt{\Sigma^*_Y\left(\hat{\theta}\right)}\right]\right\} \quad \text{(A.42)}$$

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

Proof of Proposition 12 Note that $\hat{\mu}^H_{\alpha,n}\geq\mu_{Y,n}\left(\hat{\theta}_n;P\right)$ if and only if $\mu_{Y,n}\left(\hat{\theta}_n;P\right)\in CI^H_{U,-,n}$ for $CI^H_{U,-,n}=(-\infty,\hat{\mu}^H_{\alpha,n}]$. Hence, by Lemma 5, to prove that (A.23) holds it suffices to show that for all $\{n_s\}$ and $\{P_{n_s}\}$ such that conditions (1)-(3) of the lemma hold with $C_n(P)=1\left\{\hat{\theta}_n=\tilde{\theta},\mu_{Y,n}\left(\hat{\theta}_n;P_n\right)\in CI^\beta_{P,n}\right\}$, we have

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

Recall that for $F_{TN}(Y(\theta);\mu,\Sigma_Y(\theta),\mathcal{L},\mathcal{U})$ defined as in (A.32), $\hat{\mu}^H_{\alpha,n}$ solves

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

for

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

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

The proof of Proposition 5 shows that $F_{TN}\left(Y_n\left(\hat{\theta}_n\right);\mu,\widehat{\Sigma}_{Y,n}\left(\hat{\theta}_n\right),\mathcal{L}_n^H\left(\mu,\hat{\theta}_n\right),\mathcal{U}_n^H\left(\mu,\hat{\theta}_n\right)\right)$ is strictly decreasing in μ, so for a given value $\mu_{Y,0}$,

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

As in the proof of Proposition 10

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

where

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

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

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

Lemma 8 implies that $\left(Y_{n_s}^*,\widehat{\Sigma}_{Y,n_s},\mathcal{L}_{n_s}^{H*}\left(\tilde{\theta}\right),\mathcal{U}_{n_s}^{H*}\left(\tilde{\theta}\right),\hat{\theta}_{n_s}\right)\to_d\left(Y^*,\Sigma_Y^*,\mathcal{L}^{H*}\left(\tilde{\theta}\right),\mathcal{U}^{H*}\left(\tilde{\theta}\right),\hat{\theta}\right)$, where $\mathcal{L}^{H*}\left(\tilde{\theta}\right)$ and $\mathcal{U}^{H*}\left(\tilde{\theta}\right)$ are equal to $\mathcal{L}_n^{H*}\left(\tilde{\theta}\right)$ and $\mathcal{U}_n^{H*}\left(\tilde{\theta}\right)$ after replacing $(X_n,Y_n,\widehat{\Sigma}_n)$ with (X,Y,Σ^*). Then by the continuous mapping theorem and (A.42),

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

Hence, by the same argument as in the proof of Proposition 10,

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

as we aimed to show.

To prove (A.24), note that for $\widetilde{CI}^H_{U,+,n}=(\hat{\mu}^H_{\alpha,n},\infty)$,

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

and thus that the argument above proves that

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

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

$$\sum_{\tilde{\theta}\in\Theta}Pr_P\left\{\hat{\theta}_{n_s}=\tilde{\theta},\mu_{Y,n_s}\left(\hat{\theta}_{n_s};P_{n_s}\right)\in CI^{\beta}_{P,n_s}\right\}=Pr_P\left\{\mu_{Y,n_s}\left(\hat{\theta}_{n_s};P_{n_s}\right)\in CI^{\beta}_{P,n_s}\right\}+o(1),\qquad\text{(A.43)}$$

and Proposition 11 shows that

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

Lemma 6 together with (A.23) implies that

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

and

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

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

Proof of Corollary 2 Note that by construction $CI^H_{ET,n}=\left[\hat{\mu}^H_{\frac{\alpha-\beta}{2(1-\beta)},n},\hat{\mu}^H_{1-\frac{\alpha-\beta}{2(1-\beta)},n}\right]$, where $\hat{\mu}^H_{\frac{\alpha-\beta}{2(1-\beta)},n}<\hat{\mu}^H_{1-\frac{\alpha-\beta}{2(1-\beta)},n}$ provided $\frac{\alpha-\beta}{1-\beta}<1$. Hence,

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

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

so Proposition 12 immediately implies (A.25).

Equation (A.43) in the proof of Proposition 12 together with Lemma 6 implies that

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

so (A.26) 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

$$\mu_{Y,n}\left(\hat{\theta}_n;P_n\right)\in CI_{ET,n}^{H}\Rightarrow\mu_{Y,n}\left(\hat{\theta}_n;P_n\right)\in CI_{P,n}^{\beta}.$$

Hence,

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

By the first part of the proposition, this implies that

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

so (A.27) holds as well. □

G Additional Materials for Neighborhoods Application

This appendix provides additional details on the neighborhoods application discussed in Section 7 of the main text.

G.1 Simulation Design and Target Parameters

Our simulations take the census tract-level, un-shrunk estimates from the Opportunity Atlas as the true parameter values. The true parameter value in tract t, μ_t, thus corresponds to tract-level average household income rank in adulthood for children growing up in households at the 25th percentile of the income distribution.[5] We simulate estimates for tract t by drawing $\hat{\mu}_t\sim N(\mu_t,\sigma_t^2)$, for μ_t the Opportunity Atlas estimate and σ_t the Opportunity Atlas standard error. We treat the draws $\hat{\mu}_t$ as independent across tracts. Hence, in each commuting zone, for $\mathcal{T}$ the set of tracts in that CZ we generate jointly normal sets of estimates and corresponding standard errors, $\{(\hat{\mu}_t,\sigma_t):t\in\mathcal{T}\}$. Our analysis also drops the single tract in the data where $\sigma_t>1$, (i.e. a standard error in excess of 100 percentile points) which was located in the Seattle CZ. No other tract in the CZs we consider has a standard error larger that 0.2.

