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

March 19, 2025

Abstract

While policymakers and researchers are often concerned with conducting inference based on a data-dependent selection, a strictly larger class of inference problems arises when considering multiple data-dependent selections, such as when selecting on statistical significance or quantiles. Given this, we study the problem of conducting inference on populations selected according to their ranks, which we dub the inference on multiple winners problem. In this setting, we encounter both selective and simultaneous inference problems, making existing approaches either not applicable or too conservative. Instead, we propose a novel, two-step approach to the inference on multiple winners problem, with the first step modeling a key nuisance parameter driving selection, and the second step using this model to derive critical values on the errors of the winners. In simulations, our two-step approach reduces over-coverage error by up to 96% relative to existing approaches. In a stylized example on job training, we demonstrate that existing approaches partially apply, and that our novel two-step approach is broadly applicable and yields informative confidence sets. In a second application, we apply our two-step approach to revisit the winner's curse in the Creating Moves to Opportunity (CMTO) program. We find that, after correcting for the inference on multiple winners problem, we fail to reject the possibility of null effects in the majority of census tracts selected by the CMTO program.

KEYWORDS: Selective Inference, Winner's Curse, Bonferroni Inequality, Uniform Validity

JEL classification codes: C12, C13

*This paper subsumes Petrou-Zeniou's undergraduate thesis. This document greatly benefited from detailed comments from Isaiah Andrews, Anna Mikusheva, and Alex Torgovitsky. The authors are grateful for discussion from Lanier Benkard, Alex Lehner, Yueran Ma, Jesse Shapiro, and Jann Speiss. The authors would also like to extend their gratitude to workshop participants at the University of Chicago and to the social sciences computing services at the University of Chicago.

1 Introduction

Policymakers and researchers frequently seek to conduct inferences based on a data-dependent selection. A recent paper by Andrews et al. (2023) discusses three confidence sets for a version of this problem, which they dub the inference on winners problem. However, a strictly larger class of problems arises when conducting inference on multiple such selections, such as when selecting on quantiles or cutoffs, which we dub the inference on multiple winners problem. In this setting, standard approaches either do not apply, or provide overly-conservative confidence sets. Given this, we propose a novel two-step approach to inference based on and related to the approaches from Romano et al. (2014), Canay and Shaikh (2017), and Zrnic and Fithian (2024b). In the first step, we model the selection of the multiple winners using a lower confidence bound on first-moment differences. In the second step, we use this lower confidence bound to model the error terms on the winner indices, using a Bonferroni-type correction to adjust for uncertainty in the lower confidence bound from the first step. Our goal in the inference on multiple winners problem is to conduct inference on the mean of some random vector Y at the indices whose ranks according to X lie in some pre-specified set R.

In a first application, we revisit the JOBSTART demonstration and subsequent replication studies, as in Cave et al. (1993) and Miller et al. (2005). While Cave et al. (1993) study a range of job retraining interventions for low-skilled workers at 13 separate sites, Miller et al. (2005) note that the one intervention with statistically significant effects on incomes from this original study failed to produce effects of the same magnitude in subsequent studies. We consider a thought experiment, where, instead of selecting the site with the largest treatment effect for replication, Miller et al. (2005) select all sites with statistically significant, positive treatment effects. As in Andrews et al. (2023), we study the possibility of a winner's curse driving this replication failure. In this setting, we suggest that, because the econometrician does not know ex-ante whether one or multiple interventions will yield statistically significant treatment effects, conditional and hybrid inference procedures do not fully apply to the JOBSTART demonstration. Revised conditional and hybrid inference procedures can apply when one intervention happens to be selected, though are generally more conservative than the approaches described in Andrews et al. (2023). Moreover, depending on the significance cutoff, we find that multiple interventions can be selected with probability ranging from 17.6% to 53.4%. This motivates our application of our two-step approach to inference in this setting. We find that our two-step approach delivers confidence sets that are 8% narrower than projection confidence sets. Our findings match those of Andrews et al. (2023) who suggest that a winner's curse does not explain the replication failure between Cave et al. (1993) and Miller et al. (2005).

The inference on multiple winners problem is particularly relevant to the Creating Moves to Opportunity (CMTO) program from Bergman et al. (2024). Bergman et al. (2024) use the CMTO program to study the barriers to movement between low-opportunity and high-opportunity neighborhoods. In particular, Bergman et al. (2024) advertise to housing-voucher recipients selected into treatment the top third of census tracts in the Seattle commuting zone, according to the tract-level opportunity atlas developed in Chetty et al. (2018). Naturally, in selecting the top third of overall tracts (and top fifth of urban tracts) by economic mobility, an inference on multiple winners problem arises. Our novel two-step approach allows us to revisit the discussion of the CMTO program from Andrews et al. (2023), who study the aggregate effects of moving into selected, high-opportunity zones via the inference on winners setting. Our generalized setting allows us to derive simultaneous confidence sets for neighborhood effects at the tract level. Tract-level effects have previously been studied, notably in Mogstad et al. (2023). While Andrews et al. (2023) find evidence for a positive, aggregate neighborhood effects among selected tracts, Mogstad et al. (2023) show that one cannot reject the possibility that even the worst-ranked track be classified as high-opportunity in the population. However, the methods of Mogstad et al. (2023) impose simultaneous coverage on all tracts, as opposed to a subset of selected tracts. This motivates our analysis of the CMTO application. Namely, we seek to study tract-level effects at tracts labeled as high-opportunity alone. While we are able to reject the possibility of null effects at some of the tracts labeled as high-opportunity by Bergman et al. (2024), we fail to do so in 92% of high-opportunity, urban Seattle tracts. Moreover, we cannot reject the possibility of a null effect on economic mobility associated with moving from the lowest opportunity tract to an any arbitrary high-opportunity urban tract. Thus, even when carefully restricting inference to tracts of interest, we can say little about tract-level effects.

Andrews et al. (2023) provide three approaches to the inference on winners problem: conditional, projection, and hybrid. The former two approaches are described in depth in Kuchibhotla et al. (2022), who provide a review of the post-selection inference literature. Conditional approaches involve conducting inference conditional on the selection event. Lee et al. (2016) provide an exact characterization of the conditional distribution of a multivariate Gaussian conditional on polyhedral selection events, allowing for exact selective inference as in Andrews et al. (2023). Most crucially, the conditional inference procedure outlined in Andrews et al. (2023) relies on finding a univariate conditional distribution for the observed, winner's outcome, with which they derive a confidence set via test-inversion. However, when conducting inference on multiple winners, we must compute p-values on a multivariate normal truncated to an arbitrary polyhedron, which is computationally challenging. The hybrid approach suffers from the exact same issues. The projection approach to inference, building on the work on simultaneous inference from Bachoc et al. (2017) and Berk et al. (2013), is overly conservative, since it is by design robust to arbitrary selection maps. The projection approach therefore fails to apply any ex-ante knowledge of the selection rule. Our two-step approach seeks to rectify this exact issue. We demonstrate both analytically and via simulation that the two-step approach to inference on multiple winners outperforms the projection approach. Relative to projection, the two-step approach reduces over-coverage error by up to 96.6%, and reduces confidence set volume by up to 34.5%, matching our theoretical results on the asymptotic performance of our two-step confidence sets.

The paper is organized as follows: In section 2, we formally introduce the inference on multiple winners problem and present three empirically-relevant, specialized settings. In section 3, we revisit the JOBSTART demonstration and compare different approaches to inference. In section 4, we provide a more detailed discussion of the approaches to the inference on winners problem, and demonstrate that they may fail in the inference on multiple winners problem. In section 5, we construct the two-step approach to inference on multiple winners, and provide analytical results comparing its performance to the projection approach 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 **??**, 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 provide proofs, supplemental results, and a two-step approach to an important generalization of the inference on multiple winners problem. We also compare our two-step approach to existing methods from the selective inference literature, showing favorable performance.

2 Setup and Notation

In this section, we formalize the inference on multiple winners problem. We retain the notation from Andrews et al. (2023) wherever possible. We consider a finite set of indices or populations $J := \{1, \ldots, p\}$, the p-dimensional random vector Y, as well as a correlated random vector X, which are drawn jointly from a multivariate normal as below:

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

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 and define the number of selections to be $k := |R|$. We let the indices $j_1, \ldots, j_p$ order X such that $X_{j_1} \geq X_{j_2} \geq \ldots \geq X_{j_p}$. For example, j_1 denotes the index corresponding with the largest realization in X, j_2 the second largest, and so on. We define the set of selected indices $\hat{J}_R$ as follows:

$$\hat{J}_R := \{j_l; l \in R\} \tag{2}$$

If we take R to simply be J, then $\hat{J}_R$ would be the full index set J. Similarly, taking $R := \{1\}$, our selected indices would simply be the singleton $\{j_1\}$.

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

$$P_{\mu,\Sigma}\left(\mu_{Y,\jmath} \in CS_\jmath \quad \text{for all} \quad \jmath \in \hat{J}_R\right) \geq 1 - \alpha \tag{3}$$

for all μ and Σ. When $R = \{1\}$, this problem is equivalent to the inference on single winners problem.

Remark 2.1. Andrews et al. (2023) consider the case of inference on the winner, where $R = \{1\}$. Relatedly, 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. Finally, our approach easily generalizes to the case of multiple Y, as we discuss below. ■

2.1 Review of Applications

We note that, although the inference on multiple winners setting described above represents a narrow set of inference problems, it can be readily modified and generalized to incorporate a broad array of empirically-relevant settings, just as the inference on winners problem from Andrews et al. (2023). We demonstrate that many of the same applications considered in Andrews et al. (2023) remain relevant in the inference on multiple winners setting, and, additionally, a range of novel applications are also immediately tractable via inference on multiple winners approaches.

Post-Selection Inference on Quantiles: Suppose we want to conduct inference on the means of Y when selecting the top γ-quantile of observations in some random vector X. We can take $\tau := \lfloor \gamma p \rfloor$ and take $R = \{1, \ldots, \tau\}$. 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, by the proportion of thatched roofs. We may think of these proportions as drawn from a superpopulation, and seek to study superpopulation parameters at selected villages.

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. We can append a constant vector to our X take $X_c = \begin{pmatrix} X' & c\mathbb{1}'_p \end{pmatrix}'$, and $Y_c = \begin{pmatrix} Y' & c\mathbb{1}'_p \end{pmatrix}'$. We can take $R := \{1, \ldots, p\}$, 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 values 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. While Andrews et al. (2023) consider cutoff-based selections, their framework entails performing inference on a univariate normal truncated according to a given cutoff, rather than the general case of cutoff-based selection described above. Cutoff-based selection rules have also attracted interest in the econometrics literature, namely in Gu and Koenker (2023).

Inference on Statistical Significance: We notice, 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. Moreover, we can readily apply the two generalized inference on multiple winners settings as above.

Inference on Multiple Outcomes: We may be concerned with 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 ethnic or racial group, or by gender). If we are

[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 $(Y_j)_{j \in \hat{J}_R}$ is invariant to our choice of $\hat{J}_R$.

concerned with the means of several vectors $Y_1, \dots, Y_K$ at indices selected from X according to some set of ranks R. We can take $Y_r := \begin{pmatrix} Y_1' & \dots & Y_K' \end{pmatrix}'$ and take $X_r := \begin{pmatrix} X' & \dots & X' \end{pmatrix}'$ to be the vector X repeated K times. We can take a new index set $R_r := \{lk; l \in R,\ 1 \leq k \leq K\}$. Thus, Y_r, X_r, and R_r characterize the inference on multiple winners problem with the desired estimands.

