Inference on Multiple Winners with Applications to Economic Mobility*

Andreas Petrou-Zeniou
Department of Economics
MIT
apetrouz@mit.edu

Azeem M. Shaikh
Department of Economics
University of Chicago
amshaikh@uchicago.edu

September 19, 2025

Abstract

This paper considers the problem of inference on multiple winners. In our setting, a winner is defined abstractly as any population whose rank according to some random quantity, such as an estimated treatment effect, a measure of value-added, or benefit (net of cost), falls in a pre-specified range of values. As such, this framework generalizes the inference on a single winner setting previously considered in Andrews et al. (2023), in which a winner is understood to be the single population whose rank according to some random quantity is highest. We show that this richer setting accommodates a broad variety of empirically-relevant applications. We develop a two-step method for inference in the spirit of Romano et al. (2014), which we compare to existing methods or their natural generalizations to this setting. We first show the finite-sample validity of this method in a normal location model and then develop asymptotic counterparts to these results by proving uniform validity over a large class of distributions satisfying a weak uniform integrability condition. Importantly, our results permit degeneracy in the covariance matrix of the limiting distribution, which arises naturally in many applications. In an application to the literature on economic mobility, we find that it is difficult to distinguish between high and low mobility census tracts when correcting for selection. Finally, we demonstrate the practical relevance of our theoretical results through an extensive set of simulations.

*This document greatly benefited from generous comments from Isaiah Andrews, Anna Mikusheva, and Alex Torgovitsky. We are grateful for comments from seminar discussant Sarah Moon, and for discussion from Lanier Benkard, Aditya Bhardwaj, Alex Lehner, Yueran Ma, Jesse Shapiro, and Jann Speiss, as well as for comments by participants at the IAAE Annual Conference 2025 and International Seminar for Selective Inference. Xun Huang provided excellent research assistance. The authors would also like to extend their gratitude to Victor Lima, Kotaro Yoshida and students in the economics undergraduate honors workshop at the University of Chicago and to the social sciences computing services at the University of Chicago. The second author acknowledges support from the National Science Foundation through grant SES-2419008.

1 Introduction

In this paper, we consider the problem of conducting inference on multiple selections, or multiple winners. Here, we define a winner to be a population whose rank according to some random value, such as a measure of value-added or an estimated treatment effect, lies in a set specified by the analyst. We seek to provide joint confidence sets for parameters corresponding to these populations selected according to their ranks. Our framework generalizes the setting of inference on a single winner considered in Andrews et al. (2022) and Andrews et al. (2023), in that we target joint coverage of parameters among multiple selected populations. This generalization allows us to address several important applications, such as inference after cutoff-based selections, inference on quantiles, or inference on statistical significance. These settings arise frequently in applied work. For example, cutoff-based selections arise frequently in the decision-theoretic literature on optimal subset selection, as in Gu and Koenker (2023). Similarly, inference on quantiles, or equivalently inference on the τ-best winners, arises in the applied literature on economic mobility, as in Bergman et al. (2024). Inference on statistical significance is particularly relevant in the literature on publication bias, as in Andrews and Kasy (2019). Our multiple winners framework identifies a common structure in these empirically relevant problems.

Motivated by this setting, we propose a novel, two-step approach to inference. In particular, we identify a key nuisance parameter that characterizes rank-based selection. In the first step of our procedure, we construct confidence bounds for this nuisance parameter. In the second step, we use these confidence bounds to bound the errors on selected units and construct critical values. We apply a Bonferroni-type correction to account for the possibility that our first step confidence region does not cover the key nuisance parameter driving selection. In this way, our approach is most similar to that of Romano et al. (2014), who study inference in moment inequality models. We demonstrate the finite-sample validity of our methods in a normal location model. We then provide results on feasible inference when the data generating process lies in a nonparametric class of distributions. We emphasize that attaining uniform validity is nontrivial in our setting, since the population selections may be indistinguishable from units not selected in a local sense. We demonstrate uniform asymptotic validity under a weak uniform integrability condition on the class of distributions generating the data. Moreover, we show that for each distribution in a large class, our two-step approach to inference asymptotically dominates projection and simultaneous inferences in the spirit of Bachoc et al. (2017), Kuchibhotla et al. (2022), and Berk et al. (2013).

Our approach lies in a broader literature on selective inference. After constructing our two-step approach to inference, we show that some well-known, existing methods for selective inference and their natural generalizations may be ill-suited for inference after selection on ranks. Projection-based methods are robust to arbitrary selection rules, and are therefore underpowered. Tools due to Lee et al. (2016) and Andrews et al. (2023) provide a polyhedral characterization of the rank-based selections considered in this paper. However, this characterization does not provide a computationally feasible inference procedure when joint coverage is desired. Consequently, it is unclear how to generalize the conditional and hybrid approaches of Andrews et al. (2023) to settings where multiple selections are made, further motivating our approach.

Within the selective inference literature, our approach is most similar in spirit to that of Zrnic and

Fithian (2024b), who also apply a Bonferroni-type correction to a general selective inference setting. Similar Bonferroni-type corrections appear in Silvapulle (1996) and McCloskey (2017). As explained further in remark 3.6, the approach in Zrnic and Fithian (2024b) differs from ours, in that they propose projection inference localized to a set of likely selections, whereas our method constructs worst-case critical values over a confidence set for a key nuisance parameter driving selection. We also show analytically that our critical values are, with high probability, smaller than those of Zrnic and Fithian (2024b) for a broad set of data generating processes where both methods perform relatively poorly. The empirical relevance of this result is evident in our extensive simulations comparing the methods. In those simulation results, we also include comparisons with the alternative methodology developed in Zrnic and Fithian (2024a).

We consider two applications. In our first application, we use as an illustrative example a replication failure from two studies on job retraining due to Cave et al. (1993) and Miller et al. (2005). In our second application, we revisit the literature on economic mobility, and the problem of selecting high-opportunity census tracts in the spirit of Bergman et al. (2024). We consider the replication failure between Cave et al. (1993), who estimate treatment effects for an array of job retraining programs, and Miller et al. (2005), who replicate the study of Cave et al. (1993) for a selected subset of interventions. We build on the previous analysis of Andrews et al. (2023) to determine whether correcting for selection bias can explain the replication failure between the two studies. We find that, even when considering alternative selection rules based on statistical significance, a winners' curse fails to explain this replication failure. In our second application, we consider the problem of ex-post inference on tract-level measures of economic mobility, as defined in Chetty et al. (2025), for a subset of high-mobility tracts considered in the randomized trial of Bergman et al. (2024). Our analysis revisits the findings of Mogstad et al. (2023), who suggest that the selection of high-opportunity tracts may reflect noise as opposed to signal. We find similar results to Mogstad et al. (2023), in that we generally fail to reject the null that pairwise differences between arbitrary high and low opportunity tracts are zero. We obtain more positive results, however, in comparing commuting zones.

Finally, we conduct an extensive simulation study comparing our methods to existing methods in the literature, namely projection inference as defined in Andrews et al. (2023), the locally simultaneous approach of Zrnic and Fithian (2024b), and a recent approach due to Zrnic and Fithian (2024a). Our methods are able to outperform these existing methods across a broad range of simulation designs, both in terms of reducing over-coverage and in reducing confidence set length. In particular, our methods reduce over-coverage error by up to 96%, and reduce confidence set length by up to 27% relative to projection inference. We also reduce over-coverage error by up to 71% and length by up to 11% relative to locally simultaneous inference.

The paper is organized as follows: In section 2, we formally introduce the inference on multiple winners problem and present four empirically-relevant settings to which it applies. In section 3, we develop the two-step approach to inference on multiple winners. In section 4, we provide a more detailed discussion of existing approaches to the inference on winners problem, outline some of their shortcomings in the inference on multiple winners problem, and provide some analytical results comparing our two-step approach to inference to these existing methods. In section 5, we revisit the JOBSTART demonstration and compare different approaches to inference. In section 6, we apply our two-step approach to inference to the CMTO program, evaluating neighborhood effects in selected census tracts. Finally, in section 7, we present the results from a simulation study comparing the performance of the two-step and projection approaches to inference in a range of synthetic, simulation designs. In the appendix, we describe further approaches in the literature, provide proofs and supplemental results, and provide a full simulation study.

2 Setup and Notation

In this section, we formalize the inference on multiple winners problem. We consider a finite set of indices or populations $J := \{1, \dots, p\}$, the p-dimensional random vector Y, as well as a correlated random vector X, which are drawn jointly from a multivariate normal as below:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{pmatrix} \right) \tag{1}$$

To simplify notation, we denote the variance-covariance matrix of the above by Σ, and the mean by μ. We also note that Σ is a symmetric, positive semi-definite matrix. We denote the distribution generating this data by $P_{\mu,\Sigma}$.

We observe X and Y, while μ is unknown. The variance-covariance matrix, denoted by Σ, on the other hand, is taken to be known. Our goal is to devise a valid confidence set for the values of μ_Y at indices selected according to the realized values of X. In particular, we take $R \subseteq J$ to be some fixed set of ranks of interest and define the number of selections to be $k := |R|$. We define the rank of population j, according to X, as follows:

$$r_j(X) = \sum_{j' \in J} \mathbb{1}\left(X_{j'} \leq X_j\right) \tag{2}$$

We define the set of selected indices $\hat{J}_R(X)$ as follows:

$$\hat{J}_R(X) := \{j; r_j(X) \in R\} \tag{3}$$

If we take R to simply be J, then $\hat{J}_R(X)$ would be the full index set J. Similarly, taking $R := \{p\}$, our selected indices would correspond to the largest values in X. Unless stated otherwise, we will denote $\hat{J}_R(X)$ by $\hat{J}_R$ in the following analyses. We notice that, even when R is a singleton, $\hat{J}_R(X)$ need not be a singleton according to our definitions.

Our goal is to conduct simultaneous inference on the means $\mu_{Y,\hat{\jmath}}$ for all $\hat{\jmath}$ in $\hat{J}_R$, and in particular, construct a rectangular confidence set CS indexed by $\hat{J}_R$ such that:

$$P_{\mu,\Sigma}\left(\mu_{Y,\hat{\jmath}} \in CS_{\hat{\jmath}} \quad \text{for all} \quad \hat{\jmath} \in \hat{J}_R\right) \geq 1 - \alpha \tag{4}$$

for all μ and Σ. When $R = \{1\}$, this problem is equivalent to the inference on (single) winners problem considered in Andrews et al. (2023). When convenient, we may write $(\mu_{Y,j})_{j \in J_c} \in CS$ to denote $\mu_{Y,j} \in CS_j$ for all $j \in J_c$, where J_c is an arbitrary subset of J.

Remark 2.1. In addition to the case of $R = \{1\}$ considered in Andrews et al. (2023), Andrews et al.

(2022) consider the case where selection occurs with respect to an arbitrary rank, such that R is an arbitrary singleton. However, they do not consider the case where multiple rank-based selections are made. ■

2.1 Review of Applications

Before proceeding, we describe a broad array of empirically-relevant settings in which our results are of interest. In addition to the same applications considered in Andrews et al. (2023), the inference on multiple winners setting accommodates a range of novel applications.

Post-Selection Inference on Quantiles: Suppose we want to conduct inference on the components of the mean of Y corresponding to the components of X in the top γ-quantile of all components of X. We can take $\tau := \lfloor \gamma p \rfloor$ and take $R = \{p - \tau, \dots, p\}$. In other words, we seek to construct confidence sets for the elements in μ_Y corresponding to the τ-best elements in X. This setting naturally arises in our neighborhood effects application in section 6, where we study tract-level outcomes for high opportunity tracts in the setting of Bergman et al. (2024). This specialized setting also arises in Haushofer and Shapiro (2016), who conduct a randomized controlled trial in the top 40% of villages in Rarieda, Kenya, selected according to the proportion of houses with thatched roofs.

Inference After Cutoff-Based Selections: We may want to conduct inference on the values $\mu_{Y,j}$ for j such that $X_j \geq c$ for some non-negative real number c, in the so-called file drawer problem. In order to accommodate this setting in our framework, define $X_c := \begin{pmatrix} X' & c\mathbb{1}_p' \end{pmatrix}'$, and $Y_c := \begin{pmatrix} Y' & c\mathbb{1}_p' \end{pmatrix}'$. We can take $R := \{p+1, \dots, 2p\}$, and consider inference using X_c and Y_c. The indices of the top p elements in X_c correspond exactly to those indices j in X such that $X_j \geq c$, as well as a residual set of indices corresponding (non-uniquely) to elements in the appended constant vector in X_c.[1] For a concrete, empirical example, policymakers may observe the marginal value of public funds (MVPFs) of Hendren and Sprung-Keyser (2020) for a menu of policies. An MVPF exceeding one corresponds to a policy whose benefits, in dollar terms, exceeds its costs. Consequently, policymakers may choose to proceed only with policies whose MVPFs exceed one, generating an inference on multiple winners problem. Our analysis of cutoff-based selections differs from that of Andrews et al. (2023) in that we can accommodate multiple selections. In contrast, the setting of Andrews et al. (2023) can only accommodate the above when $p = 1$. Cutoff-based selection rules have also attracted interest in the literature on optimal selection, namely in Gu and Koenker (2023).

Inference on Statistical Significance: We observe, just as in Andrews et al. (2023), that we can normalize the X_j by standard errors $\sqrt{\Sigma_{X,jj}}$. Since this is simply a linear transformation of the $\begin{pmatrix} X', & Y' \end{pmatrix}'$, the standard inference on multiple winners setting still applies. By applying the previous two examples, we can accommodate inference on all units that are statistically significant at a given level according to X. Note that X and Y need not be the same. Such problems arise in the literature on publication bias, as in Andrews and Kasy (2019).

[1] The non-uniqueness in this setting is inconsequential, since those indices selected non-uniquely share a common, degenerate distribution in Y_c. In particular, the joint distribution of the $(Y_{\hat{\jmath}})_{\hat{\jmath} \in \hat{J}_R}$ is invariant to our choice of $\hat{J}_R$.

Inference on Multiple Outcomes: We may be concerned with the means of several outcomes of interest among selected populations. For example, in the CMTO application of Bergman et al. (2024), we may be concerned with mobility effects among different groups (say effects by race or gender). Denote by $Y_1, \ldots, Y_K$ the different outcomes of interest. In order to accommodate this setting in our framework, define $Y_r := \begin{pmatrix} Y_1' & \ldots & Y_K' \end{pmatrix}'$ and $X_r := \begin{pmatrix} X' & \ldots & X' \end{pmatrix}'$, where X is repeated K times. Further define a new index set $R_r := \{l \cdot q; l \in R,\ 1 \leq q \leq K\}$. The quantities Y_r, X_r, and R_r defined in this way characterize the inference on multiple winners problem with the desired estimands.

Remark 2.2. In many of these examples of applications, Σ need not be full rank. Our methods are valid in finite samples when imposing normality as in (1). Moreover, when deriving asymptotic properties of our methods, we demonstrate we can provide uniformly valid inferences over a class of data generating processes which may included distributions with degenerate covariance matrices. We discuss and provide necessary conditions for asymptotic validity in appendix B. ■

3 A Two-Step Approach to Inference on Multiple Winners

In this section, we construct a two-step approach to inference on multiple winners. We construct a confidence interval that is conceptually similar to the inference approach outlined Romano et al. (2014), and most recently in the selective inference literature, to Zrnic and Fithian (2024b). We emphasize, however, that our approach is meaningfully different from that of Zrnic and Fithian (2024b); see remark 3.6 for further discussion. In the first step, we construct confidence bound on moment differences. In the second step, we use these bounds to model the errors on selected units and derive critical values. As we will show in sections 4 and 7, our approach to inference performs well relative to existing methods.

3.1 Construction

Our goal, as usual, is to construct a rectangular confidence set $CS^{TS}_{1-\alpha}$ satisfying (4), where $CS^{TS}_{1-\alpha} = \times_{j \in \hat{J}_R} CS^{TS}_{1-\alpha,j}$. Our construction is based on the following observation, which states that:

$$\frac{|\xi_{Y,\hat{j}}|}{\sqrt{\Sigma_{Y,\hat{j}\hat{j}}}} \leq \max_{j \in J} \frac{|\xi_{Y,j}|}{\sqrt{\Sigma_{Y,jj}}} \mathbb{1}\left(\sum_{j' \in J} \mathbb{1}\left(\xi_{X,j} \geq \xi_{X,j'} + \Delta_{j'j}\right) \in R\right) \tag{5}$$

for all $\hat{j} \in \hat{J}_R$, where $\Delta_{jj'} := \mu_{X,j} - \mu_{X,j'}$ is an unobserved nuisance parameter. Our goal is therefore to construct a confidence region given by a lower bound L and upper bound U for Δ, such that $P_{\mu,\Sigma}(L \leq \Delta \leq U) \geq 1 - \beta$ for some choice of β in $(0, \alpha)$, where the inequality is interpreted element-wise. We use this lower bound to construct a probabilistic upper bound for the right-hand side of (5). We finally use a Bonferroni-type correction to correct for the possibility that the event $\{L \leq \Delta \leq U\}$ does not hold.

We construct L by first taking $d_{1-\beta}(\Sigma)$ to be the $1-\beta$-quantile of the following:

$$\max_{j,j' \in J, j \neq j', \mathrm{var}_{jj'} \neq 0} \frac{|\xi_{X,j} - \xi_{X,j'}|}{\sqrt{\mathrm{var}_{jj'}}}\ , \tag{6}$$

where $\text{var}_{jj'} := \text{Var}(\xi_{X,j} - \xi_{X,j'})$. We define L and U as follows:

$$L_{jj'} := X_j - X_{j'} - d_{1-\beta}(\Sigma)\sqrt{\text{var}_{jj'}} \tag{7}$$

$$U_{jj'} := X_j - X_{j'} + d_{1-\beta}(\Sigma)\sqrt{\text{var}_{jj'}}\ . \tag{8}$$

It follows that $P_{\mu,\Sigma}(L \leq \Delta \leq U) \geq 1-\beta$. We note that, for each $\hat{\jmath} \in \hat{J}_R$, the following inequality holds:

$$\frac{|\xi_{Y,\hat{\jmath}}|}{\sqrt{\Sigma_{Y,\hat{\jmath}\hat{\jmath}}}} \leq f(L,U) \tag{9}$$

on the event $B := \{L \leq \Delta \leq U\}$, where, for non-random vectors ℓ and u,

$$f(\ell,u) := \max_{j\in J} \frac{|\xi_{Y,j}|}{\sqrt{\Sigma_{Y,jj}}} \mathbb{1}\left(\left[\sum_{j'\in J} \mathbb{1}\left(\xi_{X,j} \geq \xi_{X,j'} + u_{j'j}\right), \sum_{j'\in J} \mathbb{1}\left(\xi_{X,j} \geq \xi_{X,j'} + \ell_{j'j}\right)\right] \cap R \neq \emptyset\right)\ .$$

Denote by $\rho_{1-\alpha+\beta}(L,U)$ the $1-\alpha+\beta$ quantile of the $f(\ell,u)$, evaluated at the random values of L and U, and consider the following, rectangular confidence interval:

$$CS^{TS}_{1-\alpha;\beta} := \times_{\hat{\jmath}\in\hat{J}_R}\left[Y_{\hat{\jmath}} - \rho_{1-\alpha+\beta}(L,U)\sqrt{\Sigma_{Y,\hat{\jmath}\hat{\jmath}}}, Y_{\hat{\jmath}} + \rho_{1-\alpha+\beta}(L,U)\sqrt{\Sigma_{Y,\hat{\jmath}\hat{\jmath}}}\right]$$

We claim the following:

Proposition 3.1. *$CS^{TS}_{1-\alpha;\beta}$ is a valid confidence set at the $(1-\alpha)$-level, such that:*

$$P_{\mu,\Sigma}\left((\mu_{Y,\hat{\jmath}})_{\hat{\jmath}\in\hat{J}_R} \in CS^{TS}_{1-\alpha;\beta}\right) \geq 1-\alpha \tag{10}$$

for all μ and Σ. In addition, it immediately follows that for any $R' \subseteq R$:

$$P_{\mu,\Sigma}\left((\mu_{Y,\hat{\jmath}})_{\hat{\jmath}\in\hat{J}_{R'}} \in CS^{TS}_{1-\alpha;\beta}\right) \geq 1-\alpha \tag{11}$$

for all μ, Σ.

We prove this result in appendix C.