As discussed in the main text, in each commuting zone we define Θ as the set of selections containing one third of the tracts in $\mathcal{T}$ (rounded down), $\Theta=\{\theta\subset\mathcal{T}:|\theta|=\lfloor|\mathcal{T}|/3\rfloor\}$. For $X(\theta)=$

[5] Specifically, we focus on kfr_pooled_pooled_p25 from in the file tract_outcomes_simple.cvs, available at `https://opportunityinsights.org/wp-content/uploads/2018/10/tract_outcomes_simple.csv`, downloaded on May 10, 2022.

$\frac{1}{|\theta|}\sum_{t\in\theta}\hat{\mu}_t$ the average estimate over tracts in θ, $\hat{\theta}=\text{argmax}_\theta X(\theta)$ thus selects the third of tracts with the largest estimates. For c_t the number of voucher households with children reported to be living in tract t,[6] we define $Y(\theta)$ as a difference between $X(\theta)$ and a c_t weighted average of $\hat{\mu}_t$ over all tracts in the commuting zone $Y(\theta)=X(\theta)-\frac{\sum_t c_t\hat{\mu}_t}{\sum_t c_t}$. Our target parameter $\mu_Y(\theta)=E[Y(\theta)]=\frac{1}{|\theta|}\sum_{t\in\theta}\mu_t-\frac{\sum_t c_t\mu_t}{\sum_t c_t}$ thus measures the difference between the weighted average outcome in the tracts where voucher households with children currently live and the average across targeted tracts.

We examine the performance of several estimators and confidence intervals for $\mu_Y(\hat{\theta})$ in each of our 50 commuting zones. To do so, for each commuting zone and each draw $\hat{\mu}=(\hat{\mu}_1,...,\hat{\mu}_{|\mathcal{T}|})$ we compute the corresponding estimators $\hat{\mu}_Y$ and confidence intervals CI. For each procedure considered (detailed in the next section) we report the median bias, $Med\left(\hat{\mu}_Y-\mu_Y\left(\hat{\theta}\right)\right)$, the median absolute error, $Med\left(\left|\hat{\mu}_Y-\mu_Y\left(\hat{\theta}\right)\right|\right)$, the coverage probability $Pr\left\{\mu_Y\left(\hat{\theta}\right)\in CI\right\}$, and the median confidence interval length, $Med(|CI|)$. Specifically, for each commuting zone and each simulation draw $s\in\{1,...,S\}$ for $S=10{,}000$ we draw $\hat{\mu}^s$ as described above. We then form $X_s(\theta)=\frac{1}{|\theta|}\sum_{t\in\theta}\hat{\mu}_t^s$ and set $\hat{\theta}_s=\text{argmax}_\theta X_s(\theta)$. For $\hat{\mu}_{Y,s}$ and CI_s the resulting point estimate and confidence interval in simulation draw s, we then approximate the average coverage in that commuting zone by the sample average $\frac{1}{S}\sum_s 1\left\{\mu_Y\left(\hat{\theta}_s\right)\in CI_s\right\}$, while $Med\left(\hat{\mu}_{Y,s}-\mu_Y\left(\hat{\theta}\right)\right)$ and the other medians are correspondingly approximated by sample medians across simulation draws. We record these quantities separately in each commuting zone, and Figure IV in the main text shows their distribution across commuting zones.

Note that in all cases we focus on unconditional performance measures, both because we think these are of substantive interest in this application, and because computing and reporting conditional quantities (e.g. conditional coverage probabilities $Pr\left\{\mu_Y\left(\hat{\theta}\right)\in CI|\hat{\theta}=\tilde{\theta}\right\}$) is difficult in this setting given the prohibitively large size of $|\Theta|$, which is equal to $|\mathcal{T}|$ choose $\lfloor|\mathcal{T}|/3\rfloor$.

G.2 Fixed-Length Projection Intervals

We report confidence intervals and, where applicable, point estimates for the conventional, projection, conditional, and hybrid approaches. Following Chetty et al. (2020) we also report results based on an empirical Bayes approach. We defer discussion of the empirical Bayes approach to the next section, while the conventional and conditional approaches are as described in the main text. Due to the large size of Θ, however, we need to modify the projection (and thus hybrid) approach for this application.

Specifically, the projection approach considered in the main text ensures that the width of the projection interval for $\mu_Y(\hat{\theta})$ is proportional to $\sqrt{\Sigma_Y(\hat{\theta})}$. To compute these intervals, we need to approximate the critical value c_α, which we do by repeatedly drawing $\xi\sim N(0,\Sigma_Y)$ and setting

[6] Computed based on the 2018 Picture of Subsidized Housing dataset from the US Department of Housing and Urban Development, (US Department of Housing and Urban Development, 2018).

c_α equal to the $1-\alpha$ quantile of $\max_\theta|\xi(\theta)|/\sqrt{\Sigma_Y(\theta)}$. In the neighborhoods application, however, solving the optimization problem $\max_\theta|\xi(\theta)|/\sqrt{\Sigma_Y(\theta)}$ requires a numerical search over Θ, and is computationally prohibitive. To sidestep this computational challenge, in this application we consider *fixed length* projection intervals which set $CI_P=\left[Y(\hat{\theta})-c_\alpha^*,Y(\hat{\theta})+c_\alpha^*\right]$.