3 Inference on Multiple Winners in 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 in 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 the value-add of the program. 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 particular, we consider the possibility that 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 econometrician 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. (2022) are difficult to apply, as we discuss in section 4.1. When only one selection is made, one must consider an alternative selection rule to that considered by Andrews et al. (2023). As a result, we suggest that our methods are better-suited for inference in this particular setting.

Empirically, we show that our two-step confidence regions for the effect of the CET program on earnings in the Cave et al. (1993) study are wider under this alternative selection rule than under the selection rule considered by Andrews et al. (2023). Nevertheless, as in Andrews et al. (2023), these confidence regions still 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). Our two-step confidence regions provide more informative inference than projection confidence regions in this application, To be precise, our two-step confidence regions are between 6% and 8% shorter than the corresponding projection confidence regions. Hybrid and conditional inference tend to provide narrower confidence sets than the two-step approach, though only in cases where a single site is selected for replication. Finally, we study 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 probability of between 17.6% and 53.4%, depending on our choice of significance level by which to select sites for replication.

3.1 JOBSTART: Empirical Findings

The JOBSTART demonstration compared 13 distinct implementations of a job training and retraining program targeted towards low-skilled school dropouts at different sites. Out of these 13 interventions, only one intervention, the CET at San Jose, had a statistically significant effect on earnings, thus motivating the replication in Miller et al. (2005). 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

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 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.[2] In the thought experiment where selection for replication is based on statistical significance, the conditional inference procedures from Andrews et al. (2023) must be revised. However, the econometrician does not know, ex-ante, how many studies will be selected for replication as statistically significant. Consequently, the econometrician will choose to apply conditional inference only when one study is statistically significant. As a result, the conditioning event chosen for conditional inference must account for the fact that the econometrician's choice of inference procedure is itself data-dependent. That is, we condition on the event that only the winner $\jmath$ in J is statistically significant at a given level. Indeed:

$$Y_\jmath/\sqrt{\Sigma_{Y,\jmath\jmath}} \geq c, \quad Y_j/\sqrt{\Sigma_{Y,jj}} \leq c \quad \text{for all } j \neq \jmath \tag{4}$$

We can construct a confidence set conditionally valid for the event 4 using the approaches developed in Andrews et al. (2023), since their propositions 1-3 and their proposition 7 can easily be applied to accomodate selection events of the form above. We present our empirical findings using our novel two-step method, the projection approach, and both the original and revised conditional and hybrid approaches. We find that the two-step approach provides confidence sets that are between 5% and 8% narrower than the projection approach. A back of the envelope calculation gives that our two-step confidence region increases the lower bound on total effects of the replication of Miller et al. (2005) by approximately $122,000. The revised conditional and hybrid approaches provide narrower-still confidence sets, but may be of limited use in the JOBSTART application due to the frequency with which we may observe statistically significant outcomes for multiple interventions.

Table 2: Corrected Confidence Regions for CET Treatment Effects

Method	CS: 5% Significance Cutoff
Two-Step	[$2476, $10618]
Two-Step (Asymmetric)	[$2191, $10114]
Projection	[$2237, $10857]
Original Conditional	[$3485, $9478]
Revised Conditional	[$2777, $9477]

Confidence sets for the CET program, correcting for cutoff-based selection. Confidence sets are presented for the 5% significance cutoff.

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 discussed in 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

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

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:

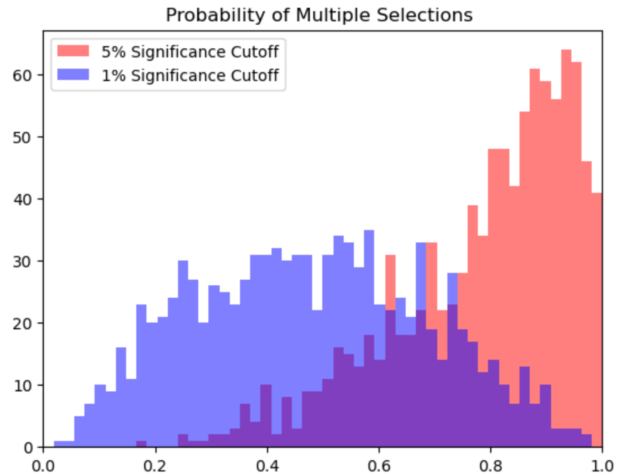


Figure 1: Probability of making multiple selections for different significance-based selection criteria. Probabilities are computer over 1000 simulation draws.

4 Existing Approaches to Inference

We now turn our attention to the existing approaches to such post-selection inference problems as above. Generally, these approaches can be divided into three classes: 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 simultaneous approach involves conducting inference that is simultaneously valid for all selection methods, as in Bachoc et al. (2017) and Berk et al. (2013). We demonstrate, in this section, that the former, conditional approach is not appropriate in the inference on multiple winners setting, while the latter approach is too conservative. In appendix A, we discuss further, alternative approaches to inference, namely locally simultaneous approaches based Zrnic and Fithian (2024b) and approaches based on inverting the zoom test of Zrnic and Fithian (2024a). In general, we find that our two-step approach outperforms most existing approaches in simulation for a broad range of data generating processes.

4.1 The Conditional Approach to Inference

The conditional inference approach outlined in Andrews et al. (2023) is an example of conditional selective inference common in the statistics literature, and 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 selection events. We find that the exact distribution of the collection of the $(Y_\jmath)_{\jmath \in \hat{J}_R}$, conditional on the selection event for the $\hat{J}_R$, is a multivariate normal truncated to a polyhedron. Given the numerical challenges of computing p-values and the associated challenges in applying test inversion to derive a confidence region for the $(\mu_{Y,\jmath})_{\jmath \in \hat{J}_R}$, conditional inference is inappropriate for the inference on multiple winners problem.

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. Moreover, we condition on a sufficient statistic Z for the nuisance parameters associated with the elements of μ not corresponding to the Y_{j_l}. That is, our goal is derive a confidence set $CS^c_{1-\alpha}$ such that:

$$P_{\mu,\Sigma}\left(\mu_{Y,j_l} \in CS^c_{1-\alpha,j_l} \quad \text{for all} \quad j_l \in \hat{J}_R \;\middle|\; (j_l)_{l\in R} = (i_l)_{l\in R},\ z\right) \geq 1-\alpha \tag{5}$$

for all μ. By the law of iterated expectations, such a confidence set satisfied 3 as well. Let e_i be a standard basis vector. To simplify the problem of characterizing selection, let R be exactly the set $\{1,\ldots,\tau\}$. Now, let B be the following matrix:

$$B' := \begin{pmatrix} & e_{i_1} \\ & e_{i_2} \\ \mathbf{0}_{k\times p} & \vdots \\ & e_{i_\tau} \end{pmatrix}$$

noticing that $B'\begin{pmatrix} X' & Y' \end{pmatrix}' = \begin{pmatrix} Y_{i_1} & Y_{i_2} & \ldots & Y_{i_\tau} \end{pmatrix}'$. Similarly, we define $c := \Sigma B(B'\Sigma B)^{-1}$ and take $Z := (I - cB')\begin{pmatrix} X' & Y' \end{pmatrix}'$. Finally, we obtain a polyhedral characterization of the selection event $(j_l)_{l\in R} = (i_l)_{l\in R}$ as follows:

$$\begin{pmatrix} (e'_{i_1} - e'_i)_{i\neq i_1} & 0 \\ (e'_{i_2} - e'_i)_{i\neq i_1, i_2} & 0 \\ \vdots & \vdots \\ (e'_{i_\tau} - e'_i)_{i\neq i_1,\ldots,i_\tau} & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} \geq 0$$

We define A to be the matrix on the left hand side.[3] We obtain the following result as an extension of the polyhedral lemma 5.1 from Lee et al. (2016):

Lemma 4.1. *The collection $(Y_\jmath)_{\jmath\in\hat{J}_R}$, 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:*

$$(Y_\jmath)_{\jmath\in\hat{J}_R} | Z = z, j_l = i_l \textit{ for } l \in R \sim TN_{\mathcal{Y}((i_l)_{l\in R}, z)}(B'\mu, B'\Sigma B) \tag{6}$$

We observe $Z = z$ and $Y_{j_l} = y_{j_l}$ for l in R. Given this, we construct the following testing procedure:

$$\begin{aligned} &\phi\left((y_{i_l})_{l\in R}; (i_l)_{l\in R}, z, (\mu_{Y,i_l})_{l\in R}\right) \\ &= \begin{cases} 1 & \text{if} \quad P_{\mu,\Sigma}\left(\left\|(Y_{j_l} - \mu_{Y,j_l})_{l\in R}\right\| \geq \left\|(y_{i_l} - \mu_{Y,i_l})_{l\in R}\right\| | Z = z, j_l = i_l \text{ for all } l \in R\right) \leq \alpha \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

The above test ϕ tests the simple null hypothesis that the true means of the $(Y_{i_l})_{l\in R}$ are $(\mu_{Y,i_l})_{l\in R}$ at level α. We can construct a confidence set for the $(\mu_{Y,i_l})_{l\in R}$ by inverting ϕ - that is by taking the set of

[3] 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 $R' := \{1,\ldots,\tau\}$. 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).

all $(\mu_{Y,i_l})_{l\in R}$ such that, given our data, we fail to reject. We denote this confidence set by $CS^c_{1-\alpha}$. 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, (5) 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, $Y_\jmath$ is distributed according to a univariate normal truncated to an interval for $\jmath$ 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 parameters $(\mu_{Y,i_l})_{l\in R}$ 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. In this section, we focus on simultaneous inference as developed in Andrews et al. (2023).

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:

$$\max_{j\in J_c}\frac{|\xi_{Y,j}|}{\sqrt{\Sigma_{Y,jj}}} \tag{7}$$

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

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

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:

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

Naturally, $CS^P_{1-\alpha}$ is a valid confidence set at the $1-\alpha$-level satisfying 3. 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. Nevertheless, the projection confidence set may be appropriate in settings where tight confidence intervals are not entirely necessary, or in cases where k is large.

[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.

4.3 Further Approaches to Inference

5 The Two Step Approach to Inference on Multiple Winners

In this section, we construct a two-step approach to inference on multiple winners that corrects for some of the shortcomings of the projection approach to inference. 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). Our approach is meaningfully different to that of Zrnic and Fithian (2024b), and is generally more powerful, as demonstrated in our simulation study in section 7 and appendix E. In the first step, we construct a lower bound on moment differences. In the second step, we use this lower bound to model selection, deriving a less conservative, Bonferroni-type critical value. While the projection approach tends to over-cover in cases where there is a clear winner, our two-step approach tends to perform quite well in such circumstances, which we demonstrate both theoretically and via simulation. Moreover, across a general range of parameters μ and Σ, the two-step approach performs, in a worst case, roughly equivalently to the projection approach. Thus, we recommend using the two-step approach in lieu of the projection approach.