Remark 3.1. In appendix A, we construct a version of our two-step confidence set that is intersected with the projection confidence set for both μ_X and μ_Y, where we formally define the projection confidence set in section 4.2. Informally, we can think of the projection confidence set $CS^P_{1-\alpha}$ as a special case of the two-step confidence set with $\beta = 0$, and with $L = -\infty$ and $U = \infty$, where here $\pm\infty$ are interpreted as vectors. Notably, when $X = Y$ it follows that this intersected two-step confidence set is a subset of projection. This case corresponds to the case where units are selected according to their ranks in Y, or equivalently, where $\hat{J}_R(X) = \hat{J}_R(Y)$. ■

Remark 3.2. In the case where R is of the form $\{p-\tau, \ldots, p\}$, we can simplify our two-step approach to inference. In particular, we can write (5) as follows:

$$\frac{|\xi_{Y,\hat{j}}|}{\sqrt{\Sigma_{Y,\hat{j}\hat{j}}}} \leq \max_{j\in J} \frac{|\xi_{Y,j}|}{\sqrt{\Sigma_{Y,jj}}} \mathbb{1}\left(\sum_{j'\in J} \mathbb{1}\left(\xi_{X,j} \geq \xi_{X,j'} + \Delta_{j'j}\right) \geq p - \tau\right).$$

Taking $\tilde{L}_{jj'} := X_j - X_{j'} - d_{1-\beta}(\Sigma)\sqrt{\text{var}_{jj'}}$, and defining $\tilde{B} := \left\{\tilde{L} \leq \Delta\right\}$, we have that on the event $\tilde{B}$:

$$\frac{|\xi_{Y,\hat{j}}|}{\sqrt{\Sigma_{Y,\hat{j}\hat{j}}}} \leq \tilde{f}\left(\tilde{L}\right),$$

where we define $\tilde{f}$ as follows:

$$\tilde{f}(l) = \max_{j\in J} \frac{|\xi_{Y,j}|}{\sqrt{\Sigma_{Y,jj}}} \mathbb{1}\left(\sum_{j'\in J} \mathbb{1}\left(\xi_{X,j} \geq \xi_{X,j'} + l_{j'j}\right) \geq p - \tau\right). \tag{12}$$

We define $\tilde{\rho}_{1-\alpha+\beta}\left(\tilde{L}\right)$, to be the $1-\alpha+\beta$ quantile of $\tilde{f}$ evaluated at $\tilde{L}$. In this case, we can write:

$$CS^{TS}_{1-\alpha;\beta} := \times_{\hat{j}\in\hat{J}_R} \left[Y_{\hat{j}} - \tilde{\rho}_{1-\alpha+\beta}\left(\tilde{L}\right)\sqrt{\Sigma_{Y,\hat{j}\hat{j}}}, Y_{\hat{j}} + \tilde{\rho}_{1-\alpha+\beta}\left(\tilde{L}\right)\sqrt{\Sigma_{Y,\hat{j}\hat{j}}}\right]$$

giving a simplified two-step approach to inference for the top-τ winners. ■

Remark 3.3. The observation in (5) generalizes to the positive and negative components of the errors, which we may denote by ξ^+_Y and ξ^-_Y, such that for $\hat{j} \in \hat{J}_R$:

$$\frac{\xi^+_{Y,\hat{j}}}{\sqrt{\Sigma_{Y,\hat{j}\hat{j}}}} \leq \max_{j\in J} \frac{\xi^+_{Y,j}}{\sqrt{\Sigma_{Y,jj}}} \mathbb{1}\left(\sum_{j'\in J} \mathbb{1}\left(\xi_{X,j} \geq \xi_{X,j'} + \Delta_{j'j}\right) \in R\right)$$

with the analogous equality holding for ξ^-_Y. Consequently, we can continue as above, constructing L as usual and deriving the same inequality as in (9) with ξ^+_Y and ξ^-_Y in place of $|\xi_Y|$. We can consequently take $\rho^+_{1-\frac{\alpha-\beta}{2}}(L,U)$ and $\rho^-_{1-\frac{\alpha-\beta}{2}}(L,U)$ to be the $\left(1-\frac{\alpha-\beta}{2}\right)$-quantile of the right-hand side of (9) with the same, respective, substitutions. We can consequently take the following, asymmetric two-step confidence set:

$$CS^{ATS}_{1-\alpha} := \times_{\hat{j}\in\hat{J}_R} \left[Y_{\hat{j}} - \rho^+_{1-\frac{\alpha-\beta}{2}}(L,U)\sqrt{\Sigma_{Y,\hat{j}\hat{j}}},\ Y_{\hat{j}} + \rho^-_{1-\frac{\alpha-\beta}{2}}(L,U)\sqrt{\Sigma_{Y,\hat{j}\hat{j}}}\right]$$

This confidence set matches the intuition in Andrews et al. (2023) that the observation corresponding to the winner is upward biased, since $\rho^+_{1-\frac{\alpha-\beta}{2}}(L,U)$ will generally be larger than $\rho^-_{1-\frac{\alpha-\beta}{2}}(L,U)$. ■

Remark 3.4. Our approach to inference can be generalized to arbitrary polyhedral selections, in the spirit of Lee et al. (2016). This generalization is possible because polyhedral selection is linearly separable in errors ξ and means μ, much as in (5). As discussed in section 4, conditional confidence regions are numerically challenging to construct when one observes multiple selections. In particular, let us suppose that we can partition the space of selections $\mathbb{R}^p$ into a finite collection of polyhedra $\mathcal{O} = \{\mathcal{O}_i\}_{i\in\mathcal{I}}$ where we take $\mathcal{O}_i := \{x; A_i x \leq b_i\}$. We take a collection of $k \times p$ matrices B_i such that, whenever we observe X satisfying $X \in \mathcal{O}_i$, we select $B_i Y$. We denote the rows of B_i by $b_{i,j}'$ for $j = 1, \dots, k$. In general, we say that we observe $\hat{B}Y$, where $\hat{B}$ is selected as above. Denoting the rows of $\hat{B}$ by $\hat{b}_j'$ for $j = 1, \dots, k$, we have that

$$\frac{\left|\hat{b}_j' \xi_Y\right|}{\sqrt{\mathrm{Var}\left(\hat{b}_j' \xi_Y\right)}} \leq \max_{i \in \mathcal{I}} \frac{|b_i' \xi_Y|}{\sqrt{\mathrm{Var}\left(b_i' \xi_Y\right)}} \mathbb{1}\left(A_i \xi_X + A_i \mu_X \leq b_i\right) \tag{13}$$

Our nuisance parameter of interest is μ_X, or more tightly, the collection $(A_i \mu_X)_{i \in \mathcal{I}}$. We can proceed as above, constructing $(1-\beta)$-level simultaneous lower confidence bounds for these nuisance parameters, and subsequently constructing the $(1-\alpha+\beta)$-quantile of the right-hand side of (13) with the first step confidence region for the nuisance parameters used to account for the unknown $A_i \mu_X$. Thus, our approach can be used as a more general tool for unconditional post-selection inference, such as in the case of the LASSO. Moreover, a broad range of selection problems analogous to the inference on multiple winners problem discussed in this paper can be cast as polyhedral selection rules. Thus, even though we develop our two-step approach to inference for rank-based selections, our methods generalize to polyhedral selections as well. ■

Remark 3.5. Our two-step approach can also be straightforwardly applied to the problem of inference for pairwise differences. Let us say that we wish to compare units ranked in R with those ranked in R'. We may modify (9) as follows:

$$\begin{aligned} &\frac{|\xi_{Y,\hat{\jmath}} - \xi_{Y,\hat{\imath}}|}{\sqrt{\Sigma_{Y,\hat{\jmath}\hat{\jmath}} + \Sigma_{Y,\hat{\imath}\hat{\imath}} - 2\Sigma_{Y,\hat{\jmath}\hat{\imath}}}} \\ &\leq \max_{i,j \in J, i \neq j} \frac{|\xi_{Y,j} - \xi_{Y,i}|}{\sqrt{\Sigma_{Y,jj} + \Sigma_{Y,ii} - 2\Sigma_{Y,ij}}} \mathbb{1}\left(\sum_{j' \in J} \mathbb{1}\left(\xi_{X,j} \geq \xi_{X,j'} + \Delta_{j'j}\right) \in R, \ \sum_{i' \in J} \mathbb{1}\left(\xi_{X,i} \geq \xi_{X,i'} + \Delta_{i'i}\right) \in R'\right) \end{aligned}$$

It is clear that our two-step technique can be applied to study the pairwise differences $\mu_{Y,\hat{\jmath}} - \mu_{Y,\hat{\imath}}$ for all pairs of $\hat{\jmath}$ in $\hat{J}_R$ and $\hat{\imath}$ in $\hat{J}_{R'}$. Moreover, as Mogstad et al. (2023) demonstrate, confidence regions for pairwise differences in unit means can be used to construct confidence sets for the ranks of these units. Thus, our methods can be applied to conduct inference for the ranks of selected units among all units, for example, or the ranks of selected units only among selected units. We leave this problem of inference for selected ranks to future research. ■

Remark 3.6. Our approach to inference is most similar to the locally simultaneous approach of Zrnic and Fithian (2024b), who suggest focusing power by localizing simultaneous inference to a set of likely selections. In particular, Zrnic and Fithian (2024b) consider, as a nuisance parameter, some particular non-random subset $\tilde{J}_R$ of J such that $\hat{J}_R$ lies in this subset with pre-specified probability, and provide the key insight that an outer confidence region $\hat{J}_R^+$ for this subset can be constructed such that $\tilde{J}_R \subset \hat{J}_R^+$ if and only if $\hat{J}_R \subseteq \tilde{J}_R$. Using this insight, they use a Bonferroni-type correction to localize simultaneous inference to $\hat{J}_R^+$ and obtain valid coverage. Our approach is similar to that of Zrnic and Fithian (2024b), in that we too construct confidence bounds for a nuisance parameter, derive worst-case critical values over this confidence set, and apply a union bound to recover valid coverage. However, our approach builds on the insight that, when considering ranked-based selections, we can model the errors on selected units more explicitly by taking pairwise differences in means as our nuisance parameter. This insight allows us to provide asymmetric confidence sets as discussed in remark 3.3. Moreover, because our critical values $\rho_{1-\alpha+\beta}(L, U)$ vary smoothly in the confidence bounds in L and U, we are able to achieve power gains relative to existing approaches, as demonstrated in section 7. We demonstrate this formally in proposition 4.5, where we show that our two-step critical values are strictly smaller than the corresponding locally-simultaneous critical values with probability at least $1-\beta$, for a broad range of data generating processes where the units in $J_R(\mu_X)$ are not well-separated from those in $J_R(\mu_X)^c$. We provide further discussion of the approach of Zrnic and Fithian (2024b) in section 4.4. ■

3.2 Feasible Inference

In this section, we provide uniform asymptotic guarantees for our methods over a nonparametric class and compare our two-step approach with the projection approach to inference, defined in section 4.2, analytically.

First, we provide notation for our asymptotic setting. We assume that we observe a sequence of random vectors $\tilde{W}_i \equiv \begin{pmatrix} \tilde{X}_i' & \tilde{Y}_i' \end{pmatrix}'$ for $i = 1, \ldots, n$ drawn i.i.d. from some distribution P in a nonparametric class of distributions $\mathcal{P}$. We denote the sample mean of the above sequence of random vectors by $\widetilde{S}_W^n := \begin{pmatrix} \widetilde{S}_X^n{}' & \widetilde{S}_Y^n{}' \end{pmatrix}'$. For all P in $\mathcal{P}$, we define

$$\mu_X(P) := \mathbb{E}_P\left(\tilde{X}_i\right) ,$$

$\mu_Y(P) := \mathbb{E}_P\left(\tilde{Y}_i\right)$ and $\mu_W(P) = (\mu_X(P)', \mu_Y(P)')'$. Let $\xi_{\tilde{X}_i}$ denote the demeaned version of $\tilde{X}_i$, and let us similarly denote the demeaned version of $\tilde{Y}_i$ by $\xi_{\tilde{Y}_i}$. We define the variance-covariance matrix as follows:

$$\Sigma_W(P) := \mathbb{E}_P\left(\begin{pmatrix} \xi_{\tilde{X}_i} \\ \xi_{\tilde{Y}_i} \end{pmatrix}\begin{pmatrix} \xi_{\tilde{X}_i}{}' & \xi_{\tilde{Y}_i}{}' \end{pmatrix}\right) \tag{14}$$

We similarly have some sequence of estimators for $\Sigma_W(P)$, which we denote by $\hat{\Sigma}^n$. In general, we can take $n\hat{\Sigma}^n = \Sigma_W(\hat{P}_n)$, where $\hat{P}_n$ denotes the empirical distribution of the data $\begin{pmatrix} \tilde{X}_i' & \tilde{Y}_i' \end{pmatrix}'$ for $i = 1, \ldots, n$.

As usual, we index $\tilde{X}_i$ and $\tilde{Y}_i$ by $J := \{1, \ldots, p\}$. We take $R \subseteq J$ to be a set of ranks of interest. We take $\hat{J}_{R;n}$ to be the set of indices given by $\hat{J}_R(\tilde{S}_{X,n})$. We denote by $CS^{TS}_{1-\alpha;\beta,n}$ the two-step confidence set using $\tilde{S}_W^n$ in lieu of the usual X and Y, and with $\hat{\Sigma}^n$ in lieu of the known Σ. We define $CS^{P}_{1-\alpha;n}$ analogously. Our first goal is to demonstrate the following uniform asymptotic validity result:

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \Pr_P\left((\mu_{Y,\hat{\jmath}}(P))_{\hat{\jmath}\in\hat{J}_{R;n}} \in CS^{TS}_{1-\alpha;\beta,n}\right) \geq 1-\alpha \tag{15}$$

We formalize uniform asymptotic validity and prove that our two-step approach to inference is uniformly, asymptotically valid in appendix B. To do so, we make one very mild assumption on $\mathcal{P}$. In particular, we impose a uniform integrability assumption, as in Romano and Shaikh (2008), which implies a uniform central limit theorem that allows us to prove uniform asymptotic validity. We formally state this assumption in assumption B.1. Under this assumption, we have that the following uniform asymptotic validity result holds:

Proposition 3.2. *Given assumption B.1, the two-step confidence set is uniformly, asymptotically valid,* *such that:*

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \Pr_P\left((\mu_{Y,\hat{\jmath}}(P))_{\hat{\jmath}\in\hat{J}_{R;n}} \in CS^{TS}_{1-\alpha;\beta,n} \right) \geq 1-\alpha \tag{16}$$

The proof of this proposition is contained in appendix B.

4 Existing Approaches to Inference

We now turn our attention to existing approaches to post-selection inference. We discuss into three classes of inference methods: conditional, projection, and hybrid. Kuchibhotla et al. (2022) provide a review of the former two approaches to inference. The conditional approach generally involves conducting inference conditional on the selection event, as in Lee et al. (2016), Markovic et al. (2018), and McCloskey (2023). The projection approach involves conducting inference that is simultaneously valid for all selection methods, as in Bachoc et al. (2017) and Berk et al. (2013). In this section, we demonstrate that the natural generalization of the conditional approach to the inference on multiple winners setting leads to computational difficulties. As a result conditional inference is too cumbersome for many applications, while the projection approach tends to be too conservative. We further explain that the hybrid approach, which is closely related to the conditional approach, inherits the same computational shortcomings. Finally, we discuss the locally simultaneous approach of Zrnic and Fithian (2024b), and provide an analytical result comparing locally simultaneous inference with our two-step approach. In appendix A, we discuss further, alternative approaches to inference, namely approaches based on inverting the zoom test of Zrnic and Fithian (2024a) and a few extensions of our two-step method. In general, we find that our two-step approach outperforms existing approaches in simulation for a broad range of data generating processes, which we demonstrate in section 7. Moreover, as we will show in proposition 4.2, our method reduces over-coverage asymptotically relative to projection, and as we will show in proposition 4.5, our two-step critical values are smaller than locally-simultaneous critical values when all populations are sufficiently close.

4.1 The Conditional Approach to Inference

The conditional inference approach outlined in Andrews et al. (2023) is an example of conditional selective, notably studied in Lee et al. (2016). In this section, we outline a generalization of the conditional approach from Andrews et al. (2023) which accounts for multiple selections. We find that the exact distribution of the collection of the $(Y_{\hat{\jmath}})_{\hat{\jmath}\in\hat{J}_R}$, conditional on the selection event for the $\hat{J}_R$, is a multivariate normal truncated to a union of convex polyhedra. Unfortunately, due to computational challenges, this characterization does not lead to a practical inference procedure for this problem.

In our analysis of conditional and hybrid inference, we will neglect ties and assume that $\hat{J}_R$ can be written as $(j_l)_{l\in R}$, such that $r_{j_l}(X) = l$. Our general approach to conditional inference involves conditioning on the selection event $j_l = i_l$ for l in R, where the $(i_l)_{l\in R}$ is a fixed family of indices. We additionally condition on a sufficient statistic Z for the nuisance parameters associated with the elements of μ not corresponding to the Y_{j_l}. Our goal is to derive a confidence set $CS^c_{1-\alpha}$ such that:

$$P_{\mu,\Sigma}\left(\mu_{Y,\hat{\jmath}} \in CS^c_{1-\alpha,\hat{\jmath}} \quad \text{for all} \quad \hat{\jmath} \in \hat{J}_R \;\middle|\; r_{\hat{\jmath}}(X) = r_{\hat{\jmath}},\ z\right) \geq 1-\alpha \tag{17}$$

for all μ, where $r_{\hat{\jmath}}$ corresponds to the observed rank attained by unit $\hat{\jmath}$. We can equivalently write this conditioning set as $j_l = i_l$ for all l in R, under unique selection. We will use this notation in the remainder of our discussion of conditional and hybrid inference. By the law of iterated expectations, such a confidence set satisfies (4) as well. To this end, we first derive an extension of the polyhedral selection lemma in Lee et al. (2016); see, in particular, lemma 5.1 therein. In order to describe our generalization of this result, we require some further notation. Let e_i be the standard basis vectors in $\mathbb{R}^p$ and, for ease of exposition, let $R := \{1, \ldots, \tau\}$.[2] Define

$$B' := \begin{pmatrix} & e'_{i_1} \\ & e'_{i_2} \\ \mathbf{0}_{\tau\times p} & \vdots \\ & e'_{i_\tau} \end{pmatrix}, \quad A := \begin{pmatrix} (e'_{i_1} - e'_i)_{i\neq i_1} & \mathbf{0}_{p-1\times p} \\ (e'_{i_2} - e'_i)_{i\neq i_1,i_2} & \mathbf{0}_{p-2\times p} \\ \vdots & \vdots \\ (e'_{i_\tau} - e'_i)_{i\neq i_1,\ldots,i_\tau} & \mathbf{0}_{p-\tau\times p} \end{pmatrix},$$

and $c := \Sigma B(B'\Sigma B)^{-1}$. Here, $(e'_{i_1} - e'_i)_{i\neq i_1}$ denotes stacking the vectors $e'_{i_1} - e'_i$ for all $i \neq i_1$. This selection event equals

$$A\begin{pmatrix} X \\ Y \end{pmatrix} \geq 0 \,.$$

Using this notation, we have the following result:

Lemma 4.1. *For, $Z := (I - cB')$, $(Y_{\hat{\jmath}})_{\hat{\jmath}\in\hat{J}_R}$, conditional on $Z = z$ and $j_l = i_l$ for l in R, is distributed according to a multivariate normal with mean $B'\mu$ and variance-covariance $B'\Sigma B$ truncated to the polyhedron $\mathcal{Y}((i_l)_{l\in R}, z) := \{x; (Ac)x \geq -Az\}$. We write:*

$$(Y_{\hat{\jmath}})_{\hat{\jmath}\in\hat{J}_R} | Z = z, j_l = i_l \textit{ for } l \in R \sim TN_{\mathcal{Y}((i_l)_{l\in R},z)}(B'\mu, B'\Sigma B) \tag{18}$$

Given $Z = z$ and $Y_{j_l} = y_{j_l}$ for l in R, consider the problem of construct a test of the null:

$$H_{0,m} : (\mu_{Y,i_l})_{l\in R} = (m_{Y,i_l})_{l\in R} \tag{19}$$

conditional on $Z = z, j_l = i_l$ for all $l \in R$. We suggest the following test of the null in (19):

$$\phi\left((y_{i_l})_{l\in R}; (i_l)_{l\in R}, z, (m_{Y,i_l})_{l\in R}\right)$$

[2] When R is not of the form $\{1, \ldots, \tau\}$, we can take the selection event to be a union of polyhedra. In particular, we take $\tau := \max R$, and take $R' := \{p - \tau, \ldots, p\}$. We condition on $j_l = i_l$ for l in R', but take the union of polyhedral conditioning sets over possible i_l. In particular, we take i_l for l in R to be fixed, and take the union of polyhedral conditioning sets where we vary $i_{l'}$ for l' in $R' \setminus R$, such that $i'_l \neq i_l$ for any l in R. Details on such unions of polyhedral selection events for ranked objects are contained in Andrews et al. (2022). However, in our setting, it is unclear how to derive a parsimonious characterization of the union of polyhedra in the spirit of algorithm 1 in Andrews et al. (2022). It is worth noting that conditional inference is especially challenging for such R, since the number of polyhedra over which we take a union can be very large, leading to well-known numerical integration issues, as per the discussion in Lee et al. (2016).

$$= \begin{cases} 1 & \text{if} \quad c_{TN,\alpha}((i_l)_{l\in R}, z, (m_{Y,i_l})_{l\in R}) \geq \left\|(y_{i_l} - m_{Y,i_l})_{l\in R}\right\| \\ 0 & \text{otherwise} \end{cases},$$

where $c_{TN,\alpha}((i_l)_{l\in R}, z, (m_{Y,i_l})_{l\in R})$ is the smallest value of x satisfying:

$$P_{m,B'\Sigma B}\left(\left\|(Y_{j_l} - m_{Y,j_l})_{l\in R}\right\| \leq x | Z = z, j_l = i_l \text{ for all } l \in R\right) \geq 1 - \alpha$$

and where the probability is calculated using the distribution in lemma 4.1.[3] The above test ϕ tests the null hypothesis that the true means of the $(Y_{i_l})_{l\in R}$ are each equal to the hypothesized values $(m_{Y,i_l})_{l\in R}$ at level α. We can therefore construct a confidence set for the $(\mu_{Y,i_l})_{l\in R}$ by inverting ϕ. For the resulting confidence set $CS^c_{1-\alpha}$, the following proposition holds:

Proposition 4.1. *For α in $(0,1)$, $CS^c_{1-\alpha}$ is a valid confidence set at level α. That is, (17) holds.*

We provide a proof in Appendix C. Now, we provide the following remark on the feasibility of conditional inference:

Remark 4.1. Andrews et al. (2023) show that, in the $k = 1$ case, $Y_{\hat{\jmath}}$ is distributed according to a univariate normal truncated to an interval for $\hat{\jmath}$ in $\hat{J}_{\{1\}}$. Given this, the test inversion procedure described above is quite tractable, since the test statistic used to compute ϕ can be easily computed using the cumulative distribution function of a univariate truncated normal. However, in the multivariate case, computing these test statistics over a high-dimensional grid for the values of $(\mu_{Y,i_l})_{l\in R}$ is a challenging numerical integration problem.[4] ■

4.2 The Projection Approach to Inference

The projection approach as described in Andrews et al. (2023) is an example of the simultaneous approach to post-selection inference, as outlined in Bachoc et al. (2017), Kuchibhotla et al. (2022), and Berk et al. (2013). Such simultaneous inference methods are appropriate for the inference on multiple winners setting, albeit rather conservative, since they are robust to arbitrary selection rules.

We take, for X and Y, $\xi_X \overset{d}{=} X - \mu_X$ and $\xi_Y \overset{d}{=} Y - \mu_Y$, where $\overset{d}{=}$ denotes equality in distribution. For any subset J_c of J, we define $c_{1-\alpha}(J_c)$ to be the $1-\alpha$-quantile:

$$\max_{j\in J_c} \frac{|\xi_{Y,j}|}{\sqrt{\Sigma_{Y,jj}}} \tag{20}$$

noting, trivially, that, for all $\hat{\jmath} \in \hat{J}$:

$$\frac{|\xi_{Y,\hat{\jmath}}|}{\sqrt{\Sigma_{Y,\hat{\jmath}\hat{\jmath}}}} \leq \max_{j\in J} \frac{|\xi_{Y,j}|}{\sqrt{\Sigma_{Y,jj}}}$$

[3] It is worth noting that in our setting $B'\mu$ equals the $(\mu_{Y,i_l})_{l\in R}$.

[4] Motivated by such concerns, Liu (2023) applies the separation of variables technique from Genz (1992) to problems of polyhedral selection to achieve performance gains in the numerical integration of multivariate Gaussians over polyhedra, but the issue of test-inversion on a potentially high-dimensional grid remains.

Unless stated otherwise, we will denote $c_{1-\alpha}(J)$ by $c_{1-\alpha}$. Consequently, we can define the following rectangular confidence set, based on the projection approach from Andrews et al. (2023) and the simultaneous approaches outlined in the post-selection inference literature:

$$CS^P_{1-\alpha} := \times_{\hat{\jmath}\in\hat{J}_R}\left[Y_{\hat{\jmath}} - c_{1-\alpha}\sqrt{\Sigma_{Y,\hat{\jmath}\hat{\jmath}}}, Y_{\hat{\jmath}} + c_{1-\alpha}\sqrt{\Sigma_{Y,\hat{\jmath}\hat{\jmath}}}\right]$$

Naturally, $CS^P_{1-\alpha}$ is a valid confidence set at the $(1-\alpha)$-level satisfying (4). This modified, projection confidence set can be easily constructed and has very simple statistical properties. However, it is generally quite conservative, particularly in cases where there may exist clear winners for all j.