We term these fixed length intervals since their length does not depend on $\hat{\theta}$. The critical value c_α^* corresponds to the $1-\alpha$ quantile of $\max_\theta|\xi(\theta)|$, again for $\xi\sim N(0,\Sigma_Y)$. Unlike the original critical value c_α, we can easily approximate c_α^* by simulation. In particular, to compute $\max_\theta|\xi(\theta)|$ it suffices to independently draw $\xi_t\sim N(0,\sigma_t^2)$ in each tract t, for σ_t the Opportunity Atlas standard error for that tract. We then sort the tract-level noise draws ξ_t and select either the top or bottom third, whichever yields a larger average in absolute value, and take $\max_\theta|\xi(\theta)|$ equal to the resulting (absolute) average. Validity of the resulting projection confidence interval follows by the same argument as before, as does validity of hybrid estimators and intervals based on this version of projection.

G.3 Empirical Bayes and Winner's Curse

Chetty et al. (2020) focus on what they term forecast-unbiased estimates. These correspond to posterior means from a correlated random effects model which treats mobility as normally distributed conditional on a set of observable tract characteristics, with a mean that changes linearly in the tract characteristics and a constant variance. Specifically, for W_t the characteristics of tract t, these estimates correspond to posterior means under the prior π that takes μ_t independent across tracts, with

$$\mu_t|W_t\sim N\left(W_t'\beta,\omega^2\right). \tag{A.44}$$

They then plug in estimates of ω and β, so in our simulations we do the same.

If we take the model (A.44) seriously and abstract from estimation of ω and β (for instance because the number of tracts is large and we plug in consistent estimates), Bayesian posterior means solve the winner's curse problem under the prior. Specifically, note that the posterior mean for μ_t given the vector of estimates $\hat{\mu}$ is simply the mean given $\hat{\mu}_t$, $E_\pi[\mu_t|\hat{\mu}]=E_\pi[\mu_t|\hat{\mu}_t]$. The law of iterated expectations implies, however, that $E_\pi[\mu_t|\hat{\mu}]$ is unbiased for μ_t conditional on $\hat{\mu}$, so for any set $\mathcal{E}$ such that $Pr_\pi\{\hat{\mu}\in\mathcal{E}\}>0$,

$$E_\pi[\mu_t-E_\pi[\mu_t|\hat{\mu}_t]|\hat{\mu}\in\mathcal{E}]=0.$$

Likewise, since we model $\hat{\mu}_t$ as normally distributed conditional on μ_t, the posterior mean is also the posterior median, so

$$Pr_\pi\{E_\pi[\mu_t|\hat{\mu}_t]>\mu_t|\hat{\mu}\in\mathcal{E}\}=\frac{1}{2},$$

and $E_\pi[\mu_t|\hat{\mu}_t]$ is median-unbiased under the prior conditional on the event $\{\hat{\mu}\in\mathcal{E}\}$. Note, however, that selection of a particular set of target tracts can be written as an event $\{\hat{\mu} \in \mathcal{E}\}$, so this argument implies that Bayesian posterior means are immune to the winner's curse under the prior. This depends crucially on the prior, however, since if we calculate the outer probability with respect to some other distribution of effect sizes $\tilde{\pi} \neq \pi$, we typically have

$$Pr_{\tilde{\pi}}\{E_{\pi}[\mu_t|\hat{\mu}_t] > \mu_t | \hat{\mu} \in \mathcal{E}\} \neq \frac{1}{2}.$$

G.4 Additional Figure for Movers Application

Figure 9 plots conventional, conditional, and projection intervals for the Opportunity Atlas application described in the main text.

In our main analysis we compare a simple average over targeted tracts to a weighted average over tracts in the same commuting zone, weighting by the number of voucher-recipient households with children. It may also be of interest to treat the targeted and non-targeted tracts more symmetrically, comparing a simple average over the target tracts to a simple average over the commuting zone as a whole. Formally, this corresponds to the alternative target parameter $\bar{\mu}_Y(\theta) = \frac{1}{|\theta|}\sum_{t\in\theta}\mu_t - \frac{1}{|\mathcal{T}|}\sum_t \mu_t$. Figures 10 and 11 report results for this alternative target in the Opportunity Atlas data. These results are qualitatively similar to our baseline results, with the notable exception that the coefficients are substantially smaller. The smaller coefficients from under this weighting reflects that voucher recipient households are located in neighborhoods with below-average mobility at baseline. Hence, a simple average of mobility across tracts in a commuting zone yields a larger average economic mobility than the weighted average we consider in our main results, and thus a smaller contrast with the targeted tracts

G.5 Evaluating the Normality Assumption in Empirical Bayes

As discussed in Appendix G.3, Bayesian approaches are immune to the winner's curse in settings where the distribution of true effects is correctly described by the prior. In Section 7 of the main text we saw that empirical Bayes credible sets display large coverage distortions in simulations calibrated to the Opportunity Atlas data, which suggests that the normal approximation is not entirely reliable in this application. This appendix formally investigates the quality of the normal approximation in this application and its relationship to the coverage of empirical Bayes credible sets.