5.1 Construction

Our goal, as usual, is to construct a rectangular confidence set $CS^{TS}_{1-\alpha}$ satisfying 3, where $CS^{TS}_{1-\alpha} = \times_{j \in \hat{J}_R} CS^{TS}_{1-\alpha,j}$. Our construction is a consequence of the observation below:

$$\frac{|\xi_{Y,\jmath}|}{\sqrt{\Sigma_{Y,\jmath\jmath}}} \leq \max_{j \in J} \frac{|\xi_{Y,j}|}{\sqrt{\Sigma_{Y,jj}}} \mathbb{1}\left(\sum_{j' \in J} \mathbb{1}\left(\xi_{X,j} \geq \xi_{X,j'} + \Delta_{j',j}\right) \in R \right) \tag{8}$$

for all $\jmath \in \hat{J}_R$. We define $\Delta_{j',j} := \mu_{X,j'} - \mu_{X,j}$, which 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 left hand side of (8). 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 $d_{1-\beta/2}(\Sigma)$ to be the $1 - \beta/2$-quantile of the following:

$$\max_{j,j' \in J, j \neq j'} \frac{|\xi_{X,j} - \xi_{X,j'}|}{\sqrt{\operatorname{Var}(\xi_{X,j} - \xi_{X,j'})}} \tag{9}$$

Because d is studentized, we have that d is homogeneous of degree zero in Σ. We define L and U as follows:

$$L_{j',j} := X_{j'} - X_j - d_{1-\beta/2}(\Sigma)\sqrt{\operatorname{Var}(\xi_{X,j} - \xi_{X,j'})} \tag{10}$$

$$U_{j',j} := X_{j'} - X_j + d_{1-\beta/2}(\Sigma)\sqrt{\operatorname{Var}(\xi_{X,j} - \xi_{X,j'})} \tag{11}$$

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

$$\frac{|\xi_{Y,\jmath}|}{\sqrt{\Sigma_{Y,\jmath\jmath}}} \leq \max_{j\in J} \frac{|\xi_{Y,j}|}{\sqrt{\Sigma_{Y,jj}}} \mathbb{1}\left(\left[\sum_{j'\in J} \mathbb{1}\left(\xi_{X,j} \geq \xi_{X,j'} + U_{j',j}\right), \sum_{j'\in J} \mathbb{1}\left(\xi_{X,j} \geq \xi_{X,j'} + L_{j',j}\right)\right] \cap R \neq \emptyset\right) \tag{12}$$

on the event $B := \{L \leq \Delta \leq U\}$. We take $1-\alpha+\beta$ quantile of the right hand side above, conditional on L and U. Thus, the distribution of the right hand side of 12 is a function of L and U. Similarly, the corresponding quantile function is a function of L and U. Thus, we denote the $1-\alpha+\beta$-quantile of the right hand side of 12 by $\rho_{1-\alpha+\beta}(L,U)$. We define the following, rectangular confidence interval:

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

We claim the following:

Proposition 5.1. *$CS^{TS}_{1-\alpha}$ is a valid confidence set at the $1-\alpha$-level, such that:*

$$P_{\mu,\Sigma}\left(\left(\mu_{Y,\jmath}\right)_{\jmath\in\hat{J}_R} \in CS^{TS}_{1-\alpha}\right) \geq 1-\alpha \tag{13}$$

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

$$P_{\mu,\Sigma}\left(\left(\mu_{Y,\jmath}\right)_{\jmath\in\hat{J}_{R'}} \in CS^{TS}_{1-\alpha}\right) \geq 1-\alpha \tag{14}$$

for all μ, Σ.

We prove this result in appendix C. We provide the following non-inferiority result in the spirit of theorem 2 of Zrnic and Fithian (2024b). 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:

$$\max_{j\in J_c} \max\left\{\frac{|\xi_{Y,j}|}{\sqrt{\Sigma_{Y,jj}}}, \frac{|\xi_{X,j}|}{\sqrt{\Sigma_{X,jj}}}\right\} \tag{15}$$

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:

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

We notice that when we select on the Y, such that $X = Y$, then $CS^{TS2}_{1-\alpha} \subseteq CS^{P}_{1-\alpha}$, providing a powerful non-inferiority result. We provide the following proposition:

Proposition 5.2. *Suppose β is sufficiently small such that $2\bar{c}_{1\alpha} \leq d_{1-\beta}(\Sigma)$. Then $CS^{TS2}_{1-\alpha}$ is a valid confidence set at the $1-\alpha$-level, such that:*

$$P_{\mu,\Sigma}\left(\left(\mu_{Y,\jmath}\right)_{\jmath\in\hat{J}_R} \in CS^{TS2}_{1-\alpha}\right) \geq 1-\alpha \tag{16}$$

for all μ and Σ.

We prove this result in appendix C.

Remark 5.1. In the case where R is of the form $\{1, \ldots, \tau\}$, we can simplify our two-step approach to inference. In particular, we can write 8 as follows:

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

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

$$\frac{|\xi_{Y,\jmath}|}{\sqrt{\Sigma_{Y,\jmath\jmath}}} \leq \max_{j \in J} \frac{|\xi_{Y,j}|}{\sqrt{\Sigma_{Y,jj}}} \mathbb{1}\left(\sum_{j' \in J} \mathbb{1}\left(\xi_{X,j} \geq \xi_{X,j'} + \tilde{L}_{j',j}\right) \geq \tau\right).$$

Allowing $\tilde{\rho}_{1-\alpha_\beta}\left(\tilde{L}\right)$ to be the $1 - \alpha + \beta$ quantile of the right hand side treating $\tilde{L}$ as fixed. In this case, we can write:

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

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

Remark 5.2. The observation in 8 generalizes to the positive and negative components of the errors, which we may denote by ξ^+_Y and ξ^-_Y, such that for $\jmath \in \hat{J}_R$:

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

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 12 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 $1 - \frac{\alpha-\beta}{2}$-quantile of the right hand side of 12 with the same, respective, substitutions. We can consequently take the following, asymmetric two-step confidence set:

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

This confidence set matches the intuition developed in Andrews et al. (2023) that the observation corresponding to the winner is upwardly biased, since $\rho^+_{1-\frac{\alpha-\beta}{2}}(L, U)$ will generally be larger than $\rho^-_{1-\frac{\alpha-\beta}{2}}(L, U)$. This also presents a key advantage of our approach relative to that of Zrnic and Fithian (2024b), since we flexibly model the errors on the winner, rather than modeling the set of likely winners and restricting simultaneous inference therein. Finally, the reasoning of proposition 5.2 generalizes to this setting, such that we can take:

$$CS^{ATS2}_{1-\alpha} := \begin{cases} \times_{\jmath \in \hat{J}_R} \left[Y_\jmath - \bar{c}_{1-\alpha}\sqrt{\Sigma_{Y,\jmath\jmath}},\ Y_\jmath + \bar{c}_{1-\alpha}\sqrt{\Sigma_{Y,\jmath\jmath}}\right] & \text{if} \quad \rho^-_{1-\frac{\alpha-\beta}{2}}(L, U), \rho^+_{1-\frac{\alpha-\beta}{2}}(L, U) \geq \bar{c}_{1-\alpha} \\ CS^{TS}_{1-\alpha} & \text{otherwise} \end{cases}$$

meaning that we can recover some of the non-inferiority properties of the standard two-step approach to inference. ■

Remark 5.3. Our approach to inference can be generalized to arbitrary polyhedral selections, in the spirit of Lee et al. (2016). This is because polyhedral selection is linearly separable in errors ξ and means μ, much as in equation 8. 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 $\mathcal{P} = \{\mathcal{P}_i\}_{i\in\mathcal{I}}$ where we take $\mathcal{P}_i := \{x; A_i x \leq b_i\}$. We take a collection of $k \times p$ matrices B_i such that, whenever we observe X satisfying $X \in \mathcal{P}_i$, we select $B_i Y$. We denote the rows of B_i by b_i, j' for $j = 1, \ldots, k$. In general, we say that we observe $\hat{B}Y$, where $\hat{B}$ is selected as above. We denote the rows of $\hat{B}$ by $\hat{b}_j'$ for $j = 1, \ldots, k$:

$$\frac{\left|\hat{b}_j' \xi_Y\right|}{\sqrt{\operatorname{Var}\left(\hat{b}_j' \xi_Y\right)}} \leq \max_{i\in\mathcal{I}} \frac{|b_i' \xi_Y|}{\sqrt{\operatorname{Var}\left(b_i' \xi_Y\right)}} \mathbb{1}\left(A_i \xi_X + A_i \mu_X \leq b_i\right) \tag{17}$$

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 a $1-\beta$-level confidence region for these nuisance parameters, and subsequently constructing the $1-\alpha+\beta$-quantile of the right hand side 17 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. ■

Remark 5.4. 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 equation 12 as follows:

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

It is clear that our two-step technique can be applied to study the pairwise differences $\mu_{Y,\jmath} - \mu_{Y,\imath}$ for all pairs of $\jmath$ in $\hat{J}_R$ and $\imath$ in $\hat{J}_{R'}$. Moreover, as Mogstad et al. (2023) demonstrate, the problem of constructing confidence sets for the ranks of units is equivalent to constructing confidence regions for pairwise difference in unit means. 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. ■

5.2 Feasible Inference and Analytical Results Comparing Projection and Two-Step Approaches

In this section, we compare the projection and two-step approaches to inference analytically in a nonparametric setting. Indeed, we show that, when conducting inference on sample means of independently and identically distributed random vectors, the two-step approach asymptotically outperforms the projection approach in cases where there exists an asymptotically clear winner. We also show that, in a worst case, the two-step approach performs, asymptotically, only marginally worse than the projection approach.

First, we construct the setting for inference on winners with sample means. 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 $\tilde{S}_{W,n} := \begin{pmatrix} \tilde{S}'_{X,n} & \tilde{S}'_{Y,n} \end{pmatrix}'$. For all P in $\mathcal{P}$, we define, for arbitrary $j = 1, \ldots, k$:

$$\mu_X(P) := \mathbb{E}_P\left(\tilde{X}_1\right)$$

with the analogous definition for $\mu_Y(P) := \mathbb{E}_P\left(\tilde{Y}_1\right)$ and $\mu_W(P)$. 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, for arbitrary n, as follows:

$$\Sigma(P) := \mathbb{E}_P\left(\left(\xi_{\tilde{X},1}{}', \xi_{\tilde{Y},1}{}'\right) \cdot \left(\xi_{\tilde{X},1}, \xi_{\tilde{Y},1}\right)'\right) \tag{18}$$

We similarly have some sequence of estimators for $\Sigma(P)$ which we denote by $\hat{\Sigma}^n$. We can denote $\Sigma(P)$ and $\Sigma_W(P)$. In general, we can take $\hat{\Sigma}^n = \Sigma(\hat{P}_n)$, 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 such the rank of $\tilde{S}_{X,n,i}$ in the vector $\tilde{S}_{X,n}$ is contained in R. Our first goal is to demonstrate the following uniform asymptotic validity result:

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} Pr_P\left(\mu_{Y,\jmath}(P) \in CS^{TS}_{1-\alpha}(\hat{\Sigma}^n) \quad \text{for all} \quad \jmath \in \hat{J}_R\right) \tag{19}$$

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 several assumptions on the properties of $\mathcal{P}$. First, 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. Moreover, in order to conduct feasible inference, we suppose that we have access to uniformly consistent estimators $\hat{\Sigma}^n$ for $\Sigma(P)$, which we formally state in assumption B.2.

Uniform asymptotic validity establishes the empirical relevance of the two-step approach to inference on multiple winners. Indeed, with only a fairly weak set of assumptions on the set of underlying distributions, we can make uniform statements on the convergence of the two step confidence set's coverage probability. Thus, we can feasibly apply our two step confidence set to finite sample settings, without the concern that the underlying data generating process yields slow convergence, which would invalidate the use of normality-based inference procedures in empirical work.

For our result on pointwise, asymptotic superirority, we impose an assumption on the pointwise rate of convergence for our variance-covariance estimator $\hat{\Sigma}^n$, essentially stating that $\hat{\Sigma}^n$ cannot converge at an arbitrarily slow rate. We formalize this in assumption C.1.