To see this analytically, let us recall the notation from section 3.2. Since $\tilde{S}^n_W$ converges weakly to $\mu_W(P)$, we can show that any $\hat{\jmath}$ in $\hat{J}_R$ will lie in $J_R(P)$ with probability approaching one, where we define $J_R(P)$ to be the set of asymptotic winners $\left\{j; \sum_{j'\in J}\mathbb{1}\left(\mu_{X,j'}(P) \le \mu_{X,j}(P)\right) \in R\right\}$. This observation underlies the following result on pointwise asymptotic superiority of the two-step method relative to projection, where we construct projection confidence sets in section 4.2.

Proposition 4.2. *Suppose that, for P in $\mathcal{P}$, the set of true winners $J_R(P)$ is a proper subset of J. We also assume that assumption B.1 holds on $\mathcal{P}$ and that $\Sigma(P)$ is full rank. Under these assumptions, the coverage probability of the two-step confidence set is pointwise, asymptotically smaller than that of the projection confidence set, in the sense that for any β sufficiently small, we have:*

$$\lim_{n\to\infty}\Pr_P\left(\left(\mu_{Y,\hat{\jmath}}(P)\right)_{\hat{\jmath}\in\hat{J}_{R;n}} \in CS^{TS}_{1-\alpha;\beta,n}\right) \le \lim_{n\to\infty}\Pr_P\left(\left(\mu_{Y,\hat{\jmath}}(P)\right)_{\hat{\jmath}\in\hat{J}_{R;n}} \in CS^P_{1-\alpha;n}\right) \tag{21}$$

We provide the proof of this proposition in appendix C.

4.3 The Hybrid Approach to Inference

McCloskey (2023) and Andrews et al. (2023) suggest an approach to selective inference related to both the conditional approaches described above. Again, ignoring ties, we assume that $\hat{J}_R$ can be written as $(j_l)_{l\in R}$, such that $r_{j_l}(X) = l$. In particular, McCloskey (2023) and Andrews et al. (2023) suggest conditioning not only the selection event, but on the event that $(\mu_{Y,j})_{j\in J} \in CS^P_{1-\beta}$ for β in $(0,\alpha)$. In particular, in our setting, one potential approach to hybrid inference would involve inverting tests of the null (19) conditional on $Z = z, j_l = i_l,\ m_{Y,i_l} \in CS^P_{1-\beta,i_l}$. We suggest using a test of the following form:

$$\begin{aligned}&\phi\left((y_{i_l})_{l\in R}; (i_l)_{l\in R}, z, (m_{Y,i_l})_{l\in R}\right)\\ &= \begin{cases} 1 & \text{if} \quad c_{TN,\frac{\alpha-\beta}{1-\beta}}((i_l)_{l\in R}, z, (m_{Y,i_l})_{l\in R}, CS^P_{1-\beta}) \ge \left\|(y_{i_l} - m_{Y,i_l})_{l\in R}\right\| \\ 0 & \text{otherwise}\end{cases},\end{aligned}$$

where $c_{TN,\frac{\alpha-\beta}{1-\beta}}((i_l)_{l\in R}, z, (m_{Y,i_l})_{l\in R}, CS^P_{1-\beta})$ denotes the smallest value of x satisfying:

$$P_{m,B'\Sigma B}\left(\left\|(Y_{j_l} - m_{Y,j_l})_{l\in R}\right\| \le x | Z = z, j_l = i_l,\ m_{Y,i_l} \in CS^P_{1-\beta,i_l} \text{ for all } l \in R\right) \ge \frac{1-\alpha}{1-\beta}.$$

Inverting this test provides the hybrid confidence set $CS^H_{1-\alpha;\beta}$, such that the following proposition holds.

Proposition 4.3. *For α in $(0,1)$ and β in $(0,\alpha)$, $CS^H_{1-\alpha;\beta}$ is an unconditionally valid confidence set at level α, such that:*

$$P_{\mu,\Sigma}\left((\mu_{Y,\hat{j}})_{\hat{j}\in\hat{J}_R} \in CS^H_{1-\alpha;\beta}\right) \geq 1-\alpha$$

To see that validity holds, it suffices to notice that:

$$\begin{aligned}
&P_{\mu,\Sigma}\left((\mu_{Y,\hat{j}})_{\hat{j}\in\hat{J}_R} \in CS^H_{1-\alpha;\beta}\right)\\
&\geq P_{\mu,\Sigma}\left((\mu_{Y,\hat{j}})_{\hat{j}\in\hat{J}_R} \in CS^H_{1-\alpha;\beta} | (\mu_{Y,\hat{j}})_{\hat{j}\in\hat{J}_R} \in CS^P_{1-\beta}\right) P_{\mu,\Sigma}\left((\mu_{Y,\hat{j}})_{\hat{j}\in\hat{J}_R} \in CS^P_{1-\beta}\right)\\
&\geq \frac{1-\alpha}{1-\beta}(1-\beta) = 1-\alpha
\end{aligned}$$

Because hybrid inference involves conditioning in the construction of a test ϕ, it is subject to the same computational concerns as conditional inference in our generalized setting. In cases where hybrid inference easily applies, as in the setting of Andrews et al. (2023), we observe that any confidence set satisfying (4) can be used in lieu of projection, including our proposed two-step approach to inference.

4.4 Locally Simultaneous Inference

We now adapt the locally simultaneous approach of Zrnic and Fithian (2024b) to the inference on multiple winners problem. As before, Zrnic and Fithian (2024b) consider some unknown set of likely selections $\tilde{J}_R$ and conduct simultaneous inference restricted to an outer confidence set for $\tilde{J}_R$, which we denote by $\hat{J}^+_R$. In general, Zrnic and Fithian (2024b) construct this $\hat{J}^+_R$ by first constructing a $1-\beta$ confidence region for the data generating process P. In our case, we can equivalently construct a $1-\beta$ confidence region for μ, or some other carefully-chosen parameter of interest. Subsequently, for each P in the confidence region described above, Zrnic and Fithian (2024b) derive a $(1-\beta)$-forecast set for the observations generated by this P, and the selections these observations imply. Taking the union of these forecast sets over all P in the aforementioned confidence set yields a forecast set for the selection which is valid at level $1-\beta$. This forecast set will provide the $\hat{J}^+_R$ described above.

For the inference on multiple winners problem, we take:

$$\hat{J}^+_R := \left\{ j; X_j \notin \bigcup_{\hat{j}\in\hat{J}_R} \left[X_{\hat{j}} - 2\bar{d}_{1-\beta}(\Sigma),\ X_{\hat{j}} + 2\bar{d}_{1-\beta}(\Sigma)\right] \right\} \tag{22}$$

where $\bar{d}_{1-\beta}(\Sigma)$ denotes the $1-\beta$ quantile of:

$$\max_{j,j'\in J, j\neq j'} |\xi_{X,j} - \xi_{X,j'}|$$

We may notice that $\bar{d}_{1-\beta}(\Sigma)$ is a version of $d_{1-\beta}(\Sigma)$ without studentization. Finally, we may define the following confidence set:

$$CS^{LS}_{1-\alpha;\beta} := \times_{\hat{\jmath}\in\hat{J}_R} \left[Y_{\hat{\jmath}} - c_{1-(\alpha-\beta)}\left(\hat{J}^+_R\right)\sqrt{\Sigma_{Y,\hat{\jmath}\hat{\jmath}}},\ Y_{\hat{\jmath}} + c_{1-(\alpha-\beta)}\left(\hat{J}^+_R\right)\sqrt{\Sigma_{Y,\hat{\jmath}\hat{\jmath}}}\right]$$

It is worth noting that if R is of the form $\{p-\tau,\dots,p\}$, we can replace the two-sided interval in (22) with a one-sided interval, as in remark 3.2. As before, we have the following proposition:

Proposition 4.4. *$CS^{LS}_{1-\alpha}$ is a valid confidence set at the $1-\alpha$-level, such that:*

$$P_{\mu,\Sigma}\left((\mu_{Y,\hat{\jmath}})_{\hat{\jmath}\in\hat{J}_R} \in CS^{LS}_{1-\alpha;\beta}\right) \geq 1-\alpha \tag{23}$$

for all μ and Σ. Again, marginal validity clearly follows.

A proof of this proposition is provided in appendix C. We also note that theorem 2 of Zrnic and Fithian (2024b) gives that $CS^{LS}_{1-\alpha;\beta}$ may be intersected with a version of the level $1-\alpha$ projection confidence set, delivering a finite-sample, non-inferiority result for their methods. Comparing the non-intersected versions of two-step and locally-simultaneous inferences yields the following result, where we take $\bar{d}_{1-\beta}(\Sigma)$ in lieu of $d_{1-\beta}(\Sigma)\sqrt{\text{var}_{jj'}}$ in (7):

Proposition 4.5. *Let $R=\{1\}$, and let μ be any mean vector such that:*

$$\max_{j,j'\in J} |\mu_{X,j} - \mu_{X,j'}| \leq \bar{d}_{1-\beta}(\Sigma) \tag{24}$$

It follows that, with probability at least $1-\beta$, $\tilde{\rho}_{1-\alpha+\beta}(L) < c_{1-\alpha+\beta}\left(\hat{J}^+_{\{1\}}\right)$.

We provide a proof of this result in appendix C. It follows that, even when intersecting with projection as in our proposition A.1 and as in theorem 2 of Zrnic and Fithian (2024b), our two-step confidence sets are contained by the corresponding locally-simultaneous confidence sets with probability at least $1-\beta$ whenever the condition in (24) is satisfied. Our simulations suggest that our two-step critical values are smaller than locally simultaneous critical values for a broad range of designs where (24) fails.

4.5 Further Approaches to Inference

Our review of existing approaches has thus far been limited to conditional, projection, hybrid, and locally simultaneous procedures. A recent approach due to Zrnic and Fithian (2024a) provides an unconditional approach to inference that recovers, or nearly recovers, uncorrected inference when the set of winners is clear. In particular, Zrnic and Fithian (2024a) construct tests of point hypotheses of the form $H_0 : \mathbb{E}(Y) = \mu_Y,\ \mathbb{E}(X) = \mu_X$. They provide an acceptance region that is largest for populations that are unlikely to be selected. We describe this approach formally in appendix A, and include these approaches in our extended simulation study in appendix D which reflect the superiority result discussed above. In appendix A, we also provide some further approaches that extend our two-step approach to inference, or that apply our observation in (5) to construct confidence sets in the spirit of Zrnic and Fithian (2024b).

5 Application: the JOBSTART Demonstration

In this section, we revisit the JOBSTART demonstration due to Cave et al. (1993) and subsequent replication failure in Miller et al. (2005), which has been previously studied in the selective inference literature by Andrews et al. (2023). The JOBSTART demonstration was a randomized controlled trial taking place between 1985 and 1988 across 13 sites, with the intention of studying the effects of a vocational training program on the employment outcomes of young, low-skilled school dropouts. The treatment group was given access to a suite of JOBSTART services which were inaccessible to those in the control group. Among the sites included in the JOBSTART study, only one site, the Center for Employment Training (CET) in San Jose, saw a large and statistically significant estimate of the effect on earnings. Cave et al. (1993) note that they cannot attribute the unique success of CET to a particular feature of the program, but suggest that the CET's strong connections with San Jose employers, or their robust placement efforts, may explain some of its value-add. Motivated by the success of the CET program, Miller et al. (2005) replicate the intervention at 12 sites. They find that, even in replication sites deemed to have high fidelity to the original CET program of Cave et al. (1993), the estimated effect of the program's services on enrollees' earnings was not statistically significant.

Andrews et al. (2023) study the possibility that this replication failure is due to a winner's curse. In particular, they consider the possibility that the estimates of the effect of CET on earnings in the original JOBSTART demonstration are upwardly biased by virtue of CET being the site with the largest estimated effect. We consider a complementary thought experiment. In particular, since the exact mechanism by which Miller et al. (2005) select a program for replication is unknown, we consider the possibility that the replication failure between the two studies can be explained by an alternative selection rule in which the sites selected by Miller et al. (2005) are chosen according to a statistical significance cutoff. Under this selection rule, it is impossible for the analyst to know ex-ante how many sites will be selected for replication. When multiple selections are made, the conditional and hybrid approaches of Andrews et al. (2023) are difficult to apply, as we discuss in section 4.1.[5]

Empirically, we show that our two-step confidence regions for the effect of the CET program on earnings exclude zero. This finding suggests that a winner's curse does not fully explain the replication failure between the studies of Cave et al. (1993) and Miller et al. (2005), even under an alternative selection rule. Finally, motivated by concerns that conditional inference may be most appropriate when one selection is made, we illustrate the frequency with which multiple sites may be selected for replication under a statistical significance cutoff rule. In simulations calibrated to the JOBSTART demonstration, we find that multiple sites can be selected as statistically significant with a mean probability of between 17.6% and 53.4%, depending on our choice of significance level by which to select sites for replication.

[5] It is not difficult to provide conditional and hybrid confidence sets that apply in the event that only one selection is made via a particular choice of conditioning set. Such concerns are discussed in appendix C of Andrews et al. (2023), but are beyond the scope of our analysis here.

5.1 JOBSTART: Empirical Findings

In the table below, we report the effects of the JOBSTART intervention on earnings at each of these 13 sites, with point estimates due to Cave et al. (1993) and standard errors due to Andrews et al. (2023):

Table 1: Treatment Effects in the JOBSTART Demonstration

Intervention	Treatment Effect	Standard Error
Atlanta Job Corps	2093	2288.40
CET/San Jose	6547***	1496.17
Chicago Commons	-1417	2168.21
Connelley (Pittsburgh)	785	1681.92
East LA Skills Center	1343	1735.51
EGOS (Denver)	401	1329.05
Phoenix Job Corps	-1325	1598.03
SET/Corpus Christi	485	971.05
El Centro (Dallas)	336	1523.33
LA Job Corps	-121	1409.79
Allentown (Buffalo)	904	1814.10
BSA (New York City)	1424	1768.44
CREC (Hartford)	-1370	1860.45

Estimated program treatment effects from the JOBSTART demonstration, as reported in Cave et al. (1993). The reported standard errors are those derived in Andrews et al. (2023).

In what follows, we denote the estimated ATEs for the thirteen interventions by Y, and index Y by j in the set of interventions J. Of all the interventions from the JOBSTART demonstration, only the CET program had a statistically significant treatment effect on earnings at the 1% level.[6] In the thought experiment where Miller et al. (2005) select statistically significant programs for replication rather than the winner, we provide confidence sets for the true treatment effects at selected sites.

We present our empirical findings using both the symmetric and asymmetric versions of our novel two-step method. We also provide a conditional confidence set under the winner selection rule from Andrews et al. (2023). We emphasize, however, that the winner selection rule leads to a different parameter and notion of coverage, so conditional inference is not directly comparable in this setting.

[6] Or, similarly, at the 5% level.

Table 2: Corrected Confidence Regions for CET Treatment Effects

Method	CS: 5% Significance Cutoff
two-step	[$2476, $10618]
two-step (Asymmetric)	[$2191, $10114]
Original Conditional	[$3485, $9478]

Confidence sets for the CET program, correcting for cutoff-based selection. Confidence sets are presented for a 5% significance cutoff. The conditional confidence set presented assumes the original, best-treatment selection rule considered in Andrews et al. (2023).

We now present results from simulations calibrated to the JOBSTART demonstration. We demonstrate that the probability of making multiple selections when using a cutoff rule is non-negligible. In such instances, conditional and hybrid confidence sets do not readily apply, as explained in section 4. In the simulation presented below, we compute the probability of making multiple selections for 1000 simulation draws. We find that in 90% of our simulations, when choosing a 1% significance cutoff rule for selection, the probability of making multiple selections lies between 0.153 and 0.828. When choosing a 5% significance cutoff, this range becomes the interval $[0.473, 0.977]$. We present a histogram of these probabilities, over a confidence region for the means μ given our observed data, below:

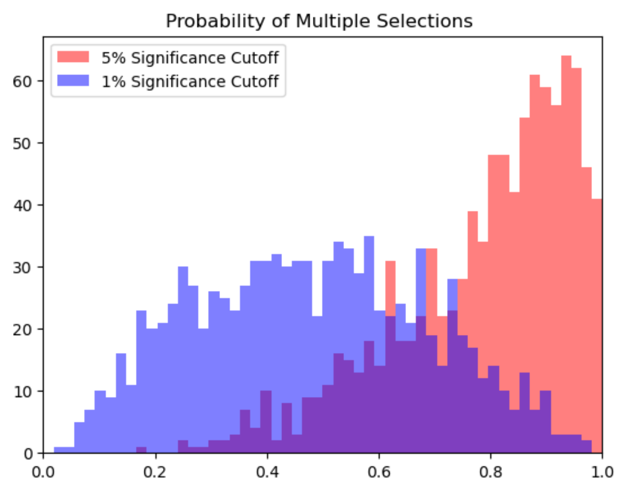


Figure 1: Probability of making multiple selections for different significance-based selection criteria. Probabilities are computed over 1000 simulation draws.

6 Application: Neighborhood Effects Revisited

In this section, we revisit the studies of Chetty et al. (2025) and Bergman et al. (2024) on the geographic nature of economic mobility. These studies have garnered substantial interest in the selective inference and multiple testing literatures, notably in Andrews et al. (2023) and Mogstad et al. (2023). Chetty et al. (2025) construct the Opportunity Atlas, a dataset containing correlational estimates of the effect of childhood neighborhood in adulthood. In particular, Chetty et al. (2025) provide a measure of economic mobility by reporting a child's expected earnings in adulthood (as a percentile) conditional on growing up in a given census tract with parents whose earnings fall in a particular income percentile. Chetty et al. (2025) also report analogous economic mobility estimates at the commuting zone (CZ) level. In a study motivated by the findings of Chetty et al. (2025), Bergman et al. (2024) report the effects of an informational intervention on the residential decisions of low-income housing voucher recipients in Seattle via the Creating Moves to Opportunity (CMTO) program. In particular, Bergman et al. (2024) use the economic mobility estimates of Chetty et al. (2025) to identify a set of high-opportunity tracts which they advertised to treated households in the CMTO program. We seek to study, in a set of exercises related to Andrews et al. (2022) and Mogstad et al. (2023), whether the CMTO can be expected to provide positive, long-run effects on the earnings of children in the treatment group, and relatedly whether the selection of high opportunity neighborhoods reflects noise as opposed to signal.

We seek to provide insight on both questions. To study the former question, we provide confidence regions for the economic mobility gains of the average, housing voucher recipient moving to an arbitrary, high-opportunity tract. We find that we fail to reject the possibility of null gains in the majority of urban Seattle tracts selected by the CMTO program. We replicate our analysis in the top fifty CZs in the US by population, and find heterogeneity in our findings. In some CZs, we are able to reject the possibility of null effects in the vast majority of selected tracts. Motivated by this surprising finding, we present analyses focused on studying whether the selection of high opportunity tracts reflects noise or signal in the estimates of Chetty et al. (2025). In particular, we study pairwise comparisons of high and low-opportunity census tracts in urban Seattle. We find that, for the majority of high-low opportunity tract pairs, we cannot reject the possibility of a null effect. We replicate this analysis for pairwise comparisons of high and low mobility commuting zones, and find the opposite. Indeed, our methods can provide informative inferences at the commuting zone level.

Our analyses are closely related to a larger literature in selective inference and multiple testing studying the mobility estimates of Chetty et al. (2025) and Bergman et al. (2024). Andrews et al. (2023) study the CMTO program of Bergman et al. (2024), showing that there exists a statistically significant, positive difference in the average mobility of high-opportunity tracts and the mobility of the average housing voucher recipient. Mogstad et al. (2023) show that one can say little about the relative ranks of census tracts or commuting zones according to economic mobility. Our analysis imposes a more strict coverage criterion than those of Andrews et al. (2023), but a less strict coverage criterion than that of Mogstad et al. (2023).

6.1 Empirical Findings

In this section, we provide an in-depth discussion of the empirical findings described above. First, we discuss tract-level effects in the CMTO program of Bergman et al. (2024). We then provide discussion of mobility effects at the commuting zone level.

Tract-level Effects in the CMTO Program

The CMTO program has attracted substantial interest in the selective inference and multiple testing literatures. We seek to provide further insights on tract-level effects, building on the work of Mogstad et al. (2023). In particular, Mogstad et al. (2023) study the problem of ranking tracts according to their true economic mobility effects. In a particularly stark finding, Mogstad et al. (2023) find that one cannot reject the possibility that the bottom-ranked tract in Seattle, according to estimated economic mobility, lies in the top third of tracts according to true economic mobility. As a result, Mogstad et al. (2023) conclude that:

"The classification of a given tract as a high upward-mobility neighbourhood may simply reflect statistical uncertainty (noise) rather than particularly high mobility (signal)."

This finding of Mogstad et al. (2023) is indeed surprising. As Chetty et al. (2025) remark, the methods of Mogstad et al. (2023) suggest that some of the poorest tracts in Los Angeles may be ranked above some of the wealthiest in terms of economic mobility. They suggest that the methods of Mogstad et al. (2023) may be too conservative. Indeed, per Chetty et al. (2025):

"This method is conservative because it assumes that the analyst is comparing all tracts in LA county (whereas in practice we focused on Watts given its well-known history of poverty and violence) and because it controls the family wise error rate (i.e., it requires that the probability that one or more of the millions of pairwise comparisons is wrong is less than 5%)."

In our analysis of tract-level effects, we focus power on selected tracts to address the former point. In particular, we seek to study tract-level effects for certain tracts of interest alone, namely high-opportunity tracts. Our two-step approach to inference focuses power onto selected tracts, while correcting for the winners' curse induced by selection. Our simulations demonstrate that our approach reduces over-coverage substantially relative to projection and provides substantially shorter confidence regions. However, we find that even when focusing inference on tracts of interest, the evidence on tract-level effects remains murky. Some aggregation of effects across tracts or loosening of the simultaneous coverage requirement may be necessary for informative inference.

In the CMTO program, Bergman et al. (2024) designate a subset of Seattle neighborhoods as high-opportunity according to their ranks. In particular, they label the top 20% and top 40% of urban and non-urban tracts, respectively, in the Seattle commuting zone as high-opportunity. This roughly corresponds to the top third of tracts in the commuting zone.

In our first analysis of the CMTO program, we compare the economic mobility estimates in each selected high-opportunity tract with an estimate of the average economic mobility of housing voucher recipients. To be precise, we allow X_j to be the economic mobility estimate for tract j in the set of all census tracts J in the Seattle commuting zone, or alternatively the set of all census tracts in urban Seattle. For each tract j, we also observe the number of housing voucher recipients residing in j, which we denote by c_j. We define Y_j as follows:

$$Y_j := X_j - \frac{\sum_{i \in J} X_i c_i}{\sum_{i \in J} c_i} \tag{25}$$

Y_j is an estimate of the expected gain or loss from moving the average housing voucher recipient to tract j.