We begin by considering formal tests for normality in each of the 50 CZs used in our simulation. Specifically, the normal prior used by empirical Bayes specifies that

$$\mu_t | W_t \sim N\left(W_t'\beta, \omega^2\right),$$

for W_t a set of (observed) track-level covariates. We observe tract-level estimates $\hat{\mu}_t \sim N\left(\mu_t, \sigma_t^2\right)$

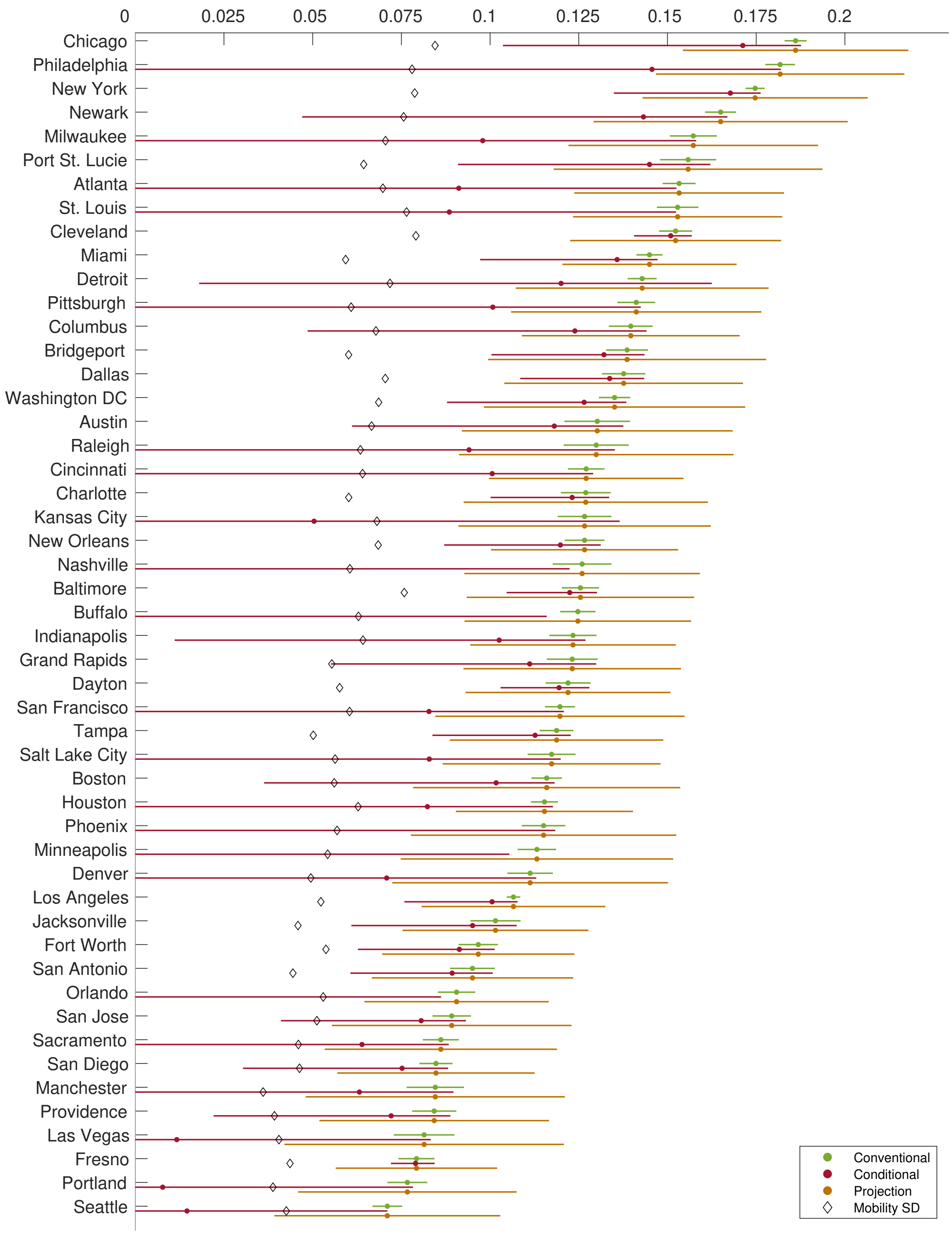

Figure 9: Estimates and confidence intervals for average economic mobility for selected census tracts based on the Chetty et al. (2020) Opportunity Atlas, relative to the within-CZ average, weighted by number of voucher recipient households with children. CZs are ordered by the magnitude of the conventional estimate. A coefficient of 0.1 implies that the target tracts are associated with a 10 percentile point higher average household income in adulthood, for children growing up in households at the 25th percentile of the income distribution, relative to the weighted average across the CZ. Diamonds plot the estimated standard deviation of mobility across all tracts in each CZ.