Remark 5.5. We also define $J_R(P) := \left\{j; \sum_{j'\in J} \mathbb{1}\left(\mu_{X,j'}(P) \geq \mu_{X,j}(P)\right) \in R\right\}$. We can show that $\jmath$ in $\hat{J}_R$ will lie in $J_R(P)$ eventually with probability one, as all $\tilde{S}_{X,n}$ converge almost surely, and jointly, to $\mu_X(P)$ for P fixed as n frows large. That is, by the strong law of large numbers:

$$\begin{pmatrix} \tilde{S}_{X,n} \\ \tilde{S}_{Y,n} \end{pmatrix} \xrightarrow{a.s.} \begin{pmatrix} \mu_{X,n}(P) \\ \mu_{Y,n}(P) \end{pmatrix}$$

for any P in $\mathcal{P}$. ■

This remark provides the intuition for our result on pointwise asymptotic superiority. Indeed, we demonstrate that:

Proposition 5.3. *Suppose that, for P in $\mathcal{P}$, the set of true winners $J_R(P)$ is a proper subset of J. Under assumptions B.1, B.2, and C.1 the coverage probability of the two-step confidence set is pointwise, asymptotically smaller than that of the projection confidence set, such that there exists β sufficiently small satisfying:*

$$\lim_{n\to\infty} Pr_P\left((\mu_{Y,\jmath}(P))_{\jmath\in\hat{J}_{R;n}} \in CS^{TS}_{1-\alpha;\beta,n}\right) \leq \lim_{n\to\infty} Pr_P\left((\mu_{Y,\jmath}(P))_{\jmath\in\hat{J}_{R;n}} \in CS^{P}_{1-\alpha;n}\right) \tag{20}$$

We provide the proof for this proposition in appendix C.

Remark 5.6. One may note that the two-step approach to inference on multiple winners can be integrated with the hybrid approach from Andrews et al. (2023) in the single winner case. We recommend using an updated hybrid approach to inference, since the two-step approach, constructed in section 5, generally outperforms the projection approach, thanks to both propositions 5.2 and 5.3. Instead of conditioning on the parameter of interest $\mu_{Y,\jmath}$, where $\jmath$ is the unique element in $\hat{J}_{\{1\}}$, lying in the projection confidence set, we recommend conditioning on the event that $\mu_{Y,\jmath}$ lies in the two-step confidence set. ■

6 Application: Neighborhood Effects Revisited

In this section, we revisit the studies of Chetty et al. (2018) 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. (2018) construct the so-called opportunity atlas, a dataset containing correlational estimates of the effect of childhood neighborhood in adulthood. In particular, Chetty et al. (2018) 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 and on a given parental income percentile. Chetty et al. (2018) also report analogous economic mobility estimates at the commuting zone (CZ) level. In a study motivated by the findings of Chetty et al. (2018), Bergman et al. (2024) study 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. (2018) 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. (2018). 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 theoretical econometrics studying the mobility estimates of Chetty et al. (2018) 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 literature within econometrics. We seek to provide further insights on tract-level effects, building on the closely related 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. (2018) 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. (2018):

"This method is conservative because it assumes that the econometrician 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 address the former point. In particular, we seek to study tract-level effects for certain tracts of interest alone, namely high-opportunity tracts. Our methods are well-suited to problems where focusing power on units of interest is desirable. 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 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 at 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:

$$Y_j := X_j - \frac{\sum_{i \in J} X_i c_i}{\sum_{i \in J} c_i} \tag{21}$$

Y_j can be thought of as a point estimate of the gain from moving to tract j for the average housing voucher recipient.

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 $\mu_{Y,\jmath}$ for $\jmath$ 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.

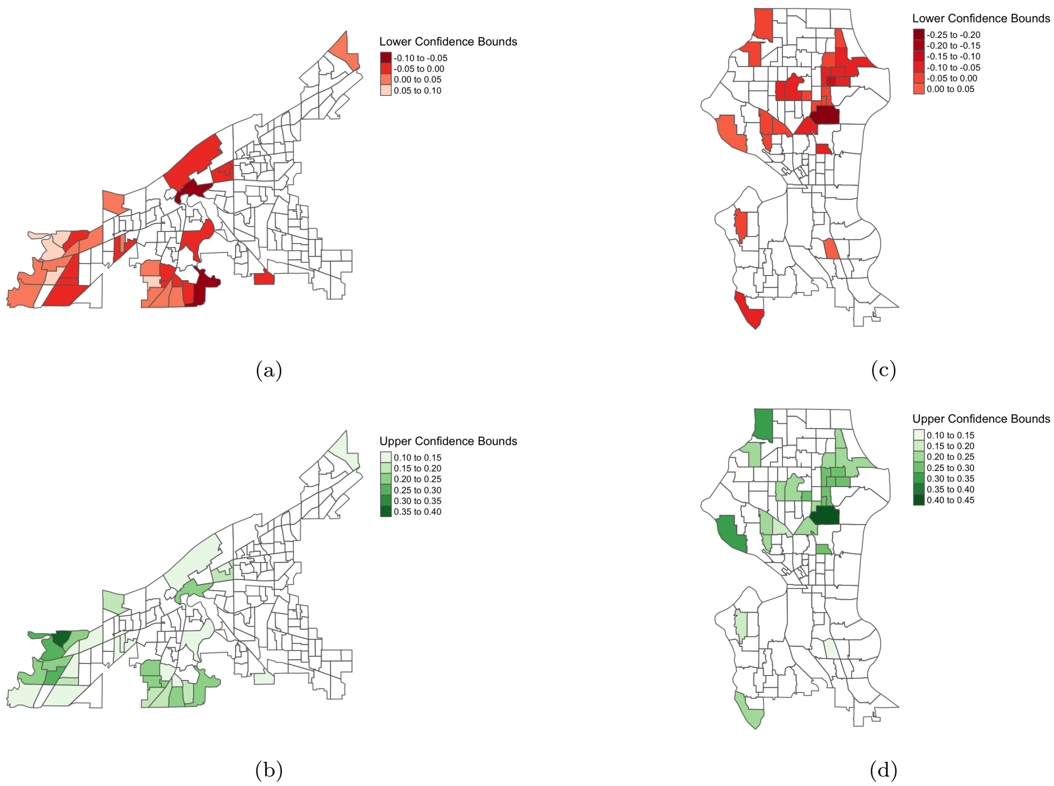


Figure 2: Lower and upper confidence bounds for economic mobility gains in selected urban tracts. Subfigures (a) and (b) display lower and upper confidence bounds on neighborhood effects in Cleveland, respectively. Subfigures (c) and (d) display lower and upper confidence bounds on neighborhood effects 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. Our two-step confidence regions are essentially equivalent to projection in urban Seattle (providing being 1.2% shorter confidence intervals on tract-level effects), while in urban Cleveland, our two-step confidence regions are a modest but non-negligible 3.5% shorter than the projection confidence regions.

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

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

Their confidence region for $\bar{\mu}_{Y,\hat{\jmath}_R}$ 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 $\mu_{Y,\jmath}$. Nevertheless, we believe that our analysis is complementary to that of Andrews et al. (2022), in that it provides insight into which of the selected tracts most credibly drive the positive aggregate effects. In contrast to Andrews et al. (2022), we find that we cannot reject the possibility of null neighborhood effects at most of the selected census tracts. Given both the results in Mogstad et al. (2023), and the simple fact that the variance of a sample mean is less than the variance of the constituent random variable forming the mean, these results are not surprising. ■

Our results suggest that, while one can draw informative conclusions on the aggregate effects of the CMTO program on economic mobility, these results are less clear when disaggregating. Moreover, while it is certainly plausible and in fact likely that treated households in the CMTO program move to a high-opportunity tract uniformly at random or approximately so, our tract-level confidence regions allow for inference robust to violations of such movement patterns.[5] Our findings suggest that, in a worst-case, we cannot reject the possibility of the CMTO program having zero or even negative effects on treated families' economic mobility outcomes.[6]

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).

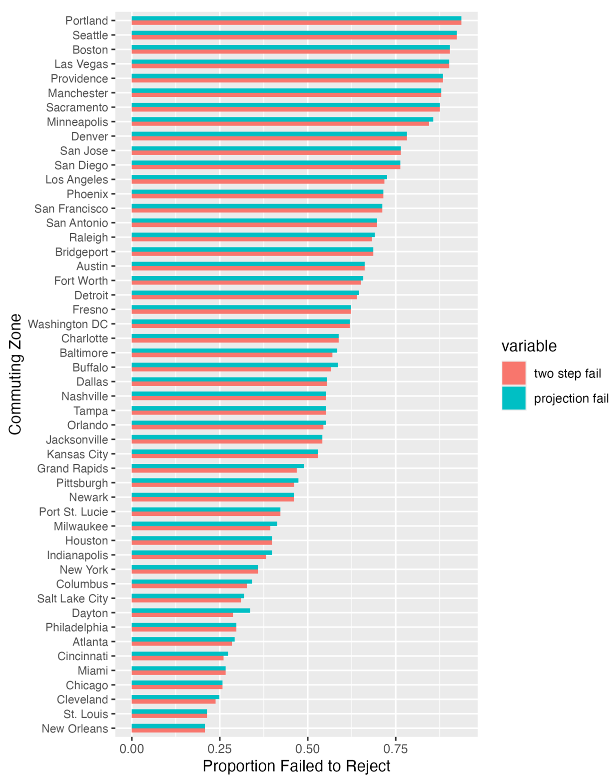


Figure 3: Proportion of selected tracts for which we fail to reject the possibility of a null effect on economic mobility relative to the average housing voucher recipient, by commuting zone. We provide results for all top-50 CZs by population, using both the two-step and projection confidence sets.

[5]Of course, in such an analysis, we assume that conditional on moving to a given high-opportunity tract, a treated household selects an individual house or apartment within that tract uniformly at random. After all, one can imagine further disaggregation of the neighborhood mobility estimates of Chetty et al. (2018). While our analysis is not robust to such concerns, we address the first-order problem of non-uniform movement patterns between tracts.

[6]Our failure to reject could be attributed to the conservativeness of our coverage criterion. As Chetty et al. (2018), such familywise error rate (FWER) controlling procedures tend to be quite conservative. We may, for example, opt for control of the false coverage rate as in Benjamini and Yekutieli (2005). Developing selective inference methods that satisfy a false coverage rate condition, rather than a simultaneous coverage condition, is an interesting area for future research.

In some commuting zones, our two-step confidence regions manage to reject the possibility of null effects for a larger proportion of selected top-third tracts than projection. Moreover, in many commuting zones, we are able to reject the possibility of null effects in over half of the selected tracts. These results suggest that our approach to inference is reasonably powerful, and can provide informative inference relative to projection.

In our final exercise on tract-level effects, we study pairwise comparison 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 5.4, this problem is closely related to the problem of inference for ranks. However, our analysis is notably different from that of Mogstad et al. (2023), since we restrict attention to pairwise comparisons of the tracts of interest, rather than to all pairwise comparisons. Consequently, in the analyses that follow, we can conduct inference on the effects of moving from low to high opportunity tracts, without enforcing coverage on the effects of moving between tracts that are not of interest.