In this first exercise, we take J to be the set of all urban Seattle tracts. We let $p = 132$ be the number of tracts in urban Seattle, and we take R to be $\{1, \ldots, \lfloor p/5 \rfloor = 26\}$. We are concerned with inference for the means $\mu_{Y,\hat{\jmath}}$ for $\hat{\jmath}$ in $\hat{J}_R$. In the figure below, we plot lower and upper confidence bounds for the mobility gains in selected Seattle tracts. We also apply the same exercise to Cleveland to demonstrate heterogeneity in findings between urban areas.

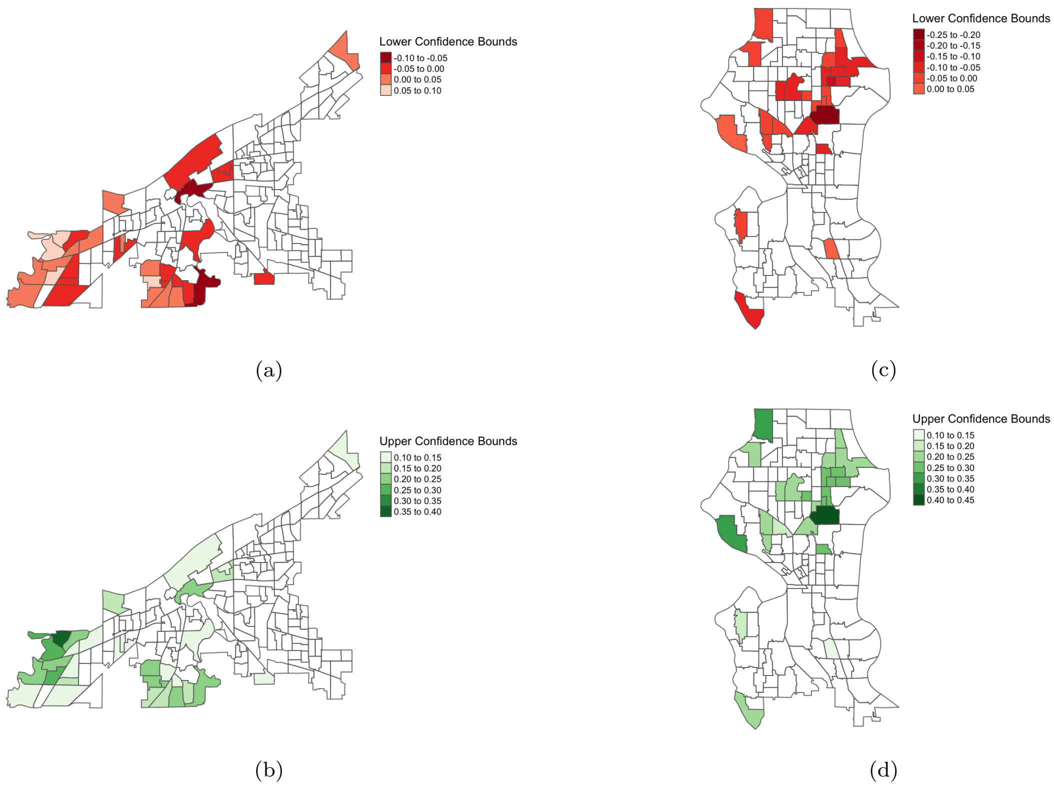


Figure 2: Lower and upper confidence bounds for economic mobility gains in selected urban tracts. Subfigures (a) and (b) display lower and upper confidence bounds on neighborhood effects in Cleveland, respectively. Subfigures (c) and (d) display lower and upper confidence bounds on neighborhood effects in Seattle, respectively.

In urban Seattle, we fail to reject the possibility of null effects at 92% of selected tracts. In urban Cleveland, we fail to reject the possibility of null effects in a comparatively small 35% of selected tracts.

Remark 6.1. Andrews et al. (2023) study the problem of inference for the mean:

$$\bar{\mu}_{Y,\hat{J}_R} := \frac{1}{\lfloor p/5 \rfloor} \sum_{\hat{j} \in \hat{J}_R} \mu_{Y,\hat{j}}$$

Their confidence region for $\bar{\mu}_{Y,\hat{J}_R}$ lies above zero, implying that they can reject the possibility of a null effect on economic mobility on aggregate. Our analysis differs from that of Andrews et al. (2023) in that we are concerned with constructing a confidence region satisfying simultaneous coverage of the $\mu_{Y,\hat{j}}$. It thereby provides insight into which of the selected tracts most credibly drive the positive aggregate effects. ■

Seattle is a fairly extreme example of this phenomenon. Repeating this analysis across the top 50 commuting zones by population in the US yields heterogeneous results. In certain commuting zones, our two-step approach to inference is reasonably powerful against the alternative of positive neighborhood effects. In this exercise, we focus on the economic mobility gains in the top-third of all tracts in a given commuting zone instead of focusing on urban tracts. We find that we fail to reject the null hypothesis of null effects in as few as 20.7% of selected tracts (in New Orleans) and as many as 93.6% of selected tracts (in Portland).

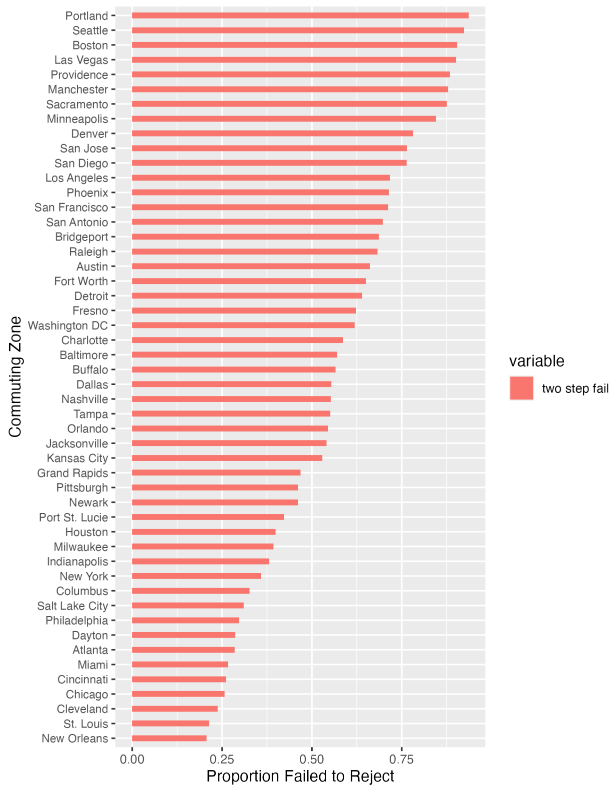


Figure 3: Proportion of selected tracts for which we fail to reject the possibility of a null effect on economic mobility relative to the average housing voucher recipient, by commuting zone. We provide results for all top-50 CZs by population.

In our final exercise on tract-level effects, we study pairwise comparisons of low and high-opportunity tracts. In particular, we consider the thought experiment of a household moving from an arbitrary low-opportunity to an arbitrary high-opportunity tract, and seek to study how often we can reject the possibility of a null effect on economic mobility associated with this move. As discussed in remark 3.5, this problem is closely related to the problem of inference for ranks. In their application to the economic mobility literature, Mogstad et al. (2023) consider the problem of inference on all pairwise comparisons in order to be agnostic on movement patterns among CMTO enrollees. Our analysis is notably different from that of Mogstad et al. (2023), since we restrict attention to pairwise comparisons of selected tracts. This allows us to be agnostic on movement patterns only among CMTO enrollees living in low-ranked tracts prior to treatment.

In particular, we consider the top and bottom fifths of urban tracts in Seattle, and for the sake of demonstrating heterogeneity between urban areas, Cleveland. In Seattle, we consider 26 high and 26 low-opportunity tracts of interest. We compare each of the top fifth and bottom fifth tracts, leading to 676 pairwise comparisons. We also consider pairwise comparisons of each top-fifth tract with the bottom-ranked tract, and a comparison of the top-ranked and bottom-ranked tracts.

Table 3: Pairwise Mobility Gains for Selected Tracts

CZ	τ-best	τ-worst	% Fail to Reject two-step	Lowest LCB two-step	Highest LCB TS two-step
Seattle	26	26	87%	[-0.33, 0.59]	[0.077, 0.50]
	26	1	100%	[-0.19, 0.60]	[-0.055, 0.66]
	1	1	100%	[-0.055, 0.66]	[-0.055, 0.66]
Cleveland	34	34	31%	[-0.076, 0.59]	[0.11, 0.49]
	34	1	9%	[-0.056, 0.38]	[0.11, 0.49]
	1	1	0%	[0.12, 0.48]	[0.12, 0.48]

Inference on the economic mobility gains associated with moving from low to high mobility urban tracts in Seattle and Cleveland.

We find that we fail to reject the possibility of a null effect associated with moving from an arbitrary low to an arbitrary high opportunity tract. This finding matches those of Mogstad et al. (2023), who suggest that one cannot reject the possibility of the bottom-ranked tract in the Seattle CZ according to estimated economic mobility lying in the top-third of tracts according to true economic mobility. Indeed, in Seattle, there remains little we can say about tract-level effects, even when attempting to focus the power of an inference procedure on certain tracts of interest.

Effects at the Commuting Zone Level

Our previous analyses sought to study tract-level effects on economic mobility in the context of CMTO. However, policymakers concerned with designing national level policies may be more concerned with targeting interventions according to commuting zone level estimates of economic mobility. We apply our methods to revisit the studies of Chetty et al. (2014) and Chetty and Hendren (2018) on the geography of economic mobility. While the analysis of Mogstad et al. (2023) suggests that it is difficult to construct a ranking of all commuting zones in the U.S. according to economic mobility, we consider a complementary exercise where we compare high and low mobility commuting zones. We find that the differences in mobility between high and low-mobility commuting zones are mostly statistically significant.

In the table below, we replicate our analysis of the mobility effects of moving from low to high opportunity areas at the commuting zone level. Our results on mobility effects at the commuting zone level are more conclusive than our findings at the tract-zone level. This is in part due to the simple fact that the standard errors on CZ level effects are smaller than those on tract level effects. We find that, for the majority of high and low-opportunity commuting zone pairs, we can indeed conclude that there exists a non-zero difference in the true mobilities of both CZs.

Table 4: Pairwise Mobility Gains for Selected Commuting Zones

Top	Bottom	% Fail to Reject two-step	Lowest LCB two-step	Highest LCB two-step
50%	50%	18.5%	[-0.28, 0.35]	[0.33, 0.44]
33%	33%	8.2%	[-0.25, 0.50]	[0.33, 0.44]
20%	20%	4.7%	[-0.19, 0.52]	[0.34, 0.44]
10%	10%	3.5%	[-0.14, 0.55]	[0.34, 0.44]
33%	50%	12.0%	[0.28, 0.35]	[0.33, 0.44]
20%	50%	8.3%	[-0.22, 0.38]	[0.33, 0.44]
10%	50%	7.1%	[-0.18, 0.41]	[0.34, 0.44]
33%	67%	19.9%	[-0.28, 0.35]	[0.33, 0.44]
20%	80%	19.1%	[-0.25, 0.29]	[0.33, 0.44]
10%	90%	17.6%	[-0.22, 0.31]	[0.34, 0.44]

Inference on the economic mobility gains associated with moving from low to high mobility commuting zones. We compare all commuting zones in the U.S.

Our results are particularly compelling in the context of Mogstad et al. (2023), who comment that:

"it is often not possible to tell apart with 95% confidence the CZs where children have opportunities to succeed from those without such opportunities."

By focusing on selected CZs, however, we are able to distinguish high and low-opportunity commuting zones with high frequency.

7 Simulation Study

In this section, we conduct an extended simulation study comparing the projection and two-step approaches to inference on multiple winners across a broad range of simulation designs. We demonstrate that the two-step approach to inference on multiple winners outperforms the projection approach almost uniformly. This outperformance is most pronounced in instances where the set of asymptotic winners is a proper subset of the set of all possible selections. In the simulation results presented in this section, we demonstrate that our methods perform favorably relative to standard approaches to inference such as projection, as well as relative to the state of the art of Zrnic and Fithian (2024b) and Zrnic and Fithian (2024a).

We consider simulations where $J = \{1, \dots, p\}$, for some natural number p. In total, we consider 28 distinct simulation designs. We present five selected simulation studies in this section, and present the results from all simulation studies in appendix D. We consider the following designs:

- **Design A** $p = 5$, $R = \{1\}$, $\mu_Y = 0$, $\mu_X = \arctan(j - 3)$
- **Design B** $p = 10$, $R = \{1\}$, $\mu_Y = 0$, $\mu_X = \arctan(j - 5.5)$
- **Design C** $p = 5$, $R = \{1\}$, $\mu_Y = 0$, $\mu_X = 0$

- **Design D** $p = 5$, $R = \{1\}$, $\mu_Y = 0$, $\mu_X = \mathbb{1}(j \in \{1, 2\})$

In all of our simulations, we consider four distinct covariance cases. In particular, we have a simple covariance case where X and Y are perfectly correlated but the X_j are independent, a low covariance case where all units are weakly correlated, a medium covariance case, and a high covariance case. In these four cases, we denote the variance covariance matrices by $\Sigma_{simple}, \Sigma_{low}, \Sigma_{medium}$, or Σ_{high}. We provide explicit formulae for these variance covariance matrices in appendix D. In the simulation results presented in this section, we take $\Sigma = \Sigma_{simple}$. We scale Σ by $1/n$ for sample size n equal to 100, 1000, and 10000. We emphasize that in all simulations, we treat Σ as known for computational simplicity.

We provide a table including results from selected designs, namely the low correlation case of designs A, B, C, and D which demonstrate the two-step approach's "clear-winner" property (see designs A and B).

Design	Confidence Set	Sample Size		
		100	1000	10000
A	Projection	0.991	0.991	0.992
	two-step	0.966	0.955	0.955
	Zoom	0.977	0.951	0.950
	Locally Simultaneous	0.971	0.955	0.955
B	Projection	0.993	0.996	0.995
	two-step	0.969	0.970	0.960
	Zoom	0.985	0.985	0.971
	Locally Simultaneous	0.979	0.976	0.962
C	Projection	0.976	0.976	0.976
	two-step	0.962	0.962	0.962
	Zoom	0.977	0.977	0.977
	Locally Simultaneous	0.979	0.979	0.979
D	Projection	0.991	0.991	0.991
	two-step	0.970	0.970	0.969
	Zoom	0.979	0.976	0.976
	Locally Simultaneous	0.978	0.978	0.979

Table 5: Coverage Probability in a Small Scale Simulation Study

We find that, between the models described above, the two-step approach substantially outperforms the projection approach when the set of winners $J_R(P)$ is clear, and specifically a proper subset of J such that $J \setminus J_R(P)$ is large. Moreover, in intermediate cases when the set of winners is moderately clear, as in designs A and B for low n, the two-step approach outperforms the approach of Zrnic and Fithian (2024a), which is based on the zoom test.[7] In general, the two-step approach also outperforms the approach of Zrnic and Fithian (2024b). Quantitatively, we have that in the four simulations presented above, the two-step approach to inference can reduce absolute overcoverage error by up to 88% relative to projection inference, 50% relative to locally simultaneous inference, and 56% relative to the zoom test. Across all simulations, including those in appendix D, the two-step approach to inference reduces overcoverage error by up to 96%

[7] Our implementation of the zoom test is based on the step-down procedure outlined in section 4 of Zrnic and Fithian (2024a).

relative to projection inference, up to 71% relative to locally simultaneous inference, and up to 67% relative to the zoom test.

Design	Confidence Set	Sample Size 100	1000	10000
A	two-step	0.835	0.780	0.780
	Zoom	0.872	0.786	0.763
	Locally Simultaneous	0.879	0.781	0.780
B	two-step	0.824	0.784	0.730
	Zoom	0.900	0.842	0.758
	Locally Simultaneous	0.888	0.827	0.731
C	two-step	0.937	0.937	0.937
	Zoom	1.003	1.003	1.003
	Locally Simultaneous	1.014	1.014	1.014
D	two-step	0.846	0.846	0.846
	Zoom	0.914	0.873	0.873
	Locally Simultaneous	0.887	0.886	0.886

Table 6: Confidence Interval Length in a Small Scale Simulation Study, as a Fraction of Projection Interval Length

Moreover, the table above demonstrates that, over a wide range of data generating processes, the two-step approach to inference provides tighter confidence regions than the projection, zoom, and locally simultaneous approaches to inference. Indeed, the interval lengths of the two-step approach to inference may be up to 27% shorter than projection inference. Moreover, two-step inference may be up to 11% shorter than inversions of the zoom test and up to 8% shorter than locally simultaneous inference.

Remark 7.1. In general, we recommend choosing $\beta = \alpha/10$. As our simulation results demonstrate, the two-step approach to inference performs quite well under such a choice of β. This choice of β is the same as in Andrews et al. (2023) and Romano et al. (2014). ■

References

Andrews, I., Bowen, D., Kitagawa, T. and McCloskey, A. (2022). Inference for losers. *AEA Papers and Proceedings*, **112** 635–42. URL https://www.aeaweb.org/articles?id=10.1257/pandp.20221065.

Andrews, I. and Kasy, M. (2019). Identification of and correction for publication bias. *American Economic Review*, **109** 2766–94. URL https://www.aeaweb.org/articles?id=10.1257/aer.20180310.

Andrews, I., Kitagawa, T. and McCloskey, A. (2023). Inference on Winners*. *The Quarterly Journal of Economics*, **139** 305–358. https://academic.oup.com/qje/article-pdf/139/1/305/55124607/qjad043.pdf, URL https://doi.org/10.1093/qje/qjad043.

Bachoc, F., Preinerstorfer, D. and Steinberger, L. (2017). Uniformly valid confidence intervals post-model-selection. 1611.01043.

Bergman, P., Chetty, R., DeLuca, S., Hendren, N., Katz, L. F. and Palmer, C. (2024). Creating moves to opportunity: Experimental evidence on barriers to neighborhood choice. *American Economic Review*, **114** 1281–1337. URL https://www.aeaweb.org/articles?id=10.1257/aer.20200407.

Berk, R., Brown, L., Buja, A., Zhang, K. and Zhao, L. (2013). Valid post-selection inference. *The Annals of Statistics*, **41** 802 – 837. URL https://doi.org/10.1214/12-AOS1077.

Cave, G., Bos, H. and Toussaint, C. (1993). Jobstart: Final report on a program for school dropouts. *Manpower Demonstration Research Corporation (MDRC)*. URL https://www.ojp.gov/ncjrs/virtual-library/abstracts/jobstart-final-report-program-school-dropouts.

Chetty, R., Friedman, J. N., Hendren, N., Jones, M. R. and Porter, S. R. (2025). The opportunity atlas: Mapping the childhood roots of social mobility. Tech. rep. URL https://opportunityinsights.org/wp-content/uploads/2018/10/atlas_paper.pdf.

Chetty, R. and Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility ii: County-level estimates*. *The Quarterly Journal of Economics*, **133** 1163–1228. https://academic.oup.com/qje/article-pdf/133/3/1163/25112629/qjy006.pdf, URL https://doi.org/10.1093/qje/qjy006.

Chetty, R., Hendren, N., Kline, P. and Saez, E. (2014). Where is the land of opportunity? the geography of intergenerational mobility in the united states *. *The Quarterly Journal of Economics*, **129** 1553–1623. https://academic.oup.com/qje/article-pdf/129/4/1553/30631636/qju022.pdf, URL https://doi.org/10.1093/qje/qju022.

Genz, A. (1992). Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics*, **1** 141–149. https://www.tandfonline.com/doi/pdf/10.1080/10618600.1992.10477010, URL https://www.tandfonline.com/doi/abs/10.1080/10618600.1992.10477010.

Gu, J. and Koenker, R. (2023). Invidious comparisons: Ranking and selection as compound decisions. *Econometrica*, **91** 1–41. https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA19304, URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA19304.

Haushofer, J. and Shapiro, J. (2016). The Short-term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya*. *The Quarterly Journal of Economics*, **131** 1973–2042. https://academic.oup.com/qje/article-pdf/131/4/1973/30636920/qjw025.pdf, URL https://doi.org/10.1093/qje/qjw025.

Hendren, N. and Sprung-Keyser, B. (2020). A unified welfare analysis of government policies*. *The Quarterly Journal of Economics*, **135** 1209–1318. https://academic.oup.com/qje/article-pdf/135/3/1209/33427738/qjaa006.pdf, URL https://doi.org/10.1093/qje/qjaa006.

Kuchibhotla, A. K., Kolassa, J. E. and Kuffner, T. A. (2022). Post-selection inference. *Annual Review of Statistics and Its Application*, **9** 505–527. URL https://www.annualreviews.org/content/journals/10.1146/annurev-statistics-100421-044639.

Lee, J. D., Sun, D. L., Sun, Y. and Taylor, J. E. (2016). Exact post-selection inference, with application to the lasso. *The Annals of Statistics*, **44** 907 – 927. URL https://doi.org/10.1214/15-AOS1371.

LIU, S. (2023). An exact sampler for inference after polyhedral model selection. 2308.10346, URL https://arxiv.org/abs/2308.10346.

MARKOVIC, J., XIA, L. and TAYLOR, J. (2018). Unifying approach to selective inference with applications to cross-validation. 1703.06559.

MCCLOSKEY, A. (2017). Bonferroni-based size-correction for nonstandard testing problems. *Journal of Econometrics*, **200** 17–35. URL https://www.sciencedirect.com/science/article/pii/S0304407617300556.

MCCLOSKEY, A. (2023). Hybrid confidence intervals for informative uniform asymptotic inference after model selection. *Biometrika*, **111** 109–127. https://academic.oup.com/biomet/article-pdf/111/1/109/56665894/asad023.pdf, URL https://doi.org/10.1093/biomet/asad023.

MILLER, C., BOS, J., PORTER, K., TSENG, F. M. and ABE, Y. (2005). The challenge of repeating success in a changing world – final report on the center for employment training replication sites. *Manpower Demonstration Research Corporation (MDRC)*. URL https://www.mdrc.org/work/publications/challenge-repeating-success-changing-world.

MOGSTAD, M., ROMANO, J. P., SHAIKH, A. M. and WILHELM, D. (2023). Inference for Ranks with Applications to Mobility across Neighbourhoods and Academic Achievement across Countries. *The Review of Economic Studies*, **91** 476–518. https://academic.oup.com/restud/article-pdf/91/1/476/56205143/rdad006.pdf, URL https://doi.org/10.1093/restud/rdad006.

ROMANO, J. P. and SHAIKH, A. M. (2008). Inference for identifiable parameters in partially identified econometric models. *Journal of Statistical Planning and Inference*, **138** 2786–2807. Special Issue in Honor of Theodore Wilbur Anderson, Jr. on the Occasion of his 90th Birthday, URL https://www.sciencedirect.com/science/article/pii/S0378375808000694.

ROMANO, J. P. and SHAIKH, A. M. (2012). On the uniform asymptotic validity of subsampling and the bootstrap. *The Annals of Statistics*, **40**. URL http://dx.doi.org/10.1214/12-AOS1051.

ROMANO, J. P., SHAIKH, A. M. and WOLF, M. (2014). A practical two-step method for testing moment inequalities. *Econometrica*, **82** 1979–2002. https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA11011, URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA11011.