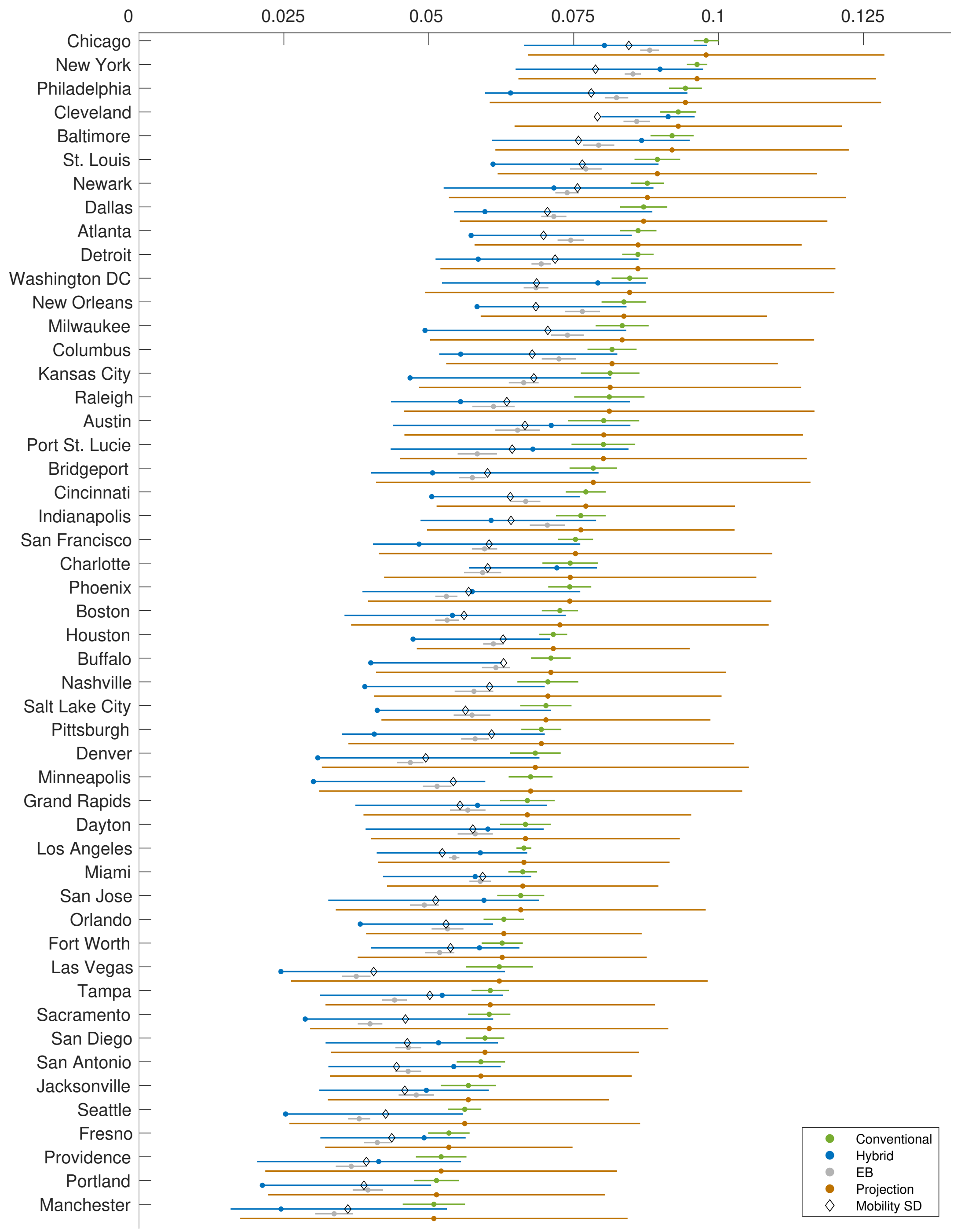

Figure 10: Estimates and confidence intervals for average economic mobility for selected census tracts based on the Chetty et al. (2020) Opportunity Atlas, relative to the within-CZ unweighted average. CZs are ordered by the magnitude of the conventional estimate. A coefficient of 0.1 implies that the target tracts are associated with a 10 percentile point higher average household income in adulthood, for children growing up in households at the 25th percentile of the income distribution, relative to the weighted average across the CZ. Diamonds plot the estimated standard deviation of mobility across all tracts in each CZ.