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	Top-k	Bottom-k	% Fail to Reject Two-Step	Lowest LCB Two-Step	Highest LCB TS Two-Step	% Shorter than Projection
Seattle	26	26	87%	[-0.33, 0.59]	[0.077, 0.50]	0%
	26	1	100%	[-0.19, 0.60]	[-0.055, 0.66]	0.3%
	1	1	100%	[-0.055, 0.66]	[-0.055, 0.66]	0.3%
Cleveland	34	34	31%	[-0.076, 0.59]	[0.11, 0.49]	0%
	34	1	9%	[-0.056, 0.38]	[0.11, 0.49]	0%
	1	1	0%	[0.12, 0.48]	[0.12, 0.48]	3.4%

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 difference in the true mobilities of both CZs. Moreover, in this application our two-step confidence regions are modestly shorter than projection. However, out two-step confidence regions allow us to make substantially more rejections. In particular, projection fails to reject the no-difference null between 1% and 26% more often than two-step.

Table 4: Pairwise Mobility Gains for Selected Commuting Zones

Top	Bottom	% Fail to Reject Two-Step	Lowest LCB Two-Step	Highest LCB TS Two-Step	% Shorter than Projection	% Fewer Fail to Reject	# Additional Rejections
50%	50%	18.5%	[-0.28, 0.35]	[0.33, 0.44]	0.7%	1.0%	267
33%	33%	8.2%	[-0.25, 0.50]	[0.33, 0.44]	0.9%	1.8%	92
20%	20%	4.7%	[-0.19, 0.52]	[0.34, 0.44]	4.4%	12.2%	140
10%	10%	3.5%	[-0.14, 0.55]	[0.34, 0.44]	7.9%	35.6%	103
33%	50%	12.0%	[0.28, 0.35]	[0.33, 0.44]	0.8%	1.3%	149
20%	50%	8.3%	[-0.22, 0.38]	[0.33, 0.44]	2.1 %	4.2%	198
10%	50%	7.1%	[-0.18, 0.41]	[0.34, 0.44]	5.0%	7.7%	159
33%	67%	19.9%	[-0.28, 0.35]	[0.33, 0.44]	0.7%	1.0%	247
20%	80%	19.1%	[-0.25, 0.29]	[0.33, 0.44]	1.9%	2.9%	500
10%	90%	17.6%	[-0.22, 0.31]	[0.34, 0.44]	4.7%	6.9%	633

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."

Our findings suggest a more complicated story. We are able to distinguish high and low-opportunity commuting zones with high frequency. The methods of Mogstad et al. (2023) depend on confidence regions for all pairwise comparisons. On the other hand, our methods focus only on pairwise comparisons of high and low-medium mobility commuting zones. As our comparison with the analogous projection approaches demonstrates, this difference translates to real gains in power. Moreover, Mogstad et al. (2023) use pairwise comparisons to construct confidence regions for ranks, likely losing power in the process. Finally, we do not construct confidence regions for ranks based on the confidence regions for selected, pairwise differences summarized above, providing a further explanation for the differences in our conclusions.

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. We observe improvements in absolute coverage probability of between three and 91%. We also observe improvements in mean confidence set volume of up to 34%.

We consider simulations where $J = \{1, \ldots, p\}$, 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 E. 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 E. 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 $1,\ 10,\ 100,\ 1000,$ and 10000. 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.

Design	Confidence Set	Sample Size				
		1	10	100	1000	10000
A	Projection	0.978	0.989	0.991	0.991	0.992
	Two-Step	0.965	0.972	0.966	0.955	0.955
	Zoom	0.979	0.989	0.977	0.951	0.950
	Locally Simultaneous	0.978	0.980	0.971	0.955	0.955
B	Projection	0.984	0.989	0.993	0.996	0.995
	Two-Step	0.968	0.965	0.969	0.970	0.960
	Zoom	0.985	0.989	0.985	0.985	0.971
	Locally Simultaneous	0.980	0.977	0.979	0.976	0.962
C	Projection	0.976	0.977	0.976	0.976	0.976
	Two-Step	0.962	0.962	0.962	0.962	0.962
	Zoom	0.977	0.977	0.977	0.977	0.977
	Locally Simultaneous	0.979	0.979	0.979	0.979	0.979
D	Projection	0.975	0.988	0.991	0.991	0.991
	Two-Step	0.964	0.976	0.970	0.970	0.969
	Zoom	0.976	0.988	0.979	0.976	0.976
	Locally Simultaneous	0.977	0.983	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, the two-step approach outperforms the approach of Zrnic and Fithian (2024a), which is based on the zoom test. However, the approach of Zrnic and Fithian (2024a) more closely matches the desired 95% coverage level in the case of the clear winner than our two step approach. 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 E, the two-step approach to inference reduces overcoverage error by up to 96% relative to projection inference.

Design	Confidence Set	Sample Size				
		1	10	100	1000	10000
A	Two-Step	0.937	0.889	0.835	0.780	0.780
	Zoom	1.003	0.966	0.872	0.786	0.763
	Locally Simultaneous	1.013	0.950	0.879	0.781	0.780
B	Two-Step	0.922	0.863	0.824	0.784	0.730
	Zoom	1.003	0.962	0.900	0.842	0.758
	Locally Simultaneous	1.000	0.926	0.888	0.827	0.731
C	Two-Step	0.937	0.937	0.937	0.937	0.937
	Zoom	1.003	1.003	1.003	1.003	1.003
	Locally Simultaneous	1.014	1.014	1.014	1.014	1.014
D	Two-Step	0.938	0.928	0.846	0.846	0.846
	Zoom	1.003	0.995	0.914	0.873	0.873
	Locally Simultaneous	1.014	1.003	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 two-step 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 is based on certain heuristics on the role of β in two step inference. Choosing $\beta = 0$ renders the two-step approach equivalent to the projection approach. The parameter β allows us to model the errors of the maximum, rather than all errors simultaneously. A choice of β close to zero limits our ability to restrict inference in the manner described above, but also requires a very minimal Bonferroni correction. A choice of β close to α provides a tighter restriction on the indices of interest, at the expense of a larger Bonferroni correction. Our suggested choice of β is the same as in Andrews et al. (2023) and Romano et al. (2014). ■

8 Conclusion

In this paper, we consider a generalization of the inference on winners setting described in Andrews et al. (2023), which we dub the inference on multiple winners setting. We review the existing approaches to inference on winners, and demonstrate that they are either not appropriate in the inference on multiple winners setting, or overly conservative. We therefore propose a novel two-step approach to inference on multiple winners based on the two-step approach to testing moment inequalities from Romano et al. (2014). We demonstrate both theoretically and via simulation that this two-step approach asymptotically outperforms existing approaches for a broad range of data generating processes where the winner is asymptotically clear.

We consider several empirical settings where the inference on multiple winners setting may be relevant.

We revisit the JOBSTART demonstration of Cave et al. (1993) and Miller et al. (2005). We suggest that in this setting, retraining sites are selected based on a statistical significance cutoff rule, meaning that multiple sites are selected with frequency ranging from 17% to 52%. Applying our two-step approach to inference in this setting provides confidence sets that are 8% shorter than projection approaches. We also revisit the Creating Moves to Opportunity program from Bergman et al. (2024), finding that we cannot reject the possibility of null neighborhood effects in many of the program's selected census tracts, regardless of the urban area studied. In this setting, we find that the two-step confidence set may provide tighter intervals on neighborhood effects, relative to the projection confidence set. Finally, we demonstrate via simulation that two-step approaches can outperform projection approaches by as much as 92% in overcoverage error and 34% in confidence set volume. Thanks to these substantial improvements on key confidence set properties, derived both analytically and via simulation, we recommend using the two-step approach in settings where selection occurs on multiple dimensions, and a hybridized two-step approach where selection is one-dimensional, as in Andrews et al. (2023).

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., 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.

Angelucci, M., Karlan, D. and Zinman, J. (2015). Microcredit impacts: Evidence from a randomized microcredit program placement experiment by compartamos banco. *American Economic Journal: Applied Economics*, **7** 151–82. URL https://www.aeaweb.org/articles?id=10.1257/app.20130537.

Attanasio, O., Augsburg, B., De Haas, R., Fitzsimons, E. and Harmgart, H. (2015). The impacts of microfinance: Evidence from joint-liability lending in mongolia. *American Economic Journal: Applied Economics*, **7** 90–122. URL https://www.aeaweb.org/articles?id=10.1257/app.20130489.

Augsburg, B., De Haas, R., Harmgart, H. and Meghir, C. (2015). The impacts of microcredit: Evidence from bosnia and herzegovina. *American Economic Journal: Applied Economics*, **7** 183–203. URL https://www.aeaweb.org/articles?id=10.1257/app.20130272.

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

Banerjee, A., Duflo, E., Glennerster, R. and Kinnan, C. (2015a). The miracle of microfinance? evidence from a randomized evaluation. *American Economic Journal: Applied Economics*, **7** 22–53. URL https://www.aeaweb.org/articles?id=10.1257/app.20130533.

Banerjee, A., Karlan, D. and Zinman, J. (2015b). Six randomized evaluations of microcredit: Introduction and further steps. *American Economic Journal: Applied Economics*, **7** 1–21. URL https://www.aeaweb.org/articles?id=10.1257/app.20140287.

Benjamini, Y. and Yekutieli, D. (2005). False discovery rate–adjusted multiple confidence intervals for selected parameters. *Journal of the American Statistical Association*, **100** 71–81. https://doi.org/10.1198/016214504000001907, URL https://doi.org/10.1198/016214504000001907.

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.

Canay, I. A. and Shaikh, A. M. (2017). *Practical and Theoretical Advances in Inference for Partially Identified Models*. Econometric Society Monographs, Cambridge University Press, 271–306.

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. (2018). The opportunity atlas: Mapping the childhood roots of social mobility. Working Paper 25147, National Bureau of Economic Research. URL http://www.nber.org/papers/w25147.

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.

Crépon, B., Devoto, F., Duflo, E. and Parienté, W. (2015). Estimating the impact of microcredit on those who take it up: Evidence from a randomized experiment in morocco. *American Economic Journal: Applied Economics*, **7** 123–50. URL https://www.aeaweb.org/articles?id=10.1257/app.20130535.

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.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, **6** 65–70. URL http://www.jstor.org/stable/4615733.

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.

Markovic, J., Xia, L. and Taylor, J. (2018). Unifying approach to selective inference with applications to cross-validation. 1703.06559.

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.

Pritchett, L. and Sandefur, J. (2015). Learning from experiments when context matters. *American Economic Review*, **105** 471–75. URL https://www.aeaweb.org/articles?id=10.1257/aer.p20151016.

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., 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.

Tarozzi, A., Desai, J. and Johnson, K. (2015). The impacts of microcredit: Evidence from ethiopia. *American Economic Journal: Applied Economics*, **7** 54–89. URL https://www.aeaweb.org/articles?id=10.1257/app.20130475.

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, I discuss three alternative approaches to the inference on multiple winners problem. Our first approach is novel, but closely related to the approach of Zrnic and Fithian (2024b). Our second approach is directly based on Zrnic and Fithian (2024b), while our final approach is based on Zrnic and Fithian (2024a). Our first approach uses a similar Bonferroni-type correction as the approach in section 5. However, this approach differs from our approach in section 5. 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 is thus very similar to that of Zrnic and Fithian (2024b), though it differs in a few subtle ways. 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). Their approach is robust to arbitrary dependence, in the spirit of the Bonferroni and Holm (1979) procedures, and is thus conservative. Generally, all three of these approaches are underpowered compared to our two-step approach to inference.