SILVAPULLE, M. J. (1996). A test in the presence of nuisance parameters. *Journal of the American Statistical Association*, **91** 1690–1693. URL http://www.jstor.org/stable/2291597.

ZRNIC, T. and FITHIAN, W. (2024a). A flexible defense against the winner's curse. 2411.18569, URL https://arxiv.org/abs/2411.18569.

ZRNIC, T. and FITHIAN, W. (2024b). Locally simultaneous inference. *The Annals of Statistics*, **52** 1227–1253. URL https://doi.org/10.1214/24-AOS2391.

A Alternative Approaches to Inference on Multiple Winners

In this section, we discuss three alternative approaches to the inference on multiple winners problem. We first show that our two-step approach to inference can be intersected with a projection confidence set in the spirit of theorem 2 Zrnic and Fithian (2024b). Our second approach to inference is novel, but closely related to the approach of Zrnic and Fithian (2024b). Our final approach is based on Zrnic and Fithian (2024a). Our first approach uses a similar Bonferroni-type correction as the approach in section 3. However, this approach differs from our approach in section 3. Instead of estimating Δ from the first stage to model the errors on the winner in the second stage, we use Δ to estimate a set of likely winners, to which we restrict simultaneous inference in the second stage. This approach can be thought of as a version of the locally simultaneous inference of Zrnic and Fithian (2024b). We similarly demonstrate that the approach of Zrnic and Fithian (2024b) can be applied in our inference on multiple winners setting. Finally, we consider the approach of Zrnic and Fithian (2024a), who consider simultaneous testing of all means, but allocate the error budget according to the suboptimality (the observed gap of a given observation from the winner). Generally, all three of these approaches are underpowered compared to our two-step approach to inference.

A.1 A weakly improved two-step approach

We provide the following non-inferiority result in the spirit of theorem 2 of Zrnic and Fithian (2024b). To do so, we will introduce some new notation. For any subset J_c of J, we define $\bar{c}_{1-\alpha}(J_c)$ to be the $(1-\alpha)$-quantile of:

$$\max_{j \in J_c} \max \left\{ \frac{|\xi_{Y,j}|}{\sqrt{\Sigma_{Y,jj}}}, \frac{|\xi_{X,j}|}{\sqrt{\Sigma_{X,jj}}} \right\} \tag{26}$$

Again, unless stated otherwise, we will denote $\bar{c}_{1-\alpha}(J)$ by $\bar{c}_{1-\alpha}$. It turns out that under some mild conditions relating $\bar{c}_{1-\alpha}$, $d_{1-\beta}$ we can provide a weak improvement to two-step inference. In particular, we define the following confidence region:

$$CS^{TS2}_{1-\alpha;\beta} := \times_{\hat{j} \in \hat{J}_R} \left[Y_j - \left(\rho_{1-\alpha+\beta}(L,U) \wedge \bar{c}_{1-\alpha}\right) \sqrt{\Sigma_{Y,\hat{j}\hat{j}}}, Y_j + \left(\rho_{1-\alpha+\beta}(L,U) \wedge \bar{c}_{1-\alpha}\right) \sqrt{\Sigma_{Y,\hat{j}\hat{j}}} \right]$$

Suppose that $X = Y$. Then $CS^{TS2}_{1-\alpha;\beta} \subseteq CS^{P}_{1-\alpha}$.

Proposition A.1. *Suppose β is sufficiently small such that $2\bar{c}_{1-\alpha} \leq d_{1-\beta}(\Sigma)$. Then $CS^{TS2}_{1-\alpha;\beta}$ is a valid confidence set at the $1-\alpha$-level, such that:*

$$P_{\mu,\Sigma}\left((\mu_{Y,\hat{j}})_{\hat{j} \in \hat{J}_R} \in CS^{TS2}_{1-\alpha;\beta} \right) \geq 1 - \alpha \tag{27}$$

for all μ and Σ.

We prove this result in appendix C.

A.2 A Two-Step, Locally Simultaneous Approach to Inference

Let us consider the parametric setting from section 2. Let $\beta \in (0, \alpha)$, and let $\beta_1 \in (0, \beta)$ with $\beta_2 := \beta - \beta_1$. For arbitrary μ and Σ, let us pick some $J_c \subseteq J$ to be the smallest set such that:[8]

$$P_{\mu,\Sigma}\left(\hat{J}_R \subseteq J_c\right) \geq 1 - \beta_1 \tag{28}$$

Similarly, we seek to construct a confidence set $\hat{J}_c$ for J_c such that $P_{\mu,\Sigma}\left(J_c \subseteq \hat{J}_c\right) \geq 1 - \beta_2$. Finally, recall that in section 4, we defined $c_{1-\alpha}(J_c)$ to be the $1-\alpha$-quantile of (20).

It remains to show how to construct $\hat{J}_c$. By (5), it follows that:

$$p_j := P_{\mu,\Sigma}\left(j \in \hat{J}_R\right) = P_{\mu,\Sigma}\left(\sum_{j' \in J} \mathbb{1}\left(\xi_{X,j} \geq \xi_{X,j'} + \Delta_{j'j}\right) \in R\right) \tag{29}$$

Since the Δ_j are unknown, we cannot compute the above probabilities. Instead, we take the familiar approach of computing upper and lower confidence bounds for the Δ_j, which we denote by U and L respectively, at level $1 - \beta_2$. We constructed such U and L in equation (7).

It follows that $P_{\mu,\Sigma}(L \leq \Delta \leq U) \geq 1 - \beta_2$. We define $p_j(l, u)$ as follows:

$$p_j(l,u) = P_{\mu,\Sigma}\left(\left[\sum_{j' \in J} \mathbb{1}\left(\xi_{X,j} \geq \xi_{X,j'} + u_{j'j}\right), \sum_{j' \in J} \mathbb{1}\left(\xi_{X,j} \geq \xi_{X,j'} + l_{j',j}\right)\right] \cap R \neq \emptyset\right)$$

We have that $P_{\mu,\Sigma}\left(p(L,U) \leq p\right) \geq 1 - \beta_2$, where the inequality is interpreted elementwise. We take $\hat{J}_c$ to be the set that minimizes $\sum_{j \in \hat{J}_c} p_j(L,U)$ subject to the requirement that this sum exceed $1 - \beta_1$. If no such set exists, we take $\hat{J}_c = J$.

Finally, we can construct the following confidence set for the selected indices:

$$CS^{LSTS}_{1-\alpha;\beta} := \times_{\hat{j} \in \hat{J}_R}\left[Y_{\hat{j}} - c_{1-(\alpha-\beta)}\left(\hat{J}_c\right)\sqrt{\Sigma_{Y,\hat{j}\hat{j}}},\ Y_{\hat{j}} + c_{1-(\alpha-\beta)}\left(\hat{J}\right)\sqrt{\Sigma_{Y,\hat{j}\hat{j}}}\right]$$

The following proposition follows:

Proposition A.2. *$CS^{LSTS}_{1-\alpha;\beta}$ is a valid confidence set at the $1-\alpha$-level, such that:*

$$P_{\mu,\Sigma}\left(\left(\mu_{Y,\hat{j}}\right)_{\hat{j} \in \hat{J}_R} \in CS^{LSTS}_{1-\alpha;\beta}\right) \geq 1 - \alpha \tag{30}$$

for all μ and Σ. Marginal validity clearly follows.

A proof of this proposition is provided in appendix C.

A.3 A Zoom Test for Inference

In this section, we provide an approach to inference on multiple winners based on the zoom test of Zrnic and Fithian (2024a). Their approach suggests allocating the error budget to near-winners by inverting the zoom

[8] If multiple J_c of equal cardinality satisfy (28), we choose the J_c minimizing the probability on the left hand side of (28).

test. The zoom test is based on an acceptance region which is increasing in the population suboptimality - the difference between a candidate's mean and the population winner's mean. While it is unclear how to precisely generalize their approach to the exact inference on multiple winners problem we discuss in this paper, Zrnic and Fithian (2024b) provide guidance on the case of selecting the top-τ winners.

Let J_R be the set of indices j such that μ_X is ranked in R at j, that is $J_R = J_R(\mu_X)$. Let the suboptimality $D_j := \min_{j' \notin J_R} |\mu_{X,j} - \mu_{X,j'}|$ for any $j \in J$. Clearly, $D_j \geq 0$ for all $j \in J$, with equality for all $j \notin J_R$. As in Zrnic and Fithian (2024a), we choose r_α such that A_α, as defined below, is a valid level $1-\alpha$ acceptance region:

$$A_\alpha := \left[\mu_{Y,j} \pm \left(r_\alpha \vee \frac{D_j}{2}\right)\right]_{j \in J} \tag{31}$$

That is, we choose r_α to be the $1-\alpha$ quantile of the random variable:

$$\max_{j \in J} |\xi_{Y,j}| \mathbb{1}\left\{|\xi_{Y,j}| > \frac{D_j}{2}\right\}$$

Clearly, under the point hypothesis $H_0(\mu_Y, \mu_X) : \mathbb{E}(Y) = \mu_Y,\ \mathbb{E}(X) = \mu_X$, the probability that $(Y_j)_{j \in J}$ lies in A_α exceeds $1-\alpha$. We define the following, joint confidence for $(\mu_{Y,\hat{j}})_{\hat{j} \in \hat{J}_R}$:

$$CS_{1-\alpha} = \{\mu_Y,\ \mu_X; (Y_j)_{j \in J} \in A_\alpha\} \tag{32}$$

For any $\hat{j} \in \hat{J}_R$, we define the marginal confidence set:

$$CS^{zoom}_{1-\alpha,\hat{j}} := \{\mu_{Y,\hat{j}}; \exists\ \tilde{\mu}_Y,\ \tilde{\mu}_X \in CS_{1-\alpha} \quad \text{s.t.} \quad \tilde{\mu}_{Y,\hat{j}} = \mu_{Y,\hat{j}}\}$$

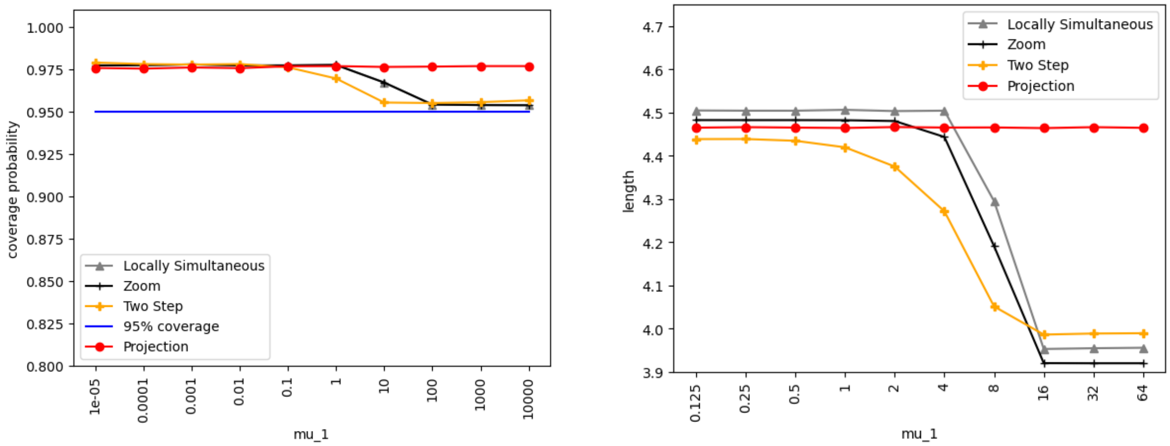
Consequently, we construct the zoom confidence set for the multiple winners:

$$CS^{zoom}_{1-\alpha} := \times_{\hat{j} \in \hat{J}_R} CS^{zoom}_{1-\alpha,\hat{j}} \tag{33}$$

Of course, it is not clear how to invert the test based on the acceptance region A_α for completely general R. Zrnic and Fithian (2024a) provide a parsimonious characterization of the confidence set based on inversion of the zoom test under certain cases, namely when R is of the form $\{1, \ldots, \tau\}$ for some $\tau \leq p$. Computationally, Zrnic and Fithian (2024b) provide a step-wise implementation of their methods which we implement in all simulations included in this paper.

A.4 Simulations Comparing Different Approaches to Inference

We now provide results from a small simulation study comparing our two-step approach to inference with the approaches above. We provide results from a more extensive simulation study in section D. In the simulations below, we consider $\Sigma = I_2$, $\mu = \begin{pmatrix}\mu_1 & 0\end{pmatrix}'$, and vary μ_1. Our two-step approach to inference substantially outperforms the approaches of Zrnic and Fithian (2024b) and Zrnic and Fithian (2024a) for intermediate values of μ_1. In particular, our approach reduces over-coverage error by as much as 68.6%. However, for extreme values of μ_1, where the winner is clear, the approaches of Zrnic and Fithian (2024b) and Zrnic and Fithian (2024a) slightly outperform our two-step approach, reducing over-coverage by up to 43.4%, although over-coverage for all methods is low in such instances. Our simulation results are plotted below:



(a) (b)

Figure 4: Confidence set coverage (a) and length (b) as μ_1 varies. Results are plotted for projection and two-step inference, as well as the locally simultaneous approach of Zrnic and Fithian (2024b), and the zoom test of Zrnic and Fithian (2024a)

A more comprehensive simulation study is provided in appendix D.

B Uniform Asymptotic Validity and Proofs

In this section, we formalize the assumptions stated in section 3.2 and introduce several lemmas supporting the proof of proposition 3.2. First, we provide the following uniform integrability assumption, which is equivalent to uniform convergence in distribution, per lemma B.1.

Assumption B.1. For $j = 1, \dots, 2p$, and for any ε, there exists K sufficiently large such that:

$$\sup_{P \in \mathcal{P}} \mathbb{E}_P \left(\frac{\left| \tilde{W}_{1,j} - \mu_{W,j}(P) \right|^2}{\Sigma_{W,jj}(P)} \mathbb{1} \left(\frac{\left| \tilde{W}_{1,j} - \mu_{W,j}(P) \right|}{\sqrt{\Sigma_{W,jj}(P)}} > K \right) \right) < \varepsilon \tag{34}$$

Our first lemma simply restates lemma 3.1 of Romano and Shaikh (2008).

Lemma B.1. *Under assumption B.1, we have uniform convergence in distribution such that there exist functions $\mu_{W,n}(P) := \mu_W(P)$ satisfying:*

$$\lim_{n \to \infty} \sup_{P \in \mathcal{P}} \sup_{S \in \mathcal{C}} \left| \Pr_P \left(\sqrt{n} \left(\tilde{S}_W^n - \mu_{W,n}(P) \right) \in S \right) - \Phi_{\Sigma(P)}(S) \right| = 0$$

where $\mathcal{C}$ denotes the set of convex subsets S of $\mathbb{R}^{2p}$ satisfying $\Phi_V(\partial S) = 0$ for all p.s.d. covariance matrices V such that $V_{j,j} = 1$ for all $j = 1, \dots, 2p$, and where $\Phi_V(\cdot)$ denotes the law of a random variable distributed according to a multivariate Gaussian with mean zero and variance-covariance V.

Proof. The result follows immediately by lemma 3.1 in Romano and Shaikh (2008). ■

We also provide the following technical lemma. Let $R_n(\tilde{S}^n_W, P_n)$ be some function of $\tilde{S}^n_W$, with the cumulative distribution function $J_n(x, P_n)$ under P_n. Let $\tilde{P}_n$ be some sequence of random distributions. The following lemma holds:

Lemma B.2. *Suppose, for any $\varepsilon > 0$, that the sequence of $\tilde{P}_n$ satisfies the following:*

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \Pr_P\left(\sup_{x\in\mathbb{R}} \left|J_n(x, \tilde{P}_n) - J_n(x, P)\right| \leq \varepsilon\right) = 1 \tag{35}$$

Then, for any $0 \leq \alpha_1$, $0 \leq \alpha_2$ such that $0 \leq \alpha_1 + \alpha_2 < 1$, the following holds:

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \Pr_P\left(J_n^{-1}(\alpha_1, \tilde{P}_n) \leq R_n(\tilde{S}^n_W, P_n) \leq J_n^{-1}(1-\alpha_2, \tilde{P}_n)\right) \geq 1 - \alpha_1 - \alpha_2 \tag{36}$$

Proof. Let us fix arbitrary $\eta > 0$. For any n sufficiently large, we have:

$$1 - \frac{\eta}{2} \leq \inf_{P\in\mathcal{P}} \Pr_P\left(\sup_{x\in\mathbb{R}} \left|J_n(x, \tilde{P}_n) - J_n(x, P)\right| \leq \frac{\eta}{2}\right)$$

This implies, by part viii of lemma A.1 of Romano and Shaikh (2012), that:

$$\inf_{P\in\mathcal{P}} \Pr_P\left(J_n^{-1}(\alpha_1, \tilde{P}_n) \leq R_n(\tilde{S}^n_W, P_n) \leq J_n^{-1}(1-\alpha_2, \tilde{P}_n)\right) \geq 1 - \alpha_1 - \alpha_2 - \eta$$

Finally, because η was arbitrary, we find that (36) holds, thus proving the lemma. ■

The following lemma concerns the asymptotic properties of L^n, and namely, whether L^n satisfies a uniform, asymptotic validity condition as a lower bound of $\Delta^n_{j'j} \equiv \Delta^n_{j'j}(P) := \mu_{X,j'}(P) - \mu_{X,j}(P)$. We naturally define L^n and U^n as follows:

$$L^n_{j'j} = \tilde{S}^n_{X,j'} - \tilde{S}^n_{X,j} - d_{1-\beta}(\hat{\Sigma}^n)\sqrt{\hat{\text{var}}^n_{j'j}}$$
$$U^n_{j'j} = \tilde{S}^n_{X,j'} - \tilde{S}^n_{X,j} + d_{1-\beta}(\hat{\Sigma}^n)\sqrt{\hat{\text{var}}^n_{j'j}}$$

where $\hat{\text{var}}^n_{j'j} = \hat{\Sigma}^n_{j'j'} + \hat{\Sigma}^n_{jj} - 2\hat{\Sigma}^n_{j'j}$, with an analogous definition for $\text{var}_{j'j}(P)$.

Lemma B.3. *Let us assume that assumption B.1 holds. Under these conditions, it follows that L^n is a uniformly, asymptotically valid lower bound for $\Delta^n(P)$. In particular:*

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \Pr_P\left(L^n \leq \Delta^n(P) \leq U^n\right) \geq 1 - \beta \tag{37}$$

Proof. We first notice that, by lemma s.6.1 and lemma s.7.1 of Romano and Shaikh (2012), and the continuous mapping theorem, we obtain following uniform consistency condition on the $\hat{\text{var}}^n_{j'j}$:

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \Pr_P\left(\left|\frac{n\hat{\text{var}}^n_{jj'}}{\text{var}_{jj'}(P)} - 1\right| > \varepsilon\right) = 0 \quad \text{for } \varepsilon > 0,\ j, j' \in J \text{ such that } \text{var}_{jj'}(P) > 0 \tag{38}$$

The remainder of our proof will use the results of lemma B.2. In particular, we define $R_n(\tilde{S}^n_W, P)$ as follows:

$$\max_{j,j'\in J, \text{var}_{j'j}(P)\neq 0} \frac{\sqrt{n}\left|(\tilde{S}^n_{X,j} - \mu_{X,j}(P)) - (\tilde{S}^n_{X,j'} - \mu_{X,j'}(P))\right|}{\sqrt{\text{var}_{j'j}(P)}} \tag{39}$$

Noting that the event that $R_n(\tilde{S}^n_W, P) \le x$ holds if and only if $\sqrt{n}(\tilde{S}^n_W - \mu_W(P))$ lies in a convex set S contained in $\mathcal{C}$, where we define $\mathcal{C}$ in lemma B.1. To see this, notice that $R_n(\tilde{S}^n_W, P) \le x$ if and only if for all j, j' in J such that $\text{var}_{j'j}(P) \neq 0$:

$$\left|(\tilde{S}^n_{X,j} - \mu_{X,j}(P)) - (\tilde{S}^n_{X,j'} - \mu_{X,j'}(P))\right| \le x\sqrt{\frac{\text{var}_{j'j}(P)}{n}}$$

It suffices to show that for any j, j' such that $\text{var}_{j'j}(P) \neq 0$, the event:

$$\left|(\tilde{S}^n_{X,j} - \mu_{X,j}(P)) - (\tilde{S}^n_{X,j'} - \mu_{X,j'}(P))\right| = x\sqrt{\frac{\text{var}_{j'j}(P)}{n}}$$

is probability zero according to Φ_V, where V is as described in lemma B.1. Clearly, this holds since we impose that $\text{var}_{j'j}(P) \neq 0$. Consequently, lemma B.1 gives us that:

$$\sup_{P\in\mathcal{P}} \sup_{x\in\mathbb{R}} \left|J_n(x, P) - J_n(x, \Phi_{\Sigma(P)})\right| = o(1)$$

We note that $\hat{\text{var}}_{jj'} = 0$ if and only if $\text{var}_{jj'}(P) = 0$. Consequently, the continuous mapping theorem, the uniform consistency result in (38), lemma s.7.1 in Romano and Shaikh (2008), and Polya's theorem provide that, for any $\varepsilon > 0$:

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \text{Pr}_P\left(\sup_{x\in\mathbb{R}}\left|J_n(x, \Phi_{\hat{\Sigma}^n}) - J_n(x, \Phi_{\Sigma(P)})\right| \le \varepsilon\right) = 1$$

or equivalently:

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \text{Pr}_P\left(\sup_{x\in\mathbb{R}}\left|J_n(x, \tilde{P}_n) - J_n(x, \Phi_{\Sigma(P)})\right| \le \varepsilon\right) = 1$$

where $\tilde{P}_n$ corresponds to the distribution of a gaussian with mean $\mu_W(P)$ and covariance $\hat{\Sigma}^n$. Thus, we can verify:

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \text{Pr}_P\left(\sup_{x\in\mathbb{R}}\left|J_n(x, \tilde{P}_n) - J_n(x, P)\right| \le \varepsilon\right) = 1$$

Applying lemma B.2 gives us that:

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \text{Pr}_P\left(J_n^{-1}(0, \tilde{P}_n) \le R_n(\tilde{S}^n_W, P) \le J_n^{-1}\left(1-\beta, \tilde{P}_n\right)\right) \ge 1-\beta$$

We can note that $J_n^{-1}(0, \tilde{P}_n) = 0$, while $J_n^{-1}(1-\beta, \tilde{P}_n) = d_{1-\beta}(\hat{\Sigma}^n)$. Consequently, algebraic manipulation of $R_n(\tilde{S}^n_W, P)$ as defined in (39) and an application of the uniform consistency result in (38) jointly imply that (37) holds, proving the lemma. ■

In addition, the following proposition demonstrates that $\rho_{1-(\alpha-\beta)}(\Delta^n(P), \Delta^n(P); \hat{\Sigma}^n)$ is an asymptotically valid upper bound for the maximum, studentized deviation between all $\hat{\jmath}$ in $\hat{J}_R$. Formally, we have the following lemma.