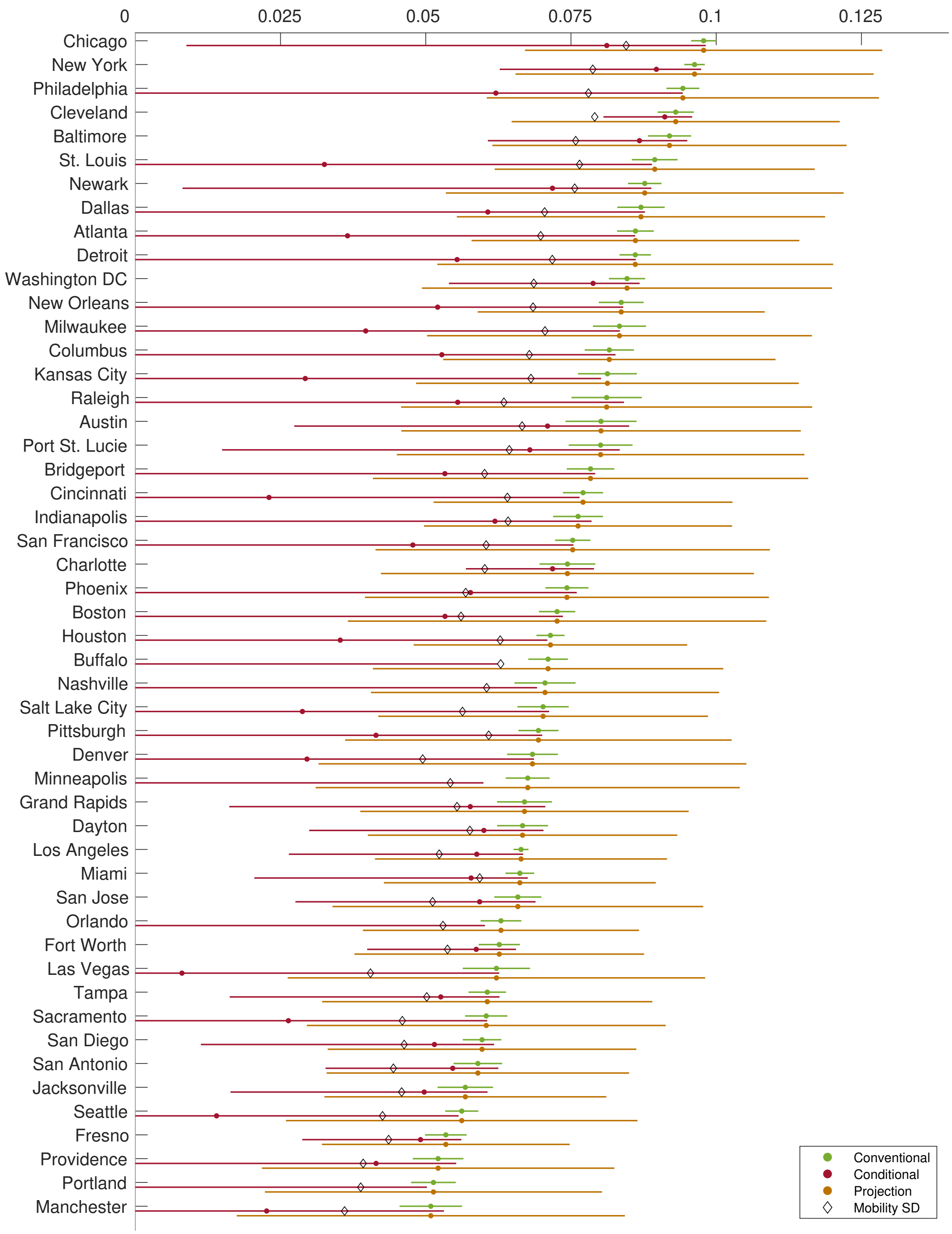

Figure 11: Estimates and confidence intervals for average economic mobility for selected census tracts based on the Chetty et al. (2020) Opportunity Atlas, relative to the within-CZ unweighted average. CZs are ordered by the magnitude of the conventional estimate. A coefficient of 0.1 implies that the target tracts are associated with a 10 percentile point higher average household income in adulthood, for children growing up in households at the 25th percentile of the income distribution, relative to the weighted average across the CZ. Diamonds plot the estimated standard deviation of mobility across all tracts in each CZ.

for σ_t^2 known. If we knew β and ω we could thus compute

$$\frac{\hat{\mu}_t - W_t'\beta}{\sqrt{\omega^2 + \sigma_t^2}},$$

which would follow a standard normal distribution if the normal prior were correct. Since β and ω are in fact unknown, within each CZ we construct an estimate $\hat{\beta}$ by regressing $\hat{\mu}_t$ on W_t, and construct an estimate $\hat{\omega}^2$ as the average squared residual minus the average of σ_t^2 (or zero in the case where $\hat{\omega}^2$ would be negative). We conduct a Kolmogorov-Smirnov (KS) test for whether

$$\frac{\hat{\mu}_t - W_t'\hat{\beta}}{\sqrt{\hat{\omega}^2 + \sigma_t^2}}$$

follows a standard normal distribution, where to account for estimation of the parameters we compare the KS statistic to a bootstrap critical value, obtained via parametric bootstrap imposing the normal model. The resulting KS test rejects normality in 37 CZs at the 5% level, and in 27 CZs in the 1% level.

We further examine the relationship between the quality of the normal approximation and the performance of empirical Bayes. Specifically, we calculate the rank correlation between the coverage of empirical Bayes credible sets and the KS statistic across the 50 CZs, obtaining a correlation of -0.15 (with a p-value of 0.13). This correlation increases to -0.21 (with a p-value of 0.033) when we recenter the KS statistic in each CZ around the mean of its bootstrap distribution. Hence CZs with a larger KS statistic tend to have worse coverage for empirical Bayes credible sets, though the low correlation suggests that the KS statistic does not fully capture the features of the distribution most important for determining the coverage of empirical Bayes credible sets.

Index	Treatment	Description
0	0	Control group with no matched donations
Match ratio		
1	1:1	An additional dollar up to the match limit
2	2:1	Two additional dollars up to the match limit
3	3:1	Three additional dollars up to the match limit
Match size		
1	$25,000	Up to $25,000 is pledged
2	$50,000	Up to $50,000 is pledged
3	$100,000	Up to $100,000 is pledged
4	Unstated	The pledged amount is not stated
Ask amount		
1	Same	The individual is asked to give as much as their largest past donation
2	25% more	The individual is asked to give 25% more than their largest past donation
3	50% more	The individual is asked to give 50% more than their largest past donation

Table 4: Treatment arms for Karlan and List (2007). Individuals were assigned to the control group or to the treatment group, in the ratio 1 : 2. Treated individuals were randomly assigned a match ratio, a match size and an ask amount with equal probability. There are 36 possible combinations, plus the control group. The leftmost column specifies a reference index used throughout this section for convenience.

H Application: Charitable Giving

Karlan and List (2007) partner with a political charity to conduct a field experiment examining the effectiveness of matching incentives at increasing charitable giving. In matched donations, a lead donor pledges to 'match' any donations made by other donors up to some threshold, effectively lowering the price of political activism for other donors.

Karlan and List (2007) use a factorial design. Potential donors, who were previous donors to the charity, were mailed a four page letter asking for a donation. The contents of the letter were randomized, with one third of the sample assigned to a control group that received a standard letter with no match. The remaining two thirds received a letter with the line "now is the time to give!" and details for a match. Treated individuals were randomly assigned with equal probability to one of 36 separate treatment arms. Treatment arms are characterized by a match ratio, a match size, and an ask amount, for which further details are given in Table 4. The outcome

Treatment	Average donation	Standard error	95% CI
(1,3,2)	1.52	0.35	[0.83,2.20]
(2,1,3)	1.51	0.46	[0.61,2.41]
(2,1,1)	1.42	0.39	[0.66,2.19]
(3,1,3)	1.40	0.36	[0.70,2.11]

Table 5: The average donations for the four best treatment arms according to the data, $n = 50{,}083$. Treatments are indexed by the indicators for (Match ratio, Match size, Ask amount) defined in Table 4. The reported 95% confidence intervals are the conventional ones that do not take selection into account.

of interest is the average dollar amount that individuals donated to the charity in the month following the solicitation.