A.1 A First 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 $J_c \subseteq J$ be the smallest set such that:[7]

$$P_{\mu,\Sigma}\left(\hat{J}_R \subseteq J_c\right) \geq 1 - \beta_1 \tag{22}$$

Similarly, we seek to construct a prediction 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 7.

It remains to show how to construct $\hat{J}_c$. By 8, it follows that:

$$p_j := P_{\mu,\Sigma}\left(j \in \hat{J}_R\right) = P_{\mu,\Sigma}\left(\sum_{j' \in J} \mathbb{1}\left(\xi_{X,j} \geq \xi_{X,j'} + \Delta_{j',j}\right) \in R\right) \tag{23}$$

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 10.

[7]If multiple equal cardinality J_c satisfy 22, we choose the J_c minimizing the probability on the left hand side of 22.

It follows that $P_{\mu,\Sigma}(L \leq \Delta \leq U) \geq 1-\beta_2$. We can now observe that this probability no longer depends on μ. Let $p_j(L,U)$ be the right hand side of 23 with L and U in place of Δ, which we take as follows:

$$p_j(L,U) = P_{\mu,\Sigma}\left(\left[\sum_{j'\in J} \mathbb{1}\left(\xi_{X,j} \geq \xi_{X,j'} + U_{j',j}\right), \sum_{j'\in J} \mathbb{1}\left(\xi_{X,j} \geq \xi_{X,j'} + L_{j',j}\right)\right] \cap R \neq \emptyset\right) \tag{24}$$

We have that $P_{\mu,\Sigma}\left(p(L,U) \leq p\right) \geq 1-\beta_2$, where 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:

$$CS^{LS1}_{1-\alpha;\beta} := \times_{j\in\hat{J}_R}\left[Y_j - c_{1-(\alpha-\beta)}\left(\hat{J}_c\right)\sqrt{\Sigma_{Y,jj}},\ Y_j + c_{1-(\alpha-\beta)}\left(\hat{J}\right)\sqrt{\Sigma_{Y,jj}}\right]$$

The following proposition follows:

Proposition A.1. *$CS^{LS1}_{1-\alpha;\beta}$ is a valid confidence set at the $1-\alpha$-level, such that:*

$$P_{\mu,\Sigma}\left(\left(\mu_{Y,j}\right)_{j\in\hat{J}_R} \in CS^{LS1}_{1-\alpha;\beta}\right) \geq 1-\alpha \tag{25}$$

for all μ and Σ. Marginal validity clearly follows.

A proof of this proposition is provided in appendix C.

A.2 A Second Locally Simultaneous Approach to 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) construct some set of likely selections $\hat{J}_c$ and conduct simultaneous inference restricted to this $\hat{J}_c$. In general, Zrnic and Fithian (2024b) construct this $\hat{J}_c$ 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 $\hat{J}_c$ described above.

For the inference on multiple winners problem, we take:

$$\hat{J}_c := \left\{j; X_j \notin \bigcup_{j\in\hat{J}_R}\left[X_j - 2d_{1-\beta/2}(\Sigma),\ X_j + 2d_{1-\beta/2}(\Sigma)\right]\right\} \tag{26}$$

Finally, we may define the following confidence set:

$$CS^{LS2}_{1-\alpha;\beta} := \times_{j\in\hat{J}_R}\left[Y_j - c_{1-(\alpha-\beta)}\left(\hat{J}_c\right)\sqrt{\Sigma_{Y,jj}},\ Y_j + c_{1-(\alpha-\beta)}\left(\hat{J}\right)\sqrt{\Sigma_{Y,jj}}\right]$$

As before, we have the following proposition:

Proposition A.2. *$CS^{LS1}_{1-\alpha}$ is a valid confidence set at the $1-\alpha$-level, such that:*

$$P_{\mu,\Sigma}\left((\mu_{Y,\jmath})_{\jmath\in\hat{J}_R} \in CS^{LS2}_{1-\alpha}\right) \geq 1-\alpha \tag{27}$$

for all μ and Σ. Again, 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 a test which Zrnic and Fithian (2024b) call the zoom 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 generalize their approach to the exact inference on multiple winners problem we discuss in this paper, they provide guidance

Let J_R be the set of indices j such that μ_X is ranked in R at j. Let the suboptimality $D_j := \min_{j'\neq J_R} |\mu_{X,j} - \mu_{X,j'}|$ 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:

$$A_\alpha := \left[\mu_{Y,j} \pm \left(r_\alpha \vee \frac{D_j}{2}\right)\right]_{j\in J} \tag{28}$$

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 $(\mu_{Y,\jmath})_{\jmath\in\hat{J}_R}$:

$$CS_{1-\alpha} = \{\mu_Y,\ \mu_X; (Y_j)_{j\in J} \in A_\alpha\} \tag{29}$$

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

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

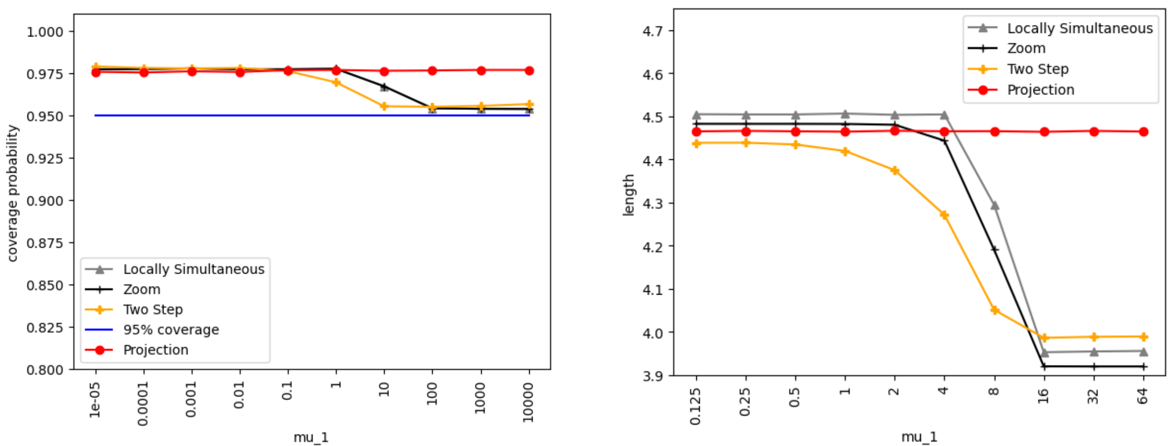
Consequently, we construct the zoom confidence set for the multiple winners:

$$CS^{zoom}_{1-\alpha} := \times_{\jmath\in\hat{J}_R} CS^{zoom}_{1-\alpha,\jmath} \tag{30}$$

Of course, it is not clear how to invert the test based on the acceptance region A_α. Zrnic and Fithian (2024a) provide computationally efficient procedures to invert the zoom test under certain special cases, namely when R is of the form $\{1,\ldots,\tau\}$ for some $\tau \leq p$. 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 E. 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 there 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 E.

B Uniform Asymptotic Validity and Proofs

In this section, we formalize the assumptions stated in section 5.2 and introduce several lemmas supporting the proof of proposition B.1. In particular, the lemmas in this section closely resemble proposition 11 from Andrews et al. (2023), in that we derive uniform asymptotic validity results on conservative, rectangular confidence intervals.

B.1 Assumptions

In this section, we provide a list of formally-stated assumptions necessary to prove propositions B.1 and 5.3. First, we provide the following uniform integrability assumption, which is equivalent to uniform convergence in distribution, per lemma B.1.

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

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

Moreover, we impose the following uniform consistency condition on our estimator $\hat{\Sigma}^n$ for $\Sigma(P)$.

Assumption B.2. We assume that $\hat{\Sigma}^n$ is a uniformly, asymptotic consistent estimator for $\Sigma(P)$, such that, for any $\varepsilon > 0$:

$$\limsup_{n \to \infty} \sup_{P \in \mathcal{P}} Pr_P \left(\left\| \hat{\Sigma}^n - \Sigma(P) \right\| > \varepsilon \right) = 0$$

where our choice of matrix norm is the max norm.

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

Lemma B.1. *Under assumptions B.1-B.2, we have uniform convergence in distribution such that there exist functions $\mu_{W,n}(P) := \mu_W(P)$ satisfying:*

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

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 $j = 1, \ldots, 2p$, 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). ■

Our second 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:

$$L^n_{j',j} = \tilde{S}^n_{X,j'} - \tilde{S}^n_{X,j} - d_{1-\beta/2}(\hat{\Sigma}^n) \sqrt{\frac{\hat{\text{var}}^n_{j'j}}{n}}$$
$$U^n_{j',j} = \tilde{S}^n_{X,j'} - \tilde{S}^n_{X,j} + d_{1-\beta/2}(\hat{\Sigma}^n) \sqrt{\frac{\hat{\text{var}}^n_{j'j}}{n}}$$

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 defintion for $\text{var}_{j',j}(P)$.

Lemma B.2. *Given assumption B.1, it follows that L^n is a uniformly, asymptotically valid lower bound for $\Delta^n(P)$. In particular:*

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

Proof. By way of contradiction, let us suppose that there exists a sequence P_{n_s} such that:

$$Pr_{P_{n_s}} \left(L^{n_s} \leq \Delta^{n_s}(P_{n_s}) \right) \to \eta < 1 - \beta/2 \tag{31}$$

Let us notice that $L^{n_s} \leq \Delta^{n_s}(P_{n_s})$ if and only if, for any j and j' in $\{1, \ldots, p\}$, we have:

$$\sqrt{\frac{n_s}{\hat{\text{var}}^{n_s}_{j'j}}} \left(\tilde{S}^{n_s}_{X,j'} - \mu_{X,j'}(P_{n_s}) - \left(\tilde{S}^{n_s}_{X,j} - \mu_{X,j}(P_{n_s}) \right) \right) \leq d_{1-\beta/2}(\hat{\Sigma}^{n_s})$$