Lemma B.4. *Under assumption B.1, the following uniform asymptotic validity condition applies to $\rho_{1-(\alpha-\beta)}$:*

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \text{Pr}_P\left(\max_{j\in J}\left\{\frac{|\xi_{\tilde{S}^n_Y,j}|}{\sqrt{\hat{\Sigma}^n_{Y,jj}}}\mathbb{1}\left(\sum_{j'\in J}\mathbb{1}\left(\xi_{\tilde{S}^n_X,j} \ge \xi_{\tilde{S}^n_X,j'} + \Delta_{j'j}\right) \in R\right)\right\} > \rho_{1-\alpha+\beta}(\Delta, \Delta; \hat{\Sigma}^n)\right) \le \alpha - \beta$$

Proof. Our proof proceeds much as the proof of lemma B.3. Indeed, we define $R_n(\tilde{S}_W^n, P)$ as follows:

$$\max_{j\in J} \frac{\left|\sqrt{n}\left(\tilde{S}_{Y,j}^n - \mu_{Y,j}(P)\right)\right|}{\sqrt{\Sigma_{Y,jj}(P)}} \mathbb{1}\left(\sum_{j'\in J} \mathbb{1}\left(\sqrt{n}\left(\tilde{S}_{X,j}^n - \mu_{X,j}(P)\right) - \sqrt{n}\left(\tilde{S}_{X,j'}^n - \mu_{X,j'}(P)\right) \geq \sqrt{n}\Delta_{j'j}(P)\right) \in R\right) \quad (40)$$

Similarly, we may define $\hat{R}_n(\tilde{S}_W^n, P)$ as follows:

$$\max_{j\in J} \frac{\left|\sqrt{n}\left(\tilde{S}_{Y,j}^n - \mu_{Y,j}(P)\right)\right|}{\sqrt{n\hat{\Sigma}_{Y,jj}}} \mathbb{1}\left(\sum_{j'\in J} \mathbb{1}\left(\sqrt{n}\left(\tilde{S}_{X,j}^n - \mu_{X,j}(P)\right) - \sqrt{n}\left(\tilde{S}_{X,j'}^n - \mu_{X,j'}(P)\right) \geq \sqrt{n}\Delta_{j'j}(P)\right) \in R\right) \quad (41)$$

The indicator in the above expression is equivalent to the indicator $\mathbb{1}(j \in \hat{J}_{R;n})$, where $\hat{J}_{R;n} = \hat{J}_R(\tilde{S}_X^n)$. Thus, we seek to show that the event:

$$\max_{j\in \hat{J}_{R;n}} \frac{\left|\sqrt{n}\left(\tilde{S}_{Y,j}^n - \mu_{Y,j}(P)\right)\right|}{\sqrt{\Sigma_{Y,jj}(P)}} \leq x \quad (42)$$

lies in $\mathcal{C}$ as defined in assumption B.1. The above event can be rewritten as:

$$A(x) := \bigcup_{J_c \in 2^J} \left(\left\{\max_{j\in J_c} \frac{\left|\sqrt{n}\left(\tilde{S}_{Y,j}^n - \mu_{Y,j}(P)\right)\right|}{\sqrt{\Sigma_{Y,jj}(P)}} \leq x\right\} \bigcap \left\{\hat{J}_{R;n} = J_c\right\}\right)$$

We may notice that:

$$\partial A(x) \subseteq \bigcup_{J_c \in 2^J} \left\{\max_{j\in J_c} \frac{\left|\sqrt{n}\left(\tilde{S}_{Y,j}^n - \mu_{Y,j}(P)\right)\right|}{\sqrt{\Sigma_{Y,jj}(P)}} = x\right\} =: S$$

The right hand side of S is a subset of the boundary of a particular hyperrectangle, and thus is such that $\Phi_V(S) = 0$ for any V as described in lemma B.1. Now, the following hold:

$$\sup_{P\in\mathcal{P}} \sup_{x\in\mathbb{R}} \left|J_n(x,P) - J_n(x, \Phi_{\Sigma(P)})\right| = o(1)$$

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \Pr_P\left(\sup_{x\in\mathbb{R}} \left|J_n(x,\tilde{P}_n) - J_n(x,\Phi_{\Sigma(P)})\right| \leq \varepsilon\right) = 1$$

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \Pr_P\left(\sup_{x\in\mathbb{R}} \left|J_n(x,\tilde{P}_n) - J_n(x,P)\right| \leq \varepsilon\right) = 1$$

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \Pr_P\left(J_n^{-1}(0,\tilde{P}_n) \leq R_n(\tilde{S}_W^n, P) \leq J_n^{-1}\left(1-\alpha+\beta, \tilde{P}_n\right)\right) \geq 1-\alpha+\beta \quad (43)$$

where $\tilde{P}_n$ denotes the distribution of a gaussian with mean $\mu_W(P)$ and variance $\hat{\Sigma}^n$. The first equality holds by an application of lemma B.1. The second equality holds by lemmas s.6.1 and s.7.1 in Romano and Shaikh (2012), and Polya's theorem. The third equality is a consequence of the first two equalities, and the final equality is a consequence of B.2. Finally, we may notice that $J_n^{-1}(0,\tilde{P}_n) = 0$, and $J_n^{-1}(1-\alpha+\beta) = \rho_{1-\alpha+\beta}(\Delta(P), \Delta(P); \hat{\Sigma}^n)$ by construction. We also notice that assumption B.1 implies that for any $\varepsilon > 0$:

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \Pr_P\left(\left|\frac{n\hat{\Sigma}_{Y,jj}^n}{\Sigma_{Y,jj}(P)} - 1\right| > \varepsilon\right) = 0$$

by lemma s.6.1 of Romano and Shaikh (2012). This result, along with (43) gives that the following holds:

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \Pr_P \left(J_n^{-1}(0, \tilde{P}_n) \leq \hat{R}_n(\tilde{S}_W^n, P) \leq J_n^{-1}\left(1-\alpha+\beta, \tilde{P}_n\right)\right) \geq 1-\alpha+\beta$$

thus proving the lemma. ■

Finally, we can prove proposition 3.2:

Proof. **Proof of Proposition 3.2.** First, we note that, by lemma B.3:

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \Pr_P \left(L^n \leq \Delta^n(P) \leq U^n \right) \geq 1-\beta \tag{44}$$

We can define the event $B^n := \{L^n \leq \Delta^n \leq U^n\}$. In addition, we note that, for any sequence of events $\{A^n(P)\}_{n=1}^{\infty}$:

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \Pr_P(A^n(P)) \geq 1-\alpha \quad \text{if and only if} \quad \limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \Pr_P(A^n(P)^c) \leq \alpha$$

Thus, we have that, by (44):

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \Pr_P \left((B^n)^c\right) \leq \beta$$

and we seek to show that:

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \Pr_P \left((\mu_{Y,\hat{\jmath}}(P))_{\hat{\jmath}\in\hat{J}_R} \notin CS_{1-\alpha;\beta,n}^{TS} \right) \leq \alpha$$

Indeed:

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \Pr_P \left((\mu_{Y,\hat{\jmath}}(P))_{\hat{\jmath}\in\hat{J}_R} \notin CS_{1-\alpha;\beta,n}^{TS} \right) \\
&\leq \limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \Pr_P \left((\mu_{Y,\hat{\jmath}}(P))_{\hat{\jmath}\in\hat{J}_R} \notin CS_{1-\alpha;\beta,n}^{TS} \cap B^n \right) + \Pr_P((B^n)^c) \\
&\leq \limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \Pr_P \left((\mu_{Y,\hat{\jmath}}(P))_{\hat{\jmath}\in\hat{J}_R} \notin CS_{1-\alpha;\beta,n}^{TS} \cap B^n \right) + \limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \Pr_P((B^n)^c) \\
&\leq \limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \Pr_P \left((\mu_{Y,\hat{\jmath}}(P))_{\hat{\jmath}\in\hat{J}_R} \notin CS_{1-\alpha;\beta,n}^{TS} \cap B^n \right) + \beta \\
&\leq \limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \Pr_P \left(\bigcup_{\hat{\jmath}\in\hat{J}_R} \left\{ \frac{\sqrt{n}\left|\xi_{\tilde{S}_Y^n,\hat{\jmath}}\right|}{\sqrt{\hat{\Sigma}_{Y,\hat{\jmath}\hat{\jmath}}^n}} > \rho_{1-\alpha+\beta}\left(L^n, U^n; \hat{\Sigma}^n\right) \right\} \cap B^n \right) + \beta \\
&\leq \limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \Pr_P \left(\bigcup_{\hat{\jmath}\in\hat{J}_R} \left\{ \frac{\sqrt{n}\left|\xi_{\tilde{S}_Y^n,\hat{\jmath}}\right|}{\sqrt{\hat{\Sigma}_{Y,\hat{\jmath}\hat{\jmath}}^n}} > \rho_{1-\alpha+\beta}\left(\Delta^n, \Delta^n; \hat{\Sigma}^n\right) \right\} \right) + \beta \\
&\leq \limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \Pr_P \left(\max_{j\in J} \left\{ \frac{\sqrt{n}|\xi_{\tilde{S}_Y^n,j}|}{\sqrt{\hat{\Sigma}_{Y,jj}^n}} \mathbb{1}\left(\sum_{j'\in J} \mathbb{1}\left(\xi_{\tilde{S}_X^n,j} \geq \xi_{\tilde{S}_X^n,j'} + \Delta_{j'j}\right) \in R \right) \right\} > \rho_{1-\alpha+\beta}(\Delta^n, \Delta^n; \hat{\Sigma}^n) \right) + \beta \\
&\leq \alpha - \beta + \beta = \alpha
\end{aligned}$$

where the third inequality follows by lemma B.3 and the final inequality holds by lemma B.4. ■

C Proofs of Other Theoretical Results

First, we present the proof of our generalized polyhedral lemma.

Proof. **Proof of Lemma 4.1.** Let $U := \begin{pmatrix} X_1' & X_2' & \dots & X_k' & Y' \end{pmatrix}'$. Following the reasoning from the proof of Lemma 5.1 from Lee et al. (2016), the following holds:

$$
\begin{aligned}
\{AU \leq b\} &= \{A(c(B'U) + Z) \leq b\} \\
&= \{(Ac)(B'U) \leq b - AZ\}
\end{aligned}
$$

yielding a set of linear constraints on $B'U$, when conditioning on Z. Thus, because Z is independent of $B'U$, we find that $B'U$, conditional on the selection event $AU \leq b$ and sufficient statistic Z,[9] is distributed according to a multivariate normal with mean $\mu_B := B'\mu$ and variance-covariance $\Sigma_B := B'\Sigma B$, truncated to the polyhedron $\mathcal{O} := \{x; (Ac)x \leq b - AZ\}$. ■

In addition, we present a proof of the finite sample validity of conditional inference:

Proof. **Proof of Proposition 4.1.** First, we notice that $\{\mu_{Y,\hat{j}}\}_{\hat{j} \in \hat{J}_R} \in CS^c_{1-\alpha}$ if and only if our test $\phi(\,\cdot\,; (i_l)_{l \in R}, z, (\mu_{Y,i_l})_{l \in R})$ fails to reject. Because this test is a valid test at level α, (17) holds. ■

Now, we provide proofs for supplementary results.

Proof. **Proof of Proposition 3.1.** We recall that $P(B) \geq 1 - \beta$, where $B := \{L \leq \Delta \leq U\}$. Moreover, on B, we have $\rho_{1-\alpha+\beta}(L, U) \geq \rho_{1-\alpha+\beta}(\Delta, \Delta)$. We write:

$$
\begin{aligned}
&P_{\mu,\Sigma}\left(\bigcup_{\hat{j} \in \hat{J}_R} \left\{\frac{|\xi_{Y,\hat{j}}|}{\sqrt{\Sigma_{Y,\hat{j}\hat{j}}}} > \rho_{1-\alpha+\beta}(L, U)\right\}\right) \\
&\leq P_{\mu,\Sigma}\left(\bigcup_{\hat{j} \in \hat{J}_R} \left\{\frac{|\xi_{Y,\hat{j}}|}{\sqrt{\Sigma_{Y,\hat{j}\hat{j}}}} > \rho_{1-\alpha+\beta}(L, U)\right\} \cap B\right) + P(B^c) \\
&\leq P_{\mu,\Sigma}\left(\bigcup_{\hat{j} \in \hat{J}_R} \left\{\frac{|\xi_{Y,\hat{j}}|}{\sqrt{\Sigma_{Y,\hat{j}\hat{j}}}} > \rho_{1-\alpha+\beta}(\Delta, \Delta)\right\} \cap B\right) + \beta \\
&\leq P_{\mu,\Sigma}\left(\bigcup_{\hat{j} \in \hat{J}_R} \left\{\frac{|\xi_{Y,\hat{j}}|}{\sqrt{\Sigma_{Y,\hat{j}\hat{j}}}} > \rho_{1-\alpha+\beta}(\Delta, \Delta)\right\}\right) + \beta \\
&\leq P_{\mu,\Sigma}\left(\max_{j \in J} \frac{|\xi_{Y,j}|}{\sqrt{\Sigma_{Y,jj}}} \mathbb{1}\left(\sum_{j' \in J} \mathbb{1}\left(\xi_{X,j} \geq \xi_{X,j'} + \Delta_{j'j}\right) \in R\right) > \rho_{1-\alpha+\beta}(\Delta, \Delta)\right) + \beta \\
&= \alpha - \beta + \beta = \alpha
\end{aligned}
$$

for all μ and Σ. Consequently, a projection argument gives that for any $R' \subseteq R$:

$$
P_{\mu,\Sigma}\left(\bigcap_{\hat{j} \in \hat{J}_{R'}} \left\{\frac{|\xi_{Y,\hat{j}}|}{\sqrt{\Sigma_{Y,\hat{j}\hat{j}}}} \leq \rho_{1-\alpha+\beta}(L, U)\right\}\right) \geq 1 - \alpha
$$

[9] Here, Z can be thought of as a sufficient statistic for the nuisance parameters in our model.

for all μ, implying that both (10) and (11) hold for all μ. ■

We now provide a proof of proposition A.1, which proceeds much as the proof of proposition 3.1.

PROOF. **Proof of Proposition A.1.** We divide our proof into cases. In the first case, we have $\rho_{1-\alpha+\beta}(\Delta, \Delta) \leq \bar{c}_{1-\alpha}$. In the second case, we have $\rho_{1-\alpha+\beta}(\Delta, \Delta) > \bar{c}_{1-\alpha}$. The proof of validity in the first case proceeds exactly as in proposition 3.1, so we omit details and focus on proving validity in the second case. As before, we take $B := \{L \leq \Delta \leq U\}$. In the second case, we have the following:

$$
\begin{aligned}
&P_{\mu,\Sigma}\left(\mu_{Y,\hat{j}} \in CS^{TS2}_{1-\alpha;\beta} \quad \text{for all } \hat{j} \in \hat{J}_R\right) \\
&\geq P_{\mu,\Sigma}\left(\left\{\mu_{Y,\hat{j}} \in CS^{TS2}_{1-\alpha;\beta} \quad \text{for all } \hat{j} \in \hat{J}_R\right\} \cap B\right) \\
&\geq P_{\mu,\Sigma}\left(\left\{\mu_{Y,\hat{j}} \in \left[Y_{\hat{j}} \pm (\rho_{1-\alpha+\beta}(\Delta,\Delta) \wedge \bar{c}_{1-\alpha})\sqrt{\Sigma_{Y,\hat{j}\hat{j}}}\right] \quad \text{for all } \hat{j} \in \hat{J}_R\right\} \cap B\right) \\
&\geq P_{\mu,\Sigma}\left(\left\{\mu_{Y,\hat{j}} \in \left[Y_{\hat{j}} \pm \bar{c}_{1-\alpha}\sqrt{\Sigma_{Y,\hat{j}\hat{j}}}\right] \quad \text{for all } \hat{j} \in \hat{J}_R\right\} \cap B\right) \\
&\geq P_{\mu,\Sigma}\left(\left\{\mu_{Y,j} \in \left[Y_j \pm \bar{c}_{1-\alpha}\sqrt{\Sigma_{Y,jj}}\right] \quad \text{for all } j \in J\right\} \cap B\right) \\
&\geq P_{\mu,\Sigma}\left(\left\{\mu_{Y,j} \in \left[Y_j \pm \bar{c}_{1-\alpha}\sqrt{\Sigma_{Y,jj}}\right], \mu_{X,j} \in \left[X_j \pm \bar{c}_{1-\alpha}\sqrt{\Sigma_{X,jj}}\right] \quad \text{for all } j \in J\right\} \cap B\right) \\
&= P_{\mu,\Sigma}\left(\mu_{Y,j} \in \left[Y_j \pm \bar{c}_{1-\alpha}\sqrt{\Sigma_{Y,jj}}\right], \mu_{X,j} \in \left[X_j \pm \bar{c}_{1-\alpha}\sqrt{\Sigma_{X,jj}}\right] \quad \text{for all } j \in J\right) \\
&\geq 1-\alpha
\end{aligned}
$$

where the final equality holds since $2\bar{c}_{1-\alpha} \leq 2d_{1-\beta}(\Sigma)$ implies that:

$$
\left\{\mu_{Y,j} \in \left[Y_j \pm \bar{c}_{1-\alpha}\sqrt{\Sigma_{Y,jj}}\right], \mu_{X,j} \in \left[X_j \pm \bar{c}_{1-\alpha}\sqrt{\Sigma_{X,jj}}\right] \quad \text{for all } j \in J\right\} \subseteq B
$$

Consequently, this casework gives that equation 27 holds. ■

Before providing a proof of proposition 4.2, we provide the following technical lemma.

Lemma C.1. *Let X_n and Y_n be random variables for $n \in \mathbb{N}$ with laws $J_{X,n}$ and $J_{Y,n}$. Suppose also that:*

1. *$X_n \leq Y_n$ with probability approaching one.*
2. *Y_n has a weak limit Y with a continuous cumulative distribution function J_Y that admits a strictly positive density on $\mathbb{R}$.*

For any $\alpha \in (0,1)$, the quantile functions satisfy:

$$
\limsup_{n\to\infty} J^{-1}_{X,n}(\alpha) - J^{-1}_{Y,n}(\alpha) \leq 0 \tag{45}
$$

PROOF. Notice that the following holds:

$$
\begin{aligned}
&\liminf_{n\to\infty} \inf_{x\in\mathbb{R}} P(X_n \leq x) - P(Y_n \leq x) \\
&= \liminf_{n\to\infty} \inf_{x\in\mathbb{R}} P(\{X_n \leq x\} \cap \{X_n \leq Y_n\}) - P(Y_n \leq x) + P(\{X_n \leq x\} \cap \{X_n > Y_n\})
\end{aligned}
$$

$$\geq \liminf_{n\to\infty} \inf_{x\in\mathbb{R}} P(\{X_n \leq x\} \cap \{X_n \leq Y_n\}) - P(Y_n \leq x)$$
$$\geq \liminf_{n\to\infty} \inf_{x\in\mathbb{R}} P(\{X_n \leq x\} \cap \{X_n \leq Y_n\}) - P(\{Y_n \leq x\} \cap \{X_n \leq Y_n\}) - P(X_n > Y_n)$$
$$\geq \liminf_{n\to\infty} -P(X_n > Y_n) = 0$$

It follows that:

$$\liminf_{n\to\infty} \inf_{x\in\mathbb{R}} J_{X,n}(x) - J_{Y,n}(x) \geq 0$$

and similarly, by Polya's theorem, that:

$$\liminf_{n\to\infty} \inf_{x\in\mathbb{R}} J_{X,n}(x) - J_Y(x) \geq 0$$

Equivalently:

$$\limsup_{n\to\infty} \sup_{x\in\mathbb{R}} J_Y(x) - J_{X,n}(x) \leq 0$$

Consequently, by part (ii) of lemma A.1 in Romano and Shaikh (2012), we obtain that for any $\varepsilon > 0$, there exists sufficiently large n such that $J_{X,n}^{-1}(\alpha) \leq J_Y^{-1}(\alpha+\varepsilon)$. Because Y admits a strictly positive density, we obtain that $\limsup_{n\to\infty} J_{X,n}^{-1}(\alpha) \leq J_Y^{-1}(\alpha)$. Similarly, we know that $\lim_{n\to\infty} J_{Y,n}^{-1}(\alpha) = J_Y^{-1}(\alpha)$, finally implying (45). ■

Finally, we prove proposition 4.2. Intuitively, we show that L^n and U^n approach the true $\Delta(P)$ in probability. Given this, we show that our modeled, selected errors $f(L^n, U^n)$ can be bounded above by the largest absolute errors in $J_R(P)$ with probability approaching one. An application of lemma C.1 concludes the proof.