In total, 50,083 individuals were contacted, of which 16,687 were randomly assigned to the control group, while 33,396 were randomly assigned to one of the 36 treatment arms. The (unconditional) average donation was $0.81 in the control group and $0.92 in the treatment group. Conditional on giving, these figures were $45.54 and $44.35, respectively. The discrepancy reflects the low response rate; only 1,034 of 50,083 individuals donated.

Table 5 reports average revenue from the four best-performing treatment arms, along with standard errors and conventional confidence intervals. The point estimates for the best-performing arm suggest that a campaign that promises a dollar-for-dollar match up to $100,000 in donations and asks individuals to donate 25% more than their largest past donation raises $1.52 per potential donor, on average, with a confidence interval of $0.83 to $2.20. This estimate and confidence interval are clearly subject to winner's curse bias, however: we are picking the best-performing arm out of 37 in the experiment, which will bias our estimates and confidence intervals upward.

Simulation Results To investigate the extent of winner's curse bias and the finite-sample performance of our corrections, we calibrate simulations to this application. We simulate datasets by resampling observations with replacement from the Karlan and List (2007) data (i.e. by drawing nonparametric bootstrap samples). In each simulated sample we re-estimate the effectiveness of each treatment arm, pick the best-performing arm, and study the performance of estimates and confidence intervals, treating the estimates for the original Karlan and List (2007) data as the true values. The underlying data here are non-normal and we re-estimate the variance in each simulation draw. Hence, these results also speak to the finite-sample performance of the normal approximation. We report results based on 10,000 simulation draws.

Since revenue does not account for the cost of the fund-raising campaign, it is impossible for the solicitation to raise a negative amount. We therefore set the parameter space for $\mu(\hat{\theta})$

Winner											
(1,3,2)	(1,4,2)	(1,4,3)	(2,1,1)	(2,1,3)	(2,2,2)	(2,3,3)	(2,4,1)	(2,4,2)	(3,1,1)	(3,1,3)	(3,3,1)
16.0%	11.4%	1.3%	13.0%	18.9%	10.8%	1.3%	1.5%	2.8%	5.1%	10.0%	3.6%

Table 6: Frequency of simulation replications where each treatment is estimated to perform best in simulations calibrated to Karlan and List (2007). Treatments are indexed by the indicators for (Match ratio, Match size, Ask amount) defined in Table 4. 31 of the 37 treatments are best in at least one replication; those that won in at least 1% of simulated samples are reported.

	Estimate		
	Conventional	Median unbiased	Hybrid
Median bias	0.61	-0.18	-0.18
Probability bias	0.50	-0.07	-0.07
Median absolute error	0.61	0.65	0.64

Table 7: Performance measures for alternative estimators in simulations calibrated to Karlan and List (2007). Probability bias is $Pr\{\hat{\mu}^{trim} > \mu(\hat{\theta})\} - \frac{1}{2}$.

to $\mathbb{R}_+$, and trim the point estimators and the confidence intervals at zero, $\hat{\mu}^{trim} \equiv \max\{0,\hat{\mu}\}$ and $CS^{trim} = [0,\infty) \cap CS$. This trimming does not affect the coverage of the confidence intervals, and also preserves the α-quantile unbiasedness of the estimators so long as the true value $\mu(\hat{\theta})$ is greater than zero.

There is substantial variability in the "winning" arm: 31 of the 37 treatments won in at least one simulation draw and 12 treatment arms won in at least 1% of simulated samples. Table 6 lists these 12 treatments. The variability of the winning arm suggests that there is scope for a winner's curse in this setting.

Table 7 examines the performance of conventional, median unbiased, and hybrid estimates, reporting (unconditional) median bias, probability bias ($Pr\{\hat{\mu}^{trim} > \mu(\hat{\theta})\} - \frac{1}{2}$), and median absolute error. Trimming the estimators at zero does not affect the reported performance measures. Conventional estimates suffer from substantial bias in this setting: they have a median bias of \$0.61, and over-estimate the revenue generated by the selected arm 100% of the time, up to rounding. The median unbiased and hybrid estimators substantially improve both measures of bias, though given the finite-sample setting they do not eliminate it completely and are both somewhat downward biased, though to a lesser degree.[7] All three estimators perform similarly in terms of median absolute error.

[7] This is a particularly challenging setting for the normal approximation, as the outcomes distribution is highly skewed due to the large number of zeros. In particular, there are on average only 20 nonzero outcomes per non-control treatment (out of approximately 930 observations in each treatment group).

Tables 8 and 9 report results for confidence intervals. Specifically, we consider the conventional, projection, conditional, and hybrid confidence intervals with nominal coverage 95%. Table 8 reports unconditional coverage and median length, while Table 9 reports conditional coverage probabilities given $\hat{\theta}$ values among the 12 treatments listed in Table 6. Conventional confidence intervals slightly undercover unconditionally, with coverage 92%. Their conditional coverage varies depending on which treatment is the winner. If the winning treatment is one of the six best-performing treatments, the conditional coverage is at least 95%, while otherwise the conventional confidence intervals under-cover with coverage probability as low as 65%. Projection

	Unconditional coverage	Median length Trimmed	Untrimmed
Conventional CS	0.92	1.88	1.88
CS_P	1.00	3.08	3.08
CS_{ET}	0.97	2.69	5.91
CS_{ET}^H	0.97	2.52	2.56

Table 8: Unconditional coverage probabilities of the confidence intervals in simulations calibrated to Karlan and List (2007). Unconditional median lengths are reported for the trimmed and untrimmed confidence intervals.