We have that:

$$\begin{aligned}
&\Bigg| Pr_{P_{n_s}} \left(\bigcap_{1 \leq j,j' \leq p} \sqrt{\frac{n_s}{\hat{\text{var}}_{j',j}}} \left(\tilde{S}^{n_s}_{X,j'} - \mu_{X,j'}(P_{n_s}) - \left(\tilde{S}^{n_s}_{X,j} - \mu_{X,j}(P_{n_s}) \right) \right) \leq d_{1-\beta/2}(\hat{\Sigma}^{n_s}) \right) \\
&\quad - Pr_{\xi_X \sim \Phi_{\Sigma(P_{n_s})}} \left(\bigcap_{1 \leq j,j' \leq p} \frac{(\xi_{X,j'} - \xi_{X,j})}{\sqrt{\text{var}_{j',j}(P_{n_s})}} \leq d_{1-\beta/2}(\Sigma(P_{n_s})) \right) \Bigg| \\
&\leq \Bigg| Pr_{P_{n_s}} \left(\bigcap_{1 \leq j,j' \leq p} \sqrt{\frac{n_s}{\hat{\text{var}}_{j',j}}} \left(\tilde{S}^{n_s}_{X,j'} - \mu_{X,j'}(P_{n_s}) - \left(\tilde{S}^{n_s}_{X,j} - \mu_{X,j}(P_{n_s}) \right) \right) \leq d_{1-\beta/2}(\hat{\Sigma}^{n_s}) \right) \\
&\quad - Pr_{\xi_X \sim \Phi_{\Sigma(P_{n_s})}} \left(\bigcap_{1 \leq j,j' \leq p} \frac{(\xi_{X,j'} - \xi_{X,j})}{\sqrt{\hat{\text{var}}_{j',j}}} \leq d_{1-\beta/2}(\hat{\Sigma}^{n_s}) \right) \Bigg| \\
&\quad + \Bigg| Pr_{\xi_X \sim \Phi_{\Sigma(P_{n_s})}} \left(\bigcap_{1 \leq j,j' \leq p} \frac{(\xi_{X,j'} - \xi_{X,j})}{\sqrt{\hat{\text{var}}_{j',j}}} \leq d_{1-\beta}(\hat{\Sigma}^{n_s}) \right) \\
&\quad - Pr_{\xi_X \sim \Phi_{\Sigma(P_{n_s})}} \left(\bigcap_{1 \leq j,j' \leq p} \frac{(\xi_{X,j'} - \xi_{X,j})}{\sqrt{\text{var}_{j',j}(P_{n_s})}} \leq d_{1-\beta/2}(\Sigma(P_{n_s})) \right) \Bigg| \\
&= o(1) + \Bigg| Pr_{\xi_X \sim \Phi_{\Sigma(P_{n_s})}} \left(\bigcap_{1 \leq j,j' \leq p} \frac{(\xi_{X,j'} - \xi_{X,j})}{\sqrt{\hat{\text{var}}_{j',j}}} \leq d_{1-\beta/2}(\hat{\Sigma}^{n_s}) \right) \\
&\quad - Pr_{\xi_X \sim \Phi_{\Sigma(P_{n_s})}} \left(\bigcap_{1 \leq j,j' \leq p} \frac{(\xi_{X,j'} - \xi_{X,j})}{\sqrt{\text{var}_{j',j}(P_{n_s})}} \leq d_{1-\beta/2}(\Sigma(P_{n_s})) \right) \Bigg|
\end{aligned}$$

where the first inequality follows by the triangle and the second equality holds by lemma B.1. We want to show that:

$$\begin{aligned}
&\Bigg| Pr_{\xi_X \sim \Phi_{\Sigma(P_{n_s})}} \left(\bigcap_{1 \leq j,j' \leq p} \frac{(\xi_{X,j'} - \xi_{X,j})}{\sqrt{\hat{\text{var}}_{j',j}}} \leq d_{1-\beta/2}(\hat{\Sigma}^{n_s}) \right) \\
&\quad - Pr_{\xi_X \sim \Phi_{\Sigma(P_{n_s})}} \left(\bigcap_{1 \leq j,j' \leq p} \frac{(\xi_{X,j'} - \xi_{X,j})}{\sqrt{\text{var}_{j',j}(P_{n_s})}} \leq d_{1-\beta/2}(\Sigma(P_{n_s})) \right) \Bigg| = o(1)
\end{aligned} \tag{32}$$

Let $a^{n_s}_{j',j} := \sqrt{\frac{\text{var}_{j',j}(P_{n_s})}{\hat{\text{var}}_{j',j}}} \frac{d_{1-\beta/2}(\Sigma(P_{n_s}))}{d_{1-\beta/2}(\hat{\Sigma}^{n_s})}$. By assumption B.2 and the continuity of d, we have that, for any $\delta > 0$, that:

$$Pr_{P_{n_s}} \left(\left| a^{n_s}_{j',j} - 1 \right| > \delta \quad \text{for any} \quad j, j' \right) = o(1)$$

Similarly, we have that 32 can be rewritten as:

$$\left| Pr_{\xi_X \sim \Phi_{\Sigma(P_{n_s})}} \left(\bigcap_{1 \le j,j' \le p} a^{n_s}_{j',j} \frac{(\xi_{X,j'} - \xi_{X,j})}{\sqrt{\text{var}_{j',j}(P_{n_s})}} \le d_{1-\beta/2}(\Sigma(P_{n_s})) \right) \right.$$
$$\left. - Pr_{\xi_X \sim \Phi_{\Sigma(P_{n_s})}} \left(\bigcap_{1 \le j,j' \le p} \frac{(\xi_{X,j'} - \xi_{X,j})}{\sqrt{\text{var}_{j',j}(P_{n_s})}} \le d_{1-\beta/2}(\Sigma(P_{n_s})) \right) \right| = o(1)$$

To simplify notation, we will take $s := n_s$, and take a further subsequence s_k. We construct this further subsequence as follows. For each s, we choose any sequence of $\delta_s > 1$ satisfying:

$$\left| Pr_{\xi_X \sim \Phi_{\Sigma(P_s)}} \left(\bigcap_{1 \le j,j' \le p} (1 \pm \delta_s) \frac{(\xi_{X,j'} - \xi_{X,j})}{\sqrt{\text{var}_{j',j}(P_s)}} \le d_{1-\beta/2}(\Sigma(P_s)) \right) \right.$$
$$\left. - Pr_{\xi_X \sim \Phi_{\Sigma(P_s)}} \left(\bigcap_{1 \le j,j' \le p} \frac{(\xi_{X,j'} - \xi_{X,j})}{\sqrt{\text{var}_{j',j}(P_s)}} \le d_{1-\beta/2}(\Sigma(P_s)) \right) \right| = o(1)$$

We choose a subsequence s_k such that:

$$Pr_{P_{s_k}} \left(\left| a^{s_k}_{j',j} - 1 \right| > \delta_{s_k} \quad \text{for any} \quad j, j' \right) = o(1)$$

It follows that on this further subsequence, we have:

$$\left| Pr_{P_{s_k}} \left(\bigcap_{1 \le j,j' \le p} \sqrt{\frac{s_k}{\hat{\text{var}}_{j',j}}} \left(\tilde{S}^{s_k}_{X,j'} - \mu_{X,j'}(P_{s_k}) - \left(\tilde{S}^{s_k}_{X,j} - \mu_{X,j}(P_{s_k}) \right) \right) \le d_{1-\beta/2}(\hat{\Sigma}^{s_k}) \right) \right.$$
$$\left. - Pr_{\xi_X \sim \Phi_{\Sigma(P_{s_k})}} \left(\bigcap_{1 \le j,j' \le p} \frac{(\xi_{X,j'} - \xi_{X,j})}{\sqrt{\text{var}_{j',j}(P_{s_k})}} \le d_{1-\beta/2}(\Sigma(P_{s_k})) \right) \right| = o(1)$$

Of course, we have that:

$$\limsup_{k \to \infty} Pr_{\xi_X \sim \Phi_{\Sigma(P_{s_k})}} \left(\bigcap_{1 \le j,j' \le p} \frac{(\xi_{X,j'} - \xi_{X,j})}{\sqrt{\text{var}_{j',j}(P_{s_k})}} \le d_{1-\beta/2}(\Sigma(P_{s_k})) \right) \ge 1 - \beta/2$$

Implying that on the further subsequence:

$$\limsup_{k \to \infty} Pr_{P_{s_k}} \left(\bigcap_{1 \le j,j' \le p} \sqrt{\frac{s_k}{\hat{\text{var}}_{j',j}}} \left(\tilde{S}^{s_k}_{X,j'} - \mu_{X,j'}(P_{s_k}) - \left(\tilde{S}^{s_k}_{X,j} - \mu_{X,j}(P_{s_k}) \right) \right) \ge d_{1-\beta/2}(\hat{\Sigma}^{s_k}) \right) \ge 1 - \beta/2 \tag{33}$$

yielding a contradiction of 31. Thus, we have:

$$\liminf_{n \to \infty} \inf_{P \in \mathcal{P}} Pr_P \left(L^n \le \Delta^n(P) \right) \ge 1 - \beta/2$$

with an analogous result holding for U^n. Applying a Bonferroni correction proves 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 $\jmath$ in $\hat{J}_R$. Formally, we have the following lemma.

Lemma B.3. *Under assumptions B.1-B.2, the following uniform asymptotic validity condition applies to $\rho_{1-(\alpha-\beta)}$:*

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} Pr_P \left(\max_{j\in J} \left\{ \frac{\sqrt{n}|\xi_{\tilde{S}_Y^n,j}|}{\sqrt{\hat{\Sigma}_{Y,jj}^n}} \mathbb{1}\left(\sum_{j'\in J} \mathbb{1}\left(\xi_{\tilde{S}_X^n,j} \geq \xi_{\tilde{S}_X^n,j'} + \Delta_{j',j} \right) \in R \right) \right\} > \rho_{1-\alpha+\beta}(\Delta^n, \Delta^n; \hat{\Sigma}^n) \right) \leq \alpha-\beta$$

Proof. By way of contradiction, let us assume that there exists a sequence of measures P_{n_s} such that:

$$\lim_{s\to\infty} Pr_{P_{n_s}} \left(\max_{j\in J} \left\{ \frac{\sqrt{n_s}|\xi_{\tilde{S}_Y^{n_s},j}|}{\sqrt{\hat{\Sigma}_{Y,jj}^{n_s}}} \mathbb{1}\left(\sum_{j'\in J} \mathbb{1}\left(\xi_{\tilde{S}_X^{n_s},j} \geq \xi_{\tilde{S}_X^{n_s},j'} + \Delta_{j',j} \right) \in R \right) \right\} > \rho_{1-\alpha+\beta}(\Delta^{n_s}, \Delta^{n_s}; \hat{\Sigma}^{n_s}) \right) = \eta > \alpha-\beta$$

Equivalently:

$$\lim_{s\to\infty} Pr_{P_{n_s}} \left(\max_{\jmath\in\hat{J}_R} \left\{ \frac{\sqrt{n_s}\left|\xi_{\tilde{S}_Y^{n_s},\jmath}\right|}{\sqrt{\hat{\Sigma}_{Y,\jmath\jmath}^{n_s}}} \right\} > \rho_{1-\alpha+\beta}(\Delta^{n_s}, \Delta^{n_s}; \hat{\Sigma}^{n_s}) \right) = \eta$$

The same reasoning as in the proof of lemma B.2 provides a further subsequence of n_s such that the above limit is less than or equal to $\alpha-\beta$, yielding a contradiction, and proving the lemma. ■

Finally, we can prove the following proposition:

Proposition B.1. *Given assumptions B.1-B.2, the two step confidence set is uniformly, asymptotically valid, such that:*

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} Pr_P \left(\left\{ \mu_{Y,\jmath}^n(P) \right\}_{\jmath\in\hat{J}_R} \in CS_{1-\alpha;n}^{TS} \right) \geq 1-\alpha \tag{34}$$

Proof. **Proof of Proposition B.1.** First, we note that, by lemma B.2:

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} Pr_P \left(L^n \leq \Delta^n(P) \leq U^n \right) \geq 1-\beta \tag{35}$$

We can define the event $B^n := \{L^n \leq \Delta^n \leq U^n\}$. In addition, we note that, for some 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 (35):

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

and we seek to show that:

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} Pr_P \left(\left(\mu_{Y,n}\left(\hat{\theta}_j^n; P\right) \right)_{j=1}^k \notin CS_{1-\alpha;n}^{TS} \right) \leq \alpha$$

Indeed:

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

where the final inequality holds by lemma B.3. ■

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,[8] 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 $\{\mu_{Y,\jmath}\}_{\jmath\in\hat{J}_R}\in CS^c_{1-\alpha}$ 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, (5) holds. ■

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

Now, we provide proofs for supplementary results.

Proof. **Proof of Proposition 5.1.** 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:

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

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

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

for all μ, implying that both (13) and (14) hold for all μ. ■

We now provide a proof of proposition 5.2, which proceeds much as the proof of proposition 5.1.

Proof. **Proof of Proposition 5.2.** 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 5.1, 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:

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

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 16 holds. ■

Finally, to prove proposition 5.3, we impose a new assumption. In particular, we impose the following pointwise rate of convergence on our variance-covariance estimator $\hat{\Sigma}^n$, which is necessary in proving almost sure convergence of variance-covariance in the proof for proposition 5.3.

Assumption C.1. We assume that, for any P in $\mathcal{P}$, there exists $\delta > 0$ such that, for any $\varepsilon > 0$:

$$Pr_P\left(\left\|\hat{\Sigma}^n(P) - \Sigma(P)\right\| > \varepsilon\right) = o(1/n^{\delta})$$

where our choice of matrix norm is again the max norm.

We now provide a proof of proposition 5.3.

Proof. **Proof of Proposition 5.3.** We begin by recalling that:

$$L^n_{j',j} = \tilde{S}^n_{X,j'} - \tilde{S}^n_{X,j} - d_{1-\beta/2}(\hat{\Sigma}^n)\sqrt{\frac{\hat{\text{var}}^n_{j'j}}{n}}$$
$$U^n_{j',j} = \tilde{S}^n_{X,j'} - \tilde{S}^n_{X,j} + d_{1-\beta/2}(\hat{\Sigma}^n)\sqrt{\frac{\hat{\text{var}}^n_{j'j}}{n}}$$

We note that, by assumption B.2, for fixed P in $\mathcal{P}$, $\left\|\hat{\Sigma}^n - \Sigma(P)\right\| = O(1)$ almost surely. Consequently, by continuity, $d_{1-\beta/2}(\hat{\Sigma}^n) = O(1)$ almost surely, while the vector $\hat{\text{var}}^n/n = o(1)$ almost surely. Thus:

$$\begin{pmatrix} L^n_{j',j} \\ U^n_{j',j} \end{pmatrix} \xrightarrow{a.s.} \begin{pmatrix} \Delta_{j',j}(P) \\ \Delta_{j',j}(P) \end{pmatrix} \tag{36}$$

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 (36) that the event $B := \{|L^n - \Delta|,\ |U^n - \Delta| \leq \delta/6\ e.v.\}$ holds almost surely. That is, $P(B) = 1$. We seek to study the convergence of $\rho_{1-\alpha+\beta}\left(L^n, U^n; \hat{\Sigma}^n/n\right)$ on B. We now seek to show that:

$$\mathbb{1}\left(\left[\sum_{j' \in J} \mathbb{1}\left(\xi^n_{X,j} \geq \xi^n_{X,j'} + U^n_{j',j}\right), \sum_{j' \in J} \mathbb{1}\left(\xi^n_{X,j} \geq \xi^n_{X,j'} + L^n_{j',j}\right)\right] \cap R \neq \emptyset\right) \to \mathbb{1}\left(j \in J_R(P)\right) \tag{37}$$

in probability, conditional on L^n and U^n. We note that ξ^n_W are drawn from a Gaussian with mean zero and variance-covariance $\hat{\Sigma}^n/n$, and are defined to be independent from the $\tilde{X}_j$ and $\tilde{Y}$. Rigorously, we want that:

$$Pr_P\left(\mathbb{1}\left(\left[\sum_{j' \in J} \mathbb{1}\left(\xi^n_{X,j} \geq \xi^n_{X,j'} + U^n_{j',j}\right), \sum_{j' \in J} \mathbb{1}\left(\xi^n_{X,j} \geq \xi^n_{X,j'} + L^n_{j',j}\right)\right] \cap R \neq \emptyset\right) \neq \mathbb{1}\left(j \in J_R(P)\right) | L^n, U^n\right) \tag{38}$$

is $o(1)$ as n approaches infinity. Consequently, we can show that,[9] on the event $\{L^n = l^n, U^n = u^n\}$,

$$Pr_P\left(\mathbb{1}\left(\left[\sum_{j'\in J}\mathbb{1}\left(\xi^n_{X,j}\geq\xi^n_{X,j'}+U^n_{j',j}\right),\sum_{j'\in J}\mathbb{1}\left(\xi^n_{X,j}\geq\xi^n_{X,j'}+L^n_{j',j}\right)\right]\cap R\neq\emptyset\right)\neq\mathbb{1}\left(j\in J_R(P)\right)|L^n,U^n\right)$$
$$=Pr_P\left(\mathbb{1}\left(\left[\sum_{j'\in J}\mathbb{1}\left(\xi^n_{X,j}\geq\xi^n_{X,j'}+u^n_{j',j}\right),\sum_{j'\in J}\mathbb{1}\left(\xi^n_{X,j}\geq\xi^n_{X,j'}+l^n_{j',j}\right)\right]\cap R\neq\emptyset\right)\neq\mathbb{1}\left(j\in J_R(P)\right)\right) \quad (39)$$

Now, we define B_n to be correspond to the event that all $\xi^n_{X,j}$ are within $\delta/6$ of zero. We note that zero is simply the probability limit of the $\xi^n_{X,j}$, implying that $P(B_n)$ approaches one as n approaches infinity. It is clear that, taking l^n and u^n arbitrarily such that $|l^n-\Delta|\leq\delta/6$ and $|u^n-\Delta|\leq\delta/6$:

$$\lim_{n\to\infty}Pr_P\left(\mathbb{1}\left(\left[\sum_{j'\in J}\mathbb{1}\left(\xi^n_{X,j}\geq\xi^n_{X,j'}+u^n_{j',j}\right),\sum_{j'\in J}\mathbb{1}\left(\xi^n_{X,j}\geq\xi^n_{X,j'}+l^n_{j',j}\right)\right]\cap R\neq\emptyset\right)\neq\mathbb{1}\left(j\in J_R(P)\right)\right)$$
$$=\lim_{n\to\infty}Pr_P\left(\left\{\mathbb{1}\left(\left[\sum_{j'\in J}\mathbb{1}\left(\xi^n_{X,j}\geq\xi^n_{X,j'}+u^n_{j',j}\right),\sum_{j'\in J}\mathbb{1}\left(\xi^n_{X,j}\geq\xi^n_{X,j'}+l^n_{j',j}\right)\right]\cap R\neq\emptyset\right)\neq\mathbb{1}\left(j\in J_R(P)\right)\right\}\cap B_n\right)$$
$$=\lim_{n\to\infty}Pr_P(\emptyset)=0$$

almost surely. Thus, (38) holds. Moreover, leveraging the above convergence in conditional probability result, we note that, by the continuous mapping theorem, the right hand side of (12) converges to:

$$\max_{j\in J_R(P)}\frac{|\xi_{Y,j}|}{\sqrt{\Sigma_{Y,jj}}}$$

in distribution, given $L^n = l^n$ and $U^n = u^n$ chosen above. We note that ξ_Y is distributed according to a multivariate Gaussian with mean zero and variance-covariance $\Sigma_Y(P)$. Thus, $\rho_{\alpha,\beta}\left(L^n, U^n; \hat{\Sigma}^n\right)$ converges to $c_{1-\alpha+\beta}\left(J_R(P)\right)$ as n grows large, on B. It naturally follows that, for sufficiently small β, $c_{1-\alpha+\beta}\left(J_R(P)\right) < c_{1-\alpha}\left(J\right)$, as $c_\gamma(J_c)$ is continuous in γ for any J_c in 2^J.

It now remains to prove that equation 20 holds. To do so, we pick some small, positive value of h such that we can construct a function $\mathbb{1}_h$ satisfying:

$$\mathbb{1}(x>y)\geq\mathbb{1}_h(x,y)\geq\mathbb{1}(x>y+h)$$

and such that $\mathbb{1}_h$ is continuous in x and y. For k in R

We also let M_n be the event where $\hat{J}_{R;n} = J_R(P)$. By previous arguments, we know that the set $\{M_n\ e.v.\}$ holds almost surely. We can now make the following calculations:

$$\lim_{n\to\infty}\mathbb{E}_P\left(\mathbb{1}_h\left(\max_{j\in\hat{J}_{R;n}}\frac{\left|\tilde{S}^n_{Y,j}-\mu_{Y,j}\left(P\right)\right|}{\sqrt{\hat{\Sigma}^n_{Y,jj}/n}},\rho_{\alpha,\beta}(L^n,U^n,\hat{\Sigma}^n/n)\right)\right)$$

[9] This result can be shown by recalling the Kolmogorov's definition of conditional probability, and subsequently applying Fubini's theorem.

$$= \mathbb{E}_P \left(\lim_{n\to\infty} \mathbb{1}_h \left(\max_{j \in \hat{J}_{R;n}} \frac{\left| \tilde{S}^n_{Y,j} - \mu_{Y,j}(P) \right|}{\sqrt{\hat{\Sigma}^n_{Y,jj}/n}}, \rho_{\alpha,\beta}(L^n, U^n, \hat{\Sigma}^n/n) \right) \right)$$

$$= \mathbb{E}_P \left(\lim_{n\to\infty} \mathbb{1}_h \left(\max_{j \in J_R(P)} \frac{\left| \tilde{S}^n_{Y,j} - \mu_{Y,j}(P) \right|}{\sqrt{\hat{\Sigma}^n_{Y,jj}/n}}, c_{1-\alpha+\beta}(J_R(P)) \right) \right)$$

$$\geq \mathbb{E}_P \left(\lim_{n\to\infty} \mathbb{1}_h \left(\max_{j \in J_R(P)} \frac{\left| \tilde{S}^n_{Y,j} - \mu_{Y,j}(P) \right|}{\sqrt{\hat{\Sigma}^n_{Y,jj}/n}}, c_{1-\alpha}(J) \right) \right)$$

$$= \lim_{n\to\infty} \mathbb{E}_P \left(\mathbb{1}_h \left(\max_{j \in \hat{J}_{R;n}} \frac{\left| \tilde{S}^n_{Y,j} - \mu_{Y,j}(P) \right|}{\sqrt{\hat{\Sigma}^n_{Y,jj}/n}}, c_{1-\alpha}(J) \right) \right)$$

That

$$\lim_{n\to\infty} \mathbb{E}_P \left(\mathbb{1}_h \left(\max_{j \in \hat{J}_{R;n}} \frac{\left| \tilde{S}^n_{Y,j} - \mu_{Y,j}(P) \right|}{\sqrt{\hat{\Sigma}^n_{Y,jj}/n}}, \rho_{\alpha,\beta}(L^n, U^n, \hat{\Sigma}^n/n) \right) \right)$$

exists holds by the central limit theorem,[10] continuous mapping theorem, and portmanteau lemma. The same reasoning applies to show the existence of:

$$\lim_{n\to\infty} \mathbb{E}_P \left(\mathbb{1}_h \left(\max_{j \in \hat{J}_{R;n}} \frac{\left| \tilde{S}^n_{Y,j} - \mu_{Y,j}(P) \right|}{\sqrt{\hat{\Sigma}^n_{Y,jj}/n}}, c_{1-\alpha}(J) \right) \right)$$

Thus, with these existence checks in mind, we have that, for arbitrary h, it follows that:

$$\lim_{n\to\infty} Pr_P \left(\max_{j \in \hat{J}_{R;n}} \frac{\left| \tilde{S}^n_{Y,j} - \mu_{Y,j}(P) \right|}{\sqrt{\hat{\Sigma}^n_{Y,jj}/n}} > \rho_{\alpha,\beta}(L^n, U^n, \hat{\Sigma}^n/n) \right)$$
$$\geq \lim_{n\to\infty} Pr_P \left(\max_{j \in \hat{J}_{R;n}} \frac{\left| \tilde{S}^n_{Y,j} - \mu_{Y,