PROOF. **Proof of Proposition 4.2.** Let us fix some arbitrary $\beta > 0$. We begin by recalling that:

$$L^n_{j'j} = \tilde{S}^n_{X,j'} - \tilde{S}^n_{X,j} - d_{1-\beta}(\hat{\Sigma}^n)\sqrt{\hat{\text{var}}^n_{j'j}}$$
$$U^n_{j'j} = \tilde{S}^n_{X,j'} - \tilde{S}^n_{X,j} + d_{1-\beta}(\hat{\Sigma}^n)\sqrt{\hat{\text{var}}^n_{j'j}}$$

We note that, by the law of large numbers and by the continuous mapping theorem, for fixed P in $\mathcal{P}$, $\left\| n\hat{\Sigma}^n - \Sigma(P) \right\| = o_P(1)$. We also know that, for any fixed Σ, $d_{1-\beta}(\Sigma)$ is $O(1)$, meaning that $d_{1-\beta}(\hat{\Sigma}^n)$ is $O_P(1)$. Finally, recalling that $\hat{\text{var}}^n_{j'j} = \hat{\Sigma}^n_{j'j'} + \hat{\Sigma}^n_{jj} - 2\hat{\Sigma}^n_{j'j}$, we notice that $\hat{\text{var}}^n_{j'j} = o_P(1)$. Thus, by the weak law of large numbers and that fact that the above facts imply that $d_{1-\beta}(\hat{\Sigma}^n)\sqrt{\hat{\text{var}}^n_{j'j}} = o_P(1)$:

$$\begin{pmatrix} L^n_{j'j} \\ U^n_{j'j} \end{pmatrix} \xrightarrow{P} \begin{pmatrix} \Delta_{j'j}(P) \\ \Delta_{j'j}(P) \end{pmatrix} \tag{46}$$

We define $\delta := \min_{j\in J_R(P), i\notin J_R(P)} |\mu_{X,j} - \mu_{X,i}|$. Naturally, because $J_R(P)$ is a proper subset of J, $\delta > 0$. It follows from (46) that the event $A_n := \{|L^n - \Delta|,\ |U^n - \Delta| \leq \delta/6\}$ satisfies $P(A_n) \to 1$. We seek to study the behavior of $\rho_{1-\alpha+\beta}\left(L^n, U^n; \hat{\Sigma}^n\right)$ on A_n. We now seek to show that:

$$\left(\mathbb{1}\left(\left[\sum_{j'\in J} \mathbb{1}\left(\xi^n_{X,j} \geq \xi^n_{X,j'} + U^n_{j'j}\right), \sum_{j'\in J} \mathbb{1}\left(\xi^n_{X,j} \geq \xi^n_{X,j'} + L^n_{j'j}\right)\right] \cap R \neq \emptyset\right)\right)_{j\in J} \leq \left(\mathbb{1}\left(j \in J_R(P)\right)\right)_{j\in J} \tag{47}$$

with probability approaching one. We let l^n and u^n be such that $|l^n - \Delta|,\ |u^n - \Delta| \leq \delta/6$. We are interested in the following probability:

$$\Pr_P\left(\mathbb{1}\left(\left[\sum_{j'\in J}\mathbb{1}\left(\xi^n_{X,j} \geq \xi^n_{X,j'} + u^n_{j'j}\right), \sum_{j'\in J}\mathbb{1}\left(\xi^n_{X,j} \geq \xi^n_{X,j'} + l^n_{j'j}\right)\right] \cap R \neq \emptyset\right)_{j\in J} \leq \mathbb{1}\left(j \in J_R(P)\right)_{j\in J}\right),$$

where ξ^n_W are taken to have a Gaussian law with mean zero and variance-covariance $\hat{\Sigma}^n$. Now, we define B_n to be correspond to the event that all $\xi^n_{X,j}$ are within $\delta/6$ of zero. Formally, $B_n := \{\max_{j\in J}|\xi^n_{X,j}| < \delta/6\}$. Since $\xi^n_{X,j} \xrightarrow{P} 0$, it follows that $P(B_n)$ approaches one as n approaches infinity. For such l^n and u^n as above, it is clear that:

$$\begin{aligned}
&\lim_{n\to\infty}\Pr_P\left(\left(\mathbb{1}\left(\left[\sum_{j'\in J}\mathbb{1}\left(\xi^n_{X,j} \geq \xi^n_{X,j'} + u^n_{j'j}\right), \sum_{j'\in J}\mathbb{1}\left(\xi^n_{X,j} \geq \xi^n_{X,j'} + l^n_{j'j}\right)\right] \cap R \neq \emptyset\right)\right)_{j\in J} \leq \left(\mathbb{1}\left(j \in J_R(P)\right)\right)_{j\in J}\right) \\
&= 1 - \lim_{n\to\infty}\Pr_P\left(\left\{\exists j \notin J_R(P); \left[\sum_{j'\in J}\mathbb{1}\left(\xi^n_{X,j} \geq \xi^n_{X,j'} + u^n_{j'j}\right), \sum_{j'\in J}\mathbb{1}\left(\xi^n_{X,j} \geq \xi^n_{X,j'} + l^n_{j'j}\right)\right] \cap R \neq \emptyset\right\}\right) \\
&= 1 - \lim_{n\to\infty}\Pr_P\left(\left\{\exists j \notin J_R(P); \left[\sum_{j'\in J}\mathbb{1}\left(\xi^n_{X,j} \geq \xi^n_{X,j'} + u^n_{j'j}\right), \sum_{j'\in J}\mathbb{1}\left(\xi^n_{X,j} \geq \xi^n_{X,j'} + l^n_{j'j}\right)\right] \cap R \neq \emptyset\right\} \cap B_n\right) \\
&\geq 1 - \lim_{n\to\infty}\Pr_P\left(\left\{\exists j \notin J_R(P); \left[\sum_{j'\in J}\mathbb{1}\left(0 \geq \Delta_{j'j} + \delta/2\right), \sum_{j'\in J}\mathbb{1}\left(0 \geq \Delta_{j'j} - \delta/2\right)\right] \cap R \neq \emptyset\right\}\right) \\
&= 1 - \Pr_P(\emptyset) = 1,
\end{aligned}$$

To see that the inequality holds, we notice that on B_n:

$$\left[\sum_{j'\in J}\mathbb{1}\left(\xi^n_{X,j} \geq \xi^n_{X,j'} + u^n_{j'j}\right), \sum_{j'\in J}\mathbb{1}\left(\xi^n_{X,j} \geq \xi^n_{X,j'} + l^n_{j'j}\right)\right] \subseteq \left[\sum_{j'\in J}\mathbb{1}\left(0 \geq \Delta_{j'j} + \delta/2\right), \sum_{j'\in J}\mathbb{1}\left(0 \geq \Delta_{j'j} - \delta/2\right)\right]$$

To see that the last equality holds, notice that $j \in J_R(P)$ if and only if:

$$\sum_{j'\in J}\mathbb{1}\left(0 \geq \Delta_{j'j}\right) \in R$$

Suppose that there exists $j' \notin J_R(P)$ such that there exists $r \in R$ satisfying:

$$r \in \left[\sum_{j''\in J}\mathbb{1}\left(0 \geq \Delta_{j''j'} + \delta/2\right), \sum_{j''\in J}\mathbb{1}\left(0 \geq \Delta_{j''j'} - \delta/2\right)\right]$$

Let $j_r(\mu_X)$ be some element in $J_R(P)$ such that $r = \sum_{j''\in J}\mathbb{1}\left(0 \geq \Delta_{j''j_r(\mu_X)}\right)$. Assume that j' is some index in $J_R(P)^c$ such that $\mu_{X,j'} < \mu_{X,j_r(\mu_X)}$. A symmetric argument holds when $\mu_{X,j'} > \mu_{X,j_r(\mu_X)}$. We have that:

$$\sum_{j''\in J}\mathbb{1}\left(0 \geq \Delta_{j''j'} + \delta/2\right) \leq \sum_{j''\in J}\mathbb{1}\left(0 \geq \Delta_{j''j_r(\mu_X)}\right) \leq \sum_{j''\in J}\mathbb{1}\left(0 \geq \Delta_{j''j'} - \delta/2\right)$$

Since the largest r elements in $\Delta_{\cdot j_r(\mu_X)}$ coincide with the largest r elements in $\Delta_{\cdot j'}$, these inequalities imply that $\mathbb{1}\left(0 \geq \Delta_{j_r(\mu_X) j_r(\mu_X)}\right) \leq \mathbb{1}\left(0 \geq \Delta_{j_r(\mu_X) j'} - \delta/2\right)$. Of course, $\mathbb{1}\left(0 \geq \Delta_{j_r(\mu_X) j_r(\mu_X)}\right) = 1$ since $\Delta_{j_r(\mu_X) j_r(\mu_X)} = 0$. It follows that $\Delta_{j_r(\mu_X) j'} \leq \delta/2$ and consequently that $|\Delta_{j_r(\mu_X) j'}| \leq \delta/2$, since we assumed that $\mu_{X,j'} < \mu_{X,j_r(\mu_X)}$. This cannot hold, since $j' \notin J_R(P)$ and thus $|\Delta_{j' j_r(\mu_X)}| \geq \delta$. Thus, the event:

$$\left\{\exists j \notin J_R(P);\ \left[\sum_{j' \in J} \mathbb{1}\left(0 \geq \Delta_{j'j} - \delta/2\right), \sum_{j' \in J} \mathbb{1}\left(0 \geq \Delta_{j'j} + \delta/2\right)\right] \cap R \neq \emptyset\right\}$$

is indeed the empty set. Consequently, whenever l^n and u^n are as above, $\rho_{1-\alpha+\beta}(l^n, u^n; \hat{\Sigma}^n) \leq c_{1-\alpha+\beta}(J_R(P)) + o(1)$ by lemma C.1. That condition 1 in the lemma holds is a consequence of the fact that, as established above:

$$\Pr_P\left(\mathbb{1}\left(\left[\sum_{j' \in J} \mathbb{1}\left(\xi^n_{X,j} \geq \xi^n_{X,j'} + u^n_{j'j}\right), \sum_{j' \in J} \mathbb{1}\left(\xi^n_{X,j} \geq \xi^n_{X,j'} + l^n_{j'j}\right)\right] \cap R \neq \emptyset\right) \leq \mathbb{1}\left(j \in J_R(P)\right)\right) \to 1$$

Condition 2 of the lemma holds since, by the continuous mapping theorem and an application of lemma s.6.1 in Romano and Shaikh (2012), we obtain:

$$\max_{j \in J_R(P)} \frac{|\xi_{Y,j}|}{\sqrt{\hat{\Sigma}^n_{jj}}} \xrightarrow{d} \max_{j \in J_R(P)} \frac{|\xi_{Y,j}|}{\sqrt{\Sigma_{jj}(P)}} \tag{48}$$

where the right hand side admits a positive-everywhere, continuous density over $\mathbb{R}_+$. Similarly, because $P(A_n) \to 1$ and because $\|n\hat{\Sigma}^n - \Sigma(P)\| = o_p(1)$, we obtain $\rho_{1-\alpha+\beta}(L^n, U^n; \hat{\Sigma}^n) \leq c_{1-\alpha+\beta}(J_R(P))$. If β is sufficiently small such that $c_{1-\alpha+\beta}(J_R(P)) < c_{1-\alpha}(J)$, we get $\rho_{1-\alpha+\beta}(L^n, U^n; \hat{\Sigma}^n) < c_{1-\alpha}(J)$. The desired result now follows. ■

Now, we provide proofs of propositions A.2 and 4.4, which concern the validity of both locally simultaneous approaches to inference.

PROOF. **Proof of Proposition A.2.** Let G be the event that $L \leq \Delta \leq U$. We know that G satisfies $P_{\mu,\Sigma}(G) \geq 1 - \beta_2$. On G, we have $p_j(L,U) \geq p_j$ for any $j = 1, \ldots, p$. Consequently, we have that, on G, $J_c \subseteq \hat{J}_c$. We will also let F be the event that $\hat{J}_R \subseteq J_c$, which satisfies $P_{\mu,\Sigma}(G) \geq 1 - \beta_1$. Thus, we have:

$$\begin{aligned}
P_{\mu,\Sigma}\left(\exists\, \hat{\jmath} \in \hat{J}_R;\ \mu_{Y,\hat{\jmath}} \notin CS^{LSTS}_{1-\alpha;\beta,\hat{\jmath}}\right) &= P_{\mu,\Sigma}\left(\max_{\hat{\jmath} \in \hat{J}_R} \frac{|\xi_{Y,\hat{\jmath}}|}{\sqrt{\Sigma_{Y,\hat{\jmath}\hat{\jmath}}}} > c_{1-\alpha+\beta}\left(\hat{J}_c\right)\right) \\
&\leq P_{\mu,\Sigma}\left(\max_{\hat{\jmath} \in \hat{J}_R} \frac{|\xi_{Y,\hat{\jmath}}|}{\sqrt{\Sigma_{Y,\hat{\jmath}\hat{\jmath}}}} > c_{1-\alpha+\beta}\left(\hat{J}_c\right) \cap G \cap F\right) + P_{\mu,\Sigma}(G^c) + P_{\mu,\Sigma}(F^c) \\
&\leq P_{\mu,\Sigma}\left(\max_{\hat{\jmath} \in \hat{J}_R} \frac{|\xi_{Y,\hat{\jmath}}|}{\sqrt{\Sigma_{Y,\hat{\jmath}\hat{\jmath}}}} > c_{1-\alpha+\beta}\left(\hat{J}_c\right) \cap G \cap F\right) + \beta_1 + \beta_2 \\
&\leq P_{\mu,\Sigma}\left(\max_{\hat{\jmath} \in J_c} \frac{|\xi_{Y,\hat{\jmath}}|}{\sqrt{\Sigma_{Y,\hat{\jmath}\hat{\jmath}}}} > c_{1-\alpha+\beta}\left(J_c\right) \cap G \cap F\right) + \beta_1 + \beta_2 \\
&\leq P_{\mu,\Sigma}\left(\max_{\hat{\jmath} \in J_c} \frac{|\xi_{Y,\hat{\jmath}}|}{\sqrt{\Sigma_{Y,\hat{\jmath}\hat{\jmath}}}} > c_{1-\alpha+\beta}\left(J_c\right)\right) + \beta_1 + \beta_2 \leq \alpha - \beta + \beta_1 + \beta_2 = \alpha
\end{aligned}$$

Thus proving the proposition. ■

Proof. **Proof of Proposition 4.4.**

Our proof follows via application of theorem 1 of Zrnic and Fithian (2024b). We can define the following set of plausible targets, given a realization $X = x$:

$$\hat{J}_R^+ = \bigcup_{x';\ \sup_{j' \neq j} |(x'_{j'} - x_{j'}) - (x'_j - x_j)| \leq 2\bar{d}_{1-\beta}(\Sigma)} \hat{J}_R(x')$$

which we can verify, via an application of the triangle inequality, is equivalent to $\hat{\Gamma}_\beta^+$ taking the acceptance region $A_\beta(\mu')$ to be as follows:

$$A_\beta(\mu') = \left\{(y,x);\ \sup_{1 \leq j,j' \leq p} \left|X_j - \mu'_j - (X_{j'} - \mu'_{j'})\right| \leq \bar{d}_{1-\beta}(\Sigma)\right\}$$

We may notice that $\sup_{j' \neq j} |(x'_{j'} - x_{j'}) - (x'_j - x_j)| \leq 2\bar{d}_{1-\beta}(\Sigma)$ is equivalent to $\sup_{j' \neq j} |x'_{j'} - x_j| \leq \bar{d}_{1-\beta}(\Sigma)$. Following the reasoning of theorem 3 in Zrnic and Fithian (2024b), we seek to find a most favorable choice of x' for an index j to be included in $\hat{J}_R(x)$. To allow $j \in \hat{J}_R(x')$, let j' be the element in $\hat{J}_R(x)$ that minimizes $|x_j - x_{j'}|$. Suppose, without loss, that $x_j < x_{j'}$. We can obtain the most favorable perturbation by taking $x'_j + \bar{d}_{1-\beta}(\Sigma)$ and $x'_{j'} - \bar{d}_{1-\beta}(\Sigma)$. Consequently, we can write:

$$\hat{J}_R^+ = \left\{j;\ \exists j' \in \hat{J}_R \text{ s.t. } |X_{j'} - X_j| \leq 2\bar{d}_{1-\beta}(\Sigma)\right\}$$

thus proving the proposition. ■

Finally, we provide a proof of proposition 4.5, which compares our two-step approach to inference to the locally simultaneous approach of Zrnic and Fithian (2024b).

Proof. **Proof of Proposition 4.5.**

Notice that, whenever $\max_{j,j' \in J} |\mu_{X,j} - \mu_{X,j'}| \leq \bar{d}_{1-\beta}(\Sigma)$, the following series of implications holds, for arbitrary $j, j' \in J$:

$$\begin{aligned}
&|\xi_{X,j} - \xi_{X,j'}| \leq \bar{d}_{1-\beta}(\Sigma) \\
&\implies |\xi_{X,j} - \xi_{X,j'}| + |\mu_{X,j} - \mu_{X,j'}| \leq 2\bar{d}_{1-\beta}(\Sigma) \\
&\implies |X_j - X_{j'}| \leq 2\bar{d}_{1-\beta}(\Sigma)
\end{aligned}$$

where the final inequality is a consequence of the triangle inequality. Consequently, $P_{\mu,\Sigma}(\max_{j,j' \in J} |X_j - X_{j'}| \leq 2\bar{d}_{1-\beta}(\Sigma)) \geq 1 - \beta$. On this event, it holds that $\hat{J}_R^+ = J$. Moreover, on the event that $|\xi_{X,j} - \xi_{X,j'}| \leq \bar{d}_{1-\beta}$, it holds that $L_{jj'} \geq -3\bar{d}_{1-\beta}(\Sigma)$. Consequently, we notice that, with probability at least $1 - \beta$, $\tilde{\rho}_{1-\alpha+\beta}(L)$ is bounded above by the $1 - \alpha + \beta$ quantile of:

$$\max_{j \in J} \frac{|\xi_{Y,j}|}{\sqrt{\Sigma_{Y,jj}}} \mathbb{1}\left(\xi_{X,j} - \xi_{X,j'} \geq -3\bar{d}_{1-\beta}(\Sigma) \quad \text{for all} \quad j' \in J\right) \leq \max_{j \in J} \frac{|\xi_{Y,j}|}{\sqrt{\Sigma_{Y,jj}}} \tag{49}$$

where the inequality is strict with positive probability. It follows that $\tilde{\rho}_{1-\alpha+\beta}(L) < c_{1-\alpha+\beta}\left(\hat{J}_R^+\right)$ with probability at least $1 - \beta$. ■

D Simulation Study Results

In this section, we present the results of a simulation study on our two-step approach to inference on multiple winners. We compare our two-step method to the existing approaches from section 4 and from appendix A, namely the locally simultaneous approach of Zrnic and Fithian (2024b) and test inversion approach of Zrnic and Fithian (2024a). two-step inference performs well in simulations and relative to these methods, a finding which is robust to correlation.

We consider the following designs:

- **Design 1** $J = \{1, \ldots, 5\}$, $R = \{1\}$, $\mu_Y = 0$, $\mu_X = \{\arctan(i-3)\}_{i=1}^{5}$
- **Design 2** $J = \{1, \ldots, 10\}$, $R = \{1\}$, $\mu_Y = 0$, $\mu_X = \{\arctan(i-5)\}_{i=1}^{10}$
- **Design 3** $J = \{1, \ldots, 5\}$, $R = \{1\}$, $\mu_Y = 0$, $\mu_X = 0$
- **Design 4** $J = \{1, \ldots, 5\}$, $R = \{1\}$, $\mu_Y = 0$, $\mu_X = \{\mathbb{1}(i = 1)\}_{i=1}^{5}$
- **Design 5** $J = \{1, \ldots, 5\}$, $R = \{1\}$, $\mu_Y = 0$, $\mu_X = \{\mathbb{1}(i \leq 2)\}_{i=1}^{5}$
- **Design 6** $J = \{1, \ldots, 5\}$, $R = \{1\}$, $\mu_Y = 0$, $\mu_X = \{\mathbb{1}(i \leq 3)\}_{i=1}^{5}$
- **Design 7** $J = \{1, \ldots, 5\}$, $R = \{1\}$, $\mu_Y = 0$, $\mu_X = \{\mathbb{1}(i \leq 4)\}_{i=1}^{5}$

In addition, we consider $\Sigma \in \{\Sigma_{simple}, \Sigma_{low}, \Sigma_{medium}, \Sigma_{high}\}$. For Σ_{simple}, we take $\Sigma_{simple,X} = \Sigma_{simple,Y} = \Sigma_{simple,XY} = Id$. For the remaining cases, we take:

$$\Sigma_{low,X} = \Sigma_{low,Y} = \Sigma_{low,XY} = 0.5 \cdot Id$$
$$\Sigma_{medium,X} = \Sigma_{medium,Y} = \Sigma_{medium,XY} = 0.25 \cdot Id + 0.5\mathbb{1}_p\mathbb{1}_p^\top$$
$$\Sigma_{high,X} = \Sigma_{high,Y} = \Sigma_{high,XY} = 0.025 \cdot Id + 0.95\mathbb{1}_p\mathbb{1}_p^\top$$

For each of the seven designs, we take $\Sigma \in \{\Sigma_{simple}, \Sigma_{low}, \Sigma_{medium}, \Sigma_{high}\}$ and scale Σ by $1/n$ for $n \in \{1, 10, 100, 1000, 10000\}$.

We present results from these simulations below. We denote the two-step approach to inference in red, projection in black, the zoom test (based on the step-wise implementation) in light gray, and locally simultaneous inference in dark gray.

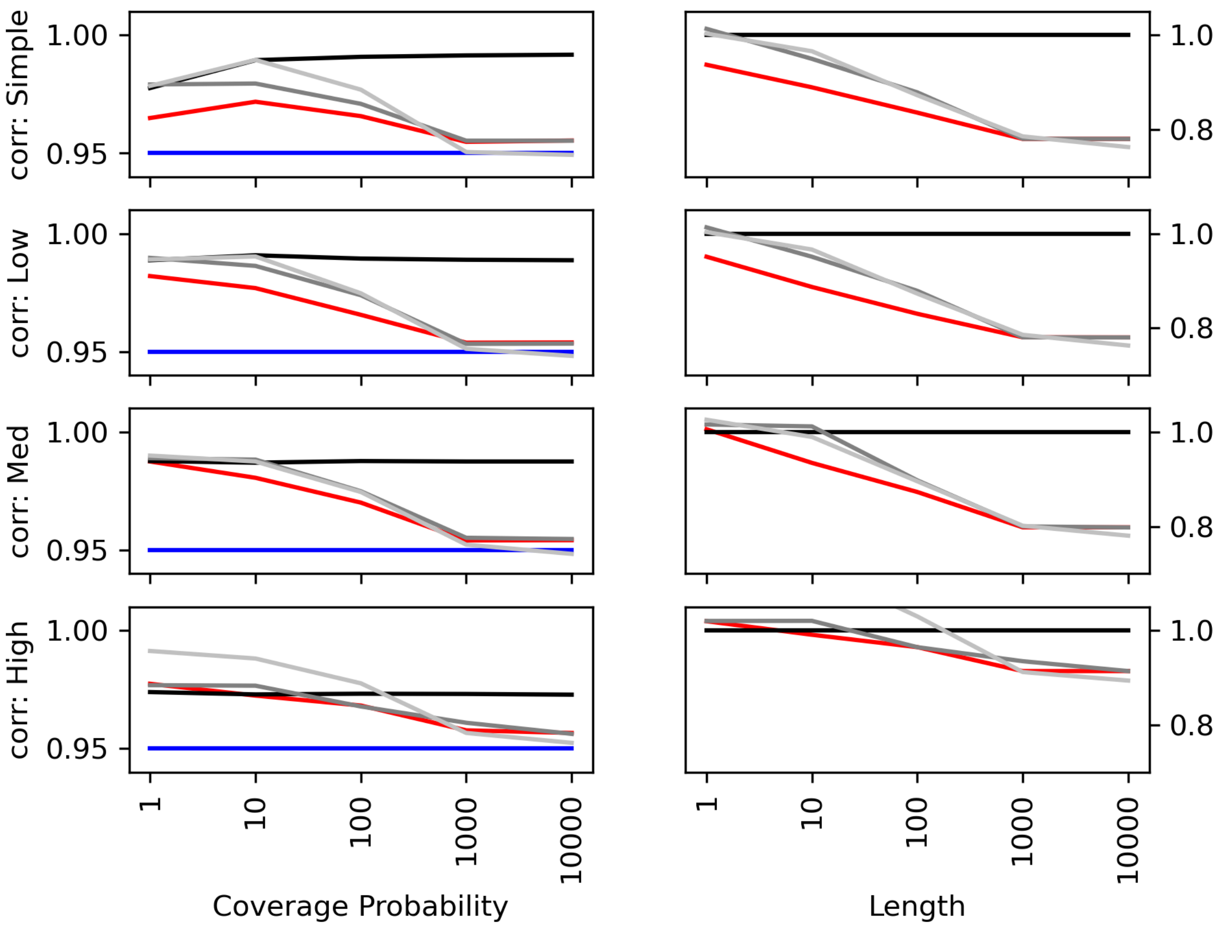


Figure 5: Coverage probability and length in design 1. CI lengths are presented as fractions of the projection CI length. We denote the two-step approach to inference in red, projection in black, the zoom test (based on the step-wise implementation) in light gray, and locally simultaneous inference in dark gray.

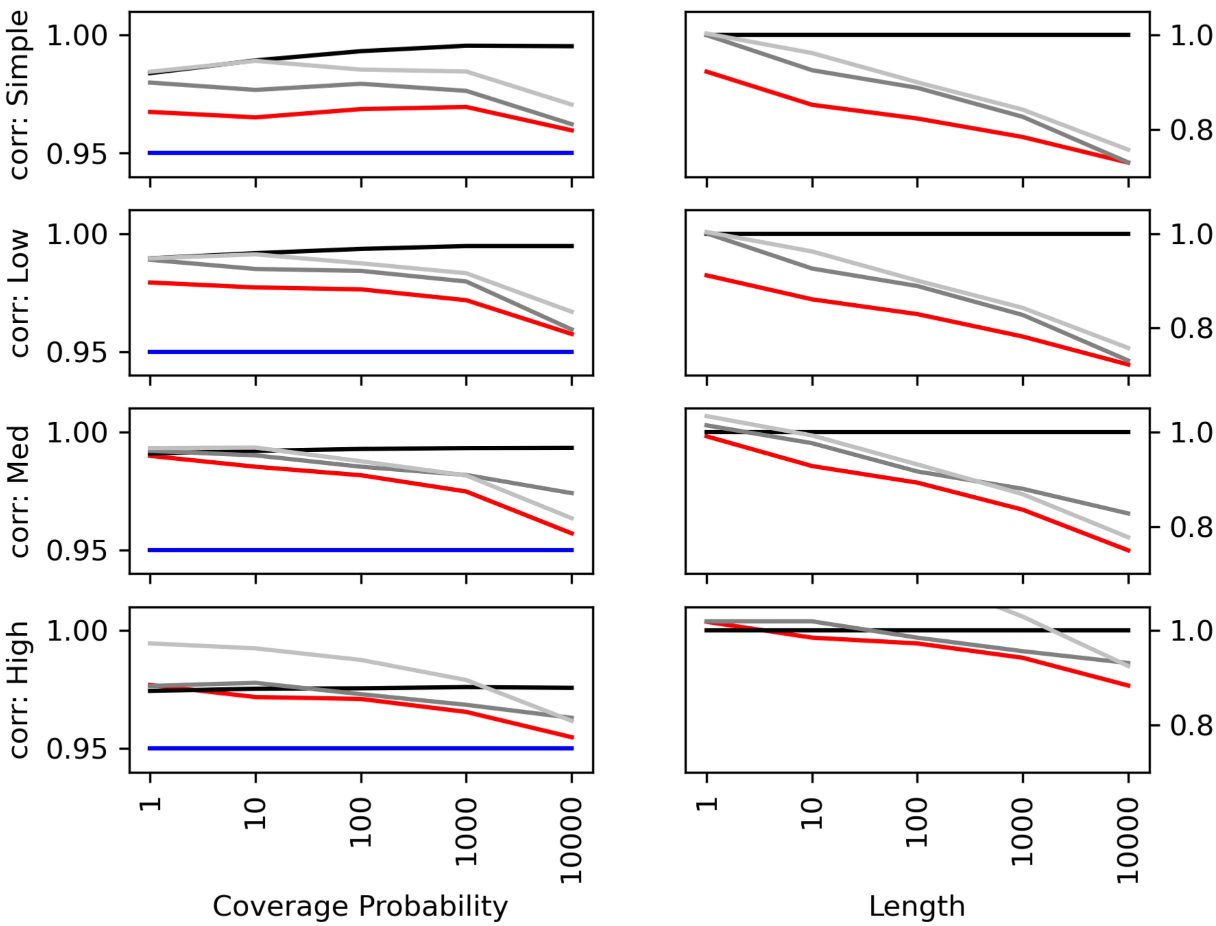


Figure 6: Coverage probability and length in design 2. CI lengths are presented as fractions of the projection CI length. We denote the two-step approach to inference in red, projection in black, the zoom test (based on the step-wise implementation) in light gray, and locally simultaneous inference in dark gray.

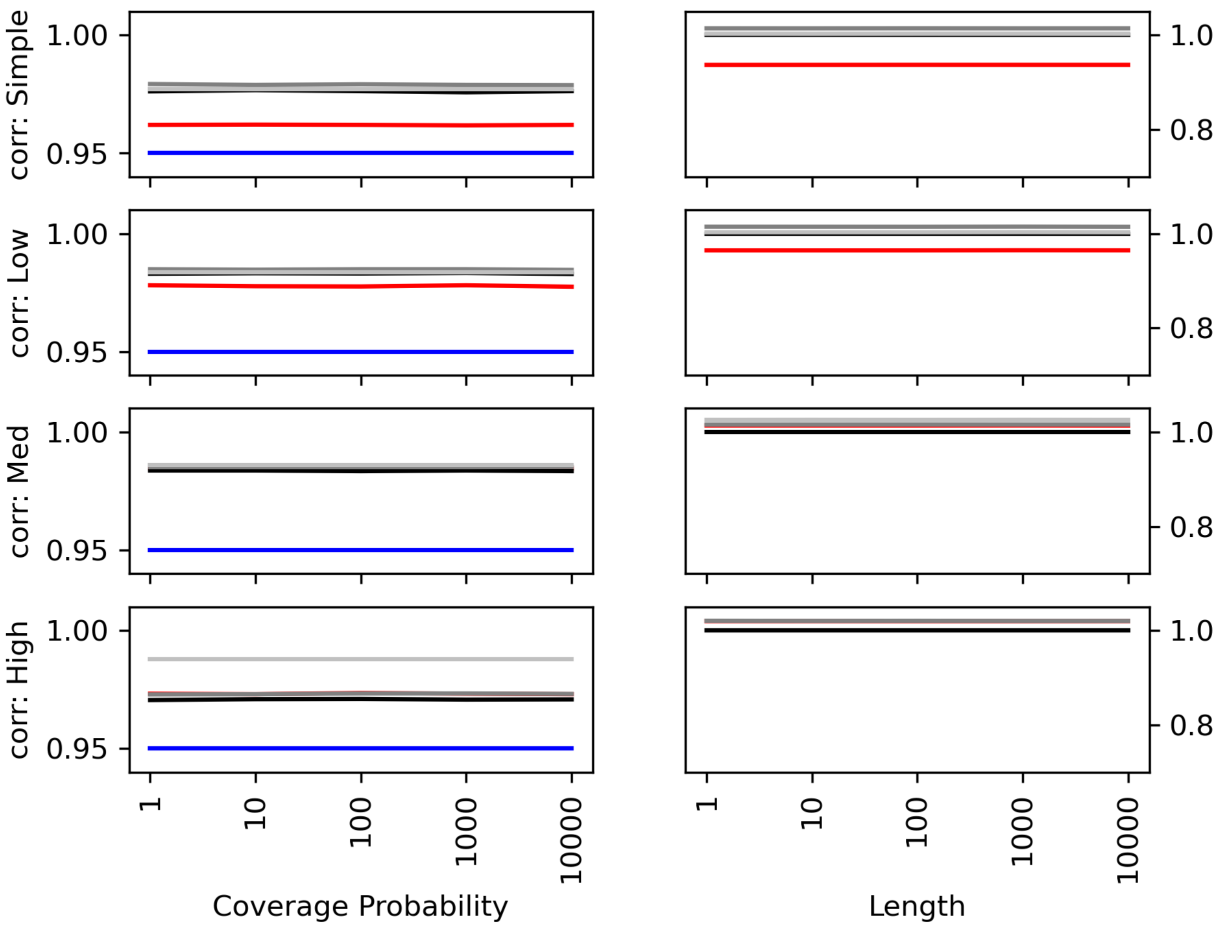


Figure 7: Coverage probability and length in design 3. CI lengths are presented as fractions of the projection CI length. We denote the two-step approach to inference in red, projection in black, the zoom test (based on the step-wise implementation) in light gray, and locally simultaneous inference in dark gray.

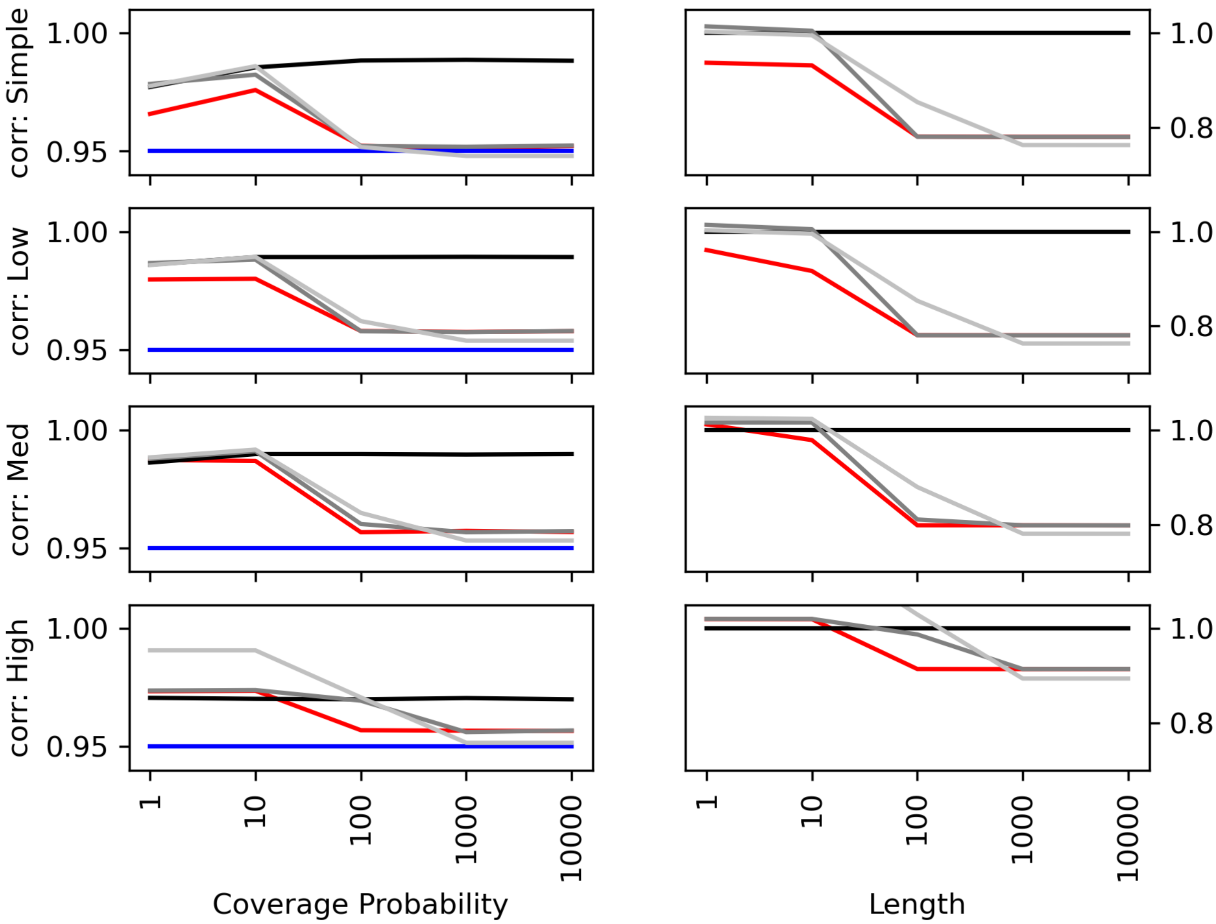


Figure 8: Coverage probability and length in design 4. CI lengths are presented as fractions of the projection CI length. We denote the two-step approach to inference in red, projection in black, the zoom test (based on the step-wise implementation) in light gray, and locally simultaneous inference in dark gray.

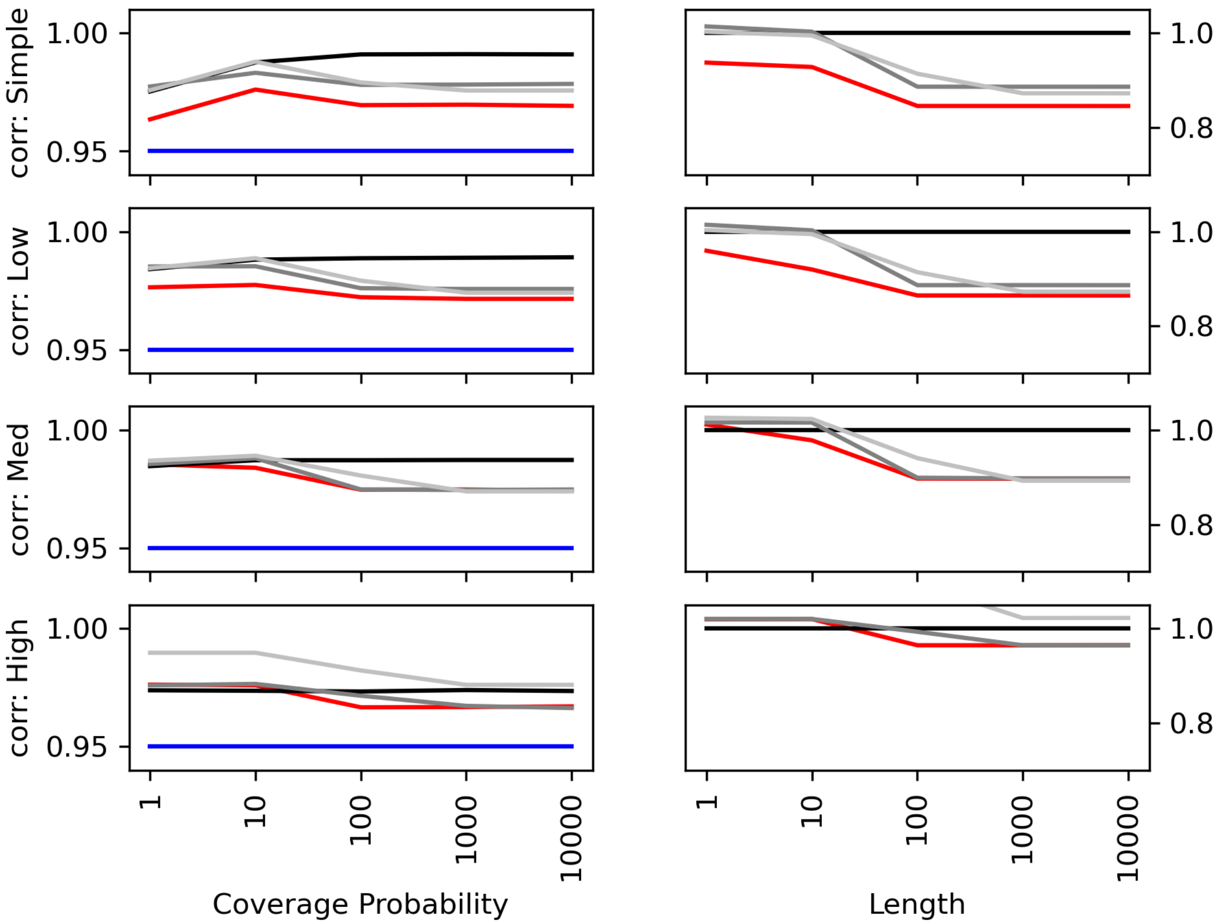


Figure 9: Coverage probability and length in design 5. CI lengths are presented as fractions of the projection CI length. We denote the two-step approach to inference in red, projection in black, the zoom test (based on the step-wise implementation) in light gray, and locally simultaneous inference in dark gray.

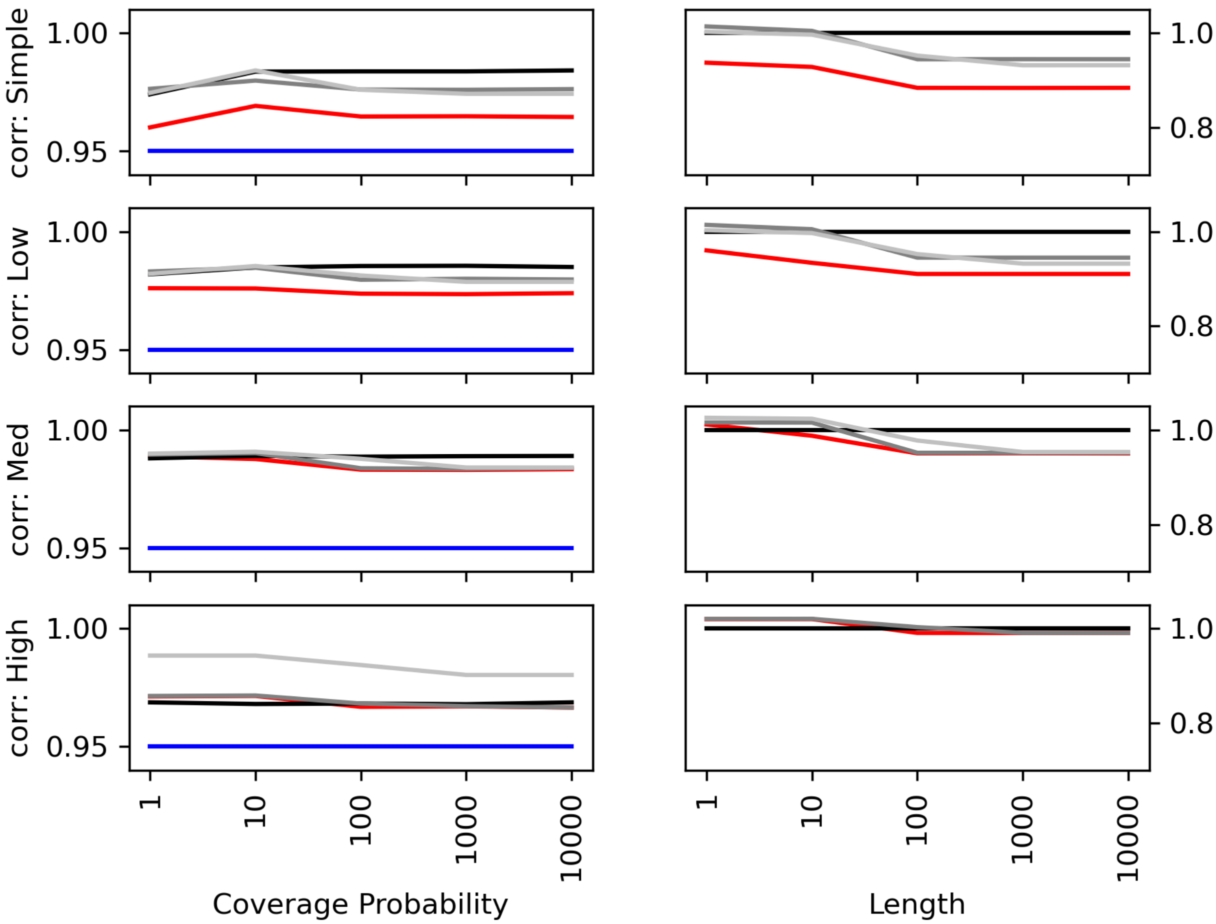


Figure 10: Coverage probability and length in design 6. CI lengths are presented as fractions of the projection CI length. We denote the two-step approach to inference in red, projection in black, the zoom test (based on the step-wise implementation) in light gray, and locally simultaneous inference in dark gray.

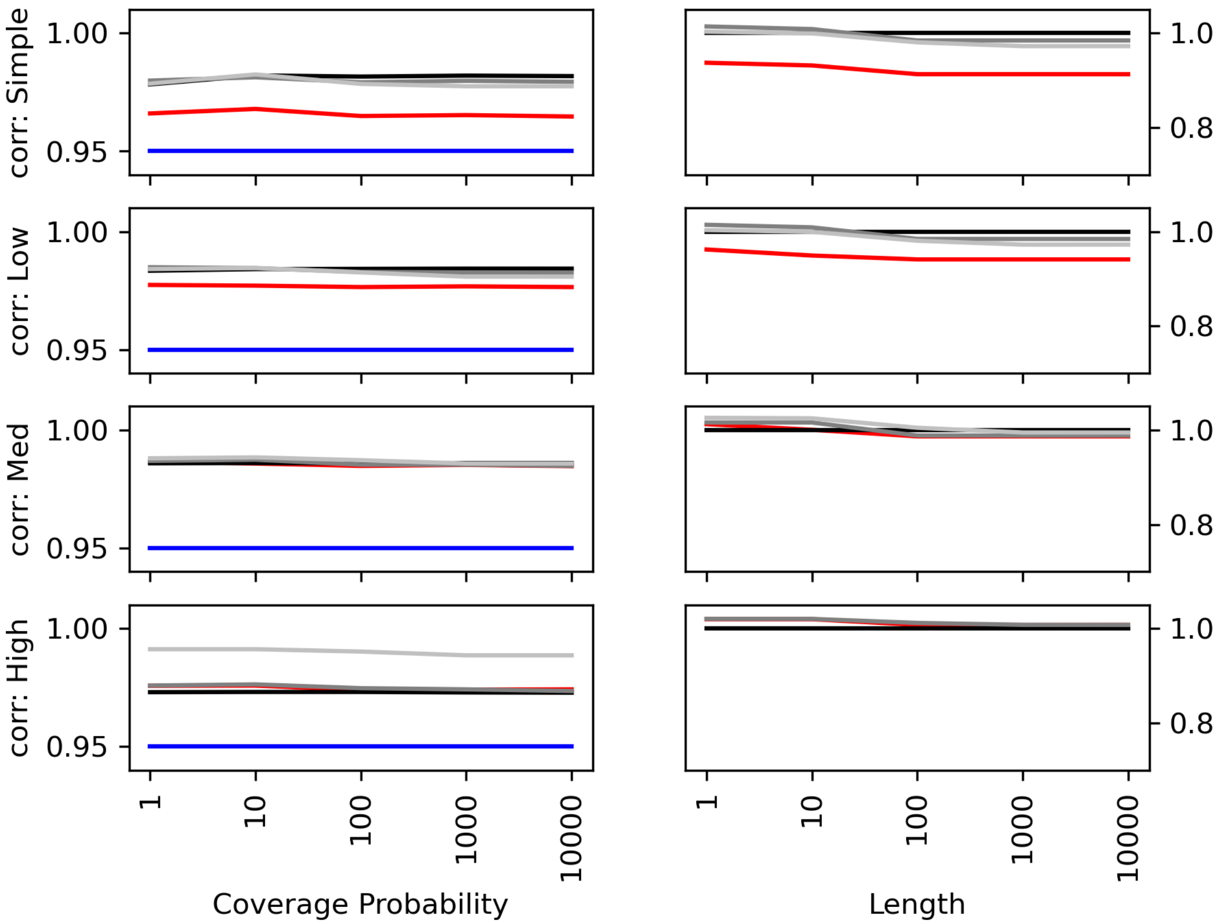


Figure 11: Coverage probability and length in design 7. CI lengths are presented as fractions of the projection CI length. We denote the two-step approach to inference in red, projection in black, the zoom test (based on the step-wise implementation) in light gray, and locally simultaneous inference in dark gray.