Treatment θ	Average donation $\mu(\theta)$	Conditional coverage Conventional CS	CS_P	CS_{ET}	CS_{ET}^H
(1,3,2)	1.52	0.95	1	0.98	0.98
(2,1,3)	1.51	0.97	1	0.97	0.97
(2,1,1)	1.42	0.94	1	0.97	0.97
(3,1,3)	1.40	0.95	1	0.97	0.97
(2,2,2)	1.34	0.96	1	0.97	0.98
(1,4,2)	1.27	0.99	1	0.97	0.97
(3,3,1)	1.26	0.84	1	0.96	0.97
(3,1,1)	1.24	0.89	1	0.97	0.97
(2,4,2)	1.22	0.79	1	0.99	0.99
(2,3,3)	1.12	0.65	1	0.98	0.98
(2,4,1)	1.10	0.81	1	0.97	0.97
(1,4,3)	1.03	0.78	1	0.96	0.97

Table 9: Conditional coverage probabilities, $Pr\{\mu(\hat{\theta})\in CS^{trim}|\hat{\theta}=\theta\}$, of the confidence intervals for each of the 12 treatments in Table 6. The treatments are sorted according to the average donation.

Treatment (1,3,2)		Estimates	Equal-tailed CI
Conventional		1.52	[0.83,2.20]
Projection		–	[0.40,2.63]
Conditional	– trimmed	0	[0, 1.42]
	– untrimmed	-7.49	[-47.66,1.42]
Hybrid		0.20	[0.19,1.47]

Table 10: Conventional and bias-corrected estimates and confidence intervals for best-performing treatment in Karlan and List (2007) data.

confidence intervals over-cover unconditionally and conditionally for these treatments, with coverage 100%. Conditional and hybrid confidence intervals slightly over-cover, with unconditional and conditional coverage about 97%, and have unconditional median (trimmed) length around 35% larger than conventional intervals and around 20% shorter than projection intervals. It is important to emphasize, however, that the conditional coverage for projection and hybrid intervals is particular to the data generating process considered here: as illustrated in Figure IV of the main text, these intervals do not ensure conditional coverage in general.

The median length of conditional intervals more than doubles if we leave their lower bound untrimmed. In contrast, the median length of the hybrid confidence intervals is basically unaffected by trimming. This is because despite the similarity of their upper bounds, the lower bound of the conditional confidence intervals tends to be negative and substantially lower than the lower bound of the hybrid confidence intervals. In other words, if the parameter space is unconstrained, the hybrid confidence intervals are substantially shorter than conditional confidence intervals. The good performance of the hybrid approach in applications with unconstrained parameter space is encouraging, and in line with the results in Section 2.

Empirical results Returning to the Karlan and List (2007) data, Table 10 reports corrected estimates and confidence intervals for the best-performing treatment in the experiment. We repeat the conventional estimate and confidence interval for comparison. The median unbiased estimate makes an aggressive downwards correction to the conventional estimate, suggesting negative revenue (-$7.49) from the winning arm if not trimmed. The conditional confidence interval is tight, ranging from 0 to $1.42, if trimmed at zero, and otherwise extremely wide, ranging from -$47.66 to $1.42. The hybrid estimate also shifts the conventional estimate downwards, but much less so. Moreover, the hybrid confidence interval is no wider than the conventional interval, and excludes both zero and the conventional estimate. These results suggest that future fundraising campaigns deploying the winning strategy in the experiment are likely to raise some revenue, but substantially less than would be expected based on the conventional estimates.

Conditional inference seems potentially natural in this application. The data highlight an interpretable combination of treatment parameters (1:1 match, $100,000 pledged, with an ask 25% above an individual's highest past donation) as best-performing, raising the question of what we can conclude about this particular treatment, given that it was the best in the experiment. This is precisely the question answered by the conditional approach. By contrast, while the hybrid approach ensures correct coverage on average across different "winning" treatments which could arise, it offers no guarantees given the particular winner observed in the Karlan and List (2007) data.

References

Andrews, D., X. Cheng, and P. Guggenberger (2020). Generic results for establishing the asymptotic size of confidence sets and tests. *Journal of Econometrics 218*, 496–531.

Andrews, I., T. Kitagawa, and A. McCloskey (2021). Inference after estimation of breaks. *Journal of Econometrics 224*, 39–59.

Cave, G., H. Bos, F. Doolittle, and C. Toussaint (1993). JOBSTART – Final report on a program for school dropouts. Technical report, Manpower Demonstration Research Corporation (MDRC), New York, NY.

Chetty, R., J. Friedman, N. Hendren, M. R. Jones, and S. R. Porter (2020). The opportunity atlas: Mapping the childhood roots of social mobility. Unpublished Manuscript.

Eaton, M. L. (1967). Some optimum properties of ranking procedures. *Annals of Mathematical Statistics 38*(1), 124–137.

Karlan, D. and J. A. List (2007). Does price matter in charitable giving? evidence from a large-scale natural field experiment. *American Economic Review 97*(5), 1774–1793.

Lee, J. D., D. L. Sun, Y. Sun, and J. E. Taylor (2016). Exact post-selection inference, with application to the LASSO. *Annals of Statistics 44*(3), 907–927.

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

Lehmann, E. L. (1966). On a theorem of bahadur and goodman. *Annals of Mathematical Statistics 37*(1), 1–6.

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

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

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

US Department of Housing and Urban Development (2018). Picture of subsidized households.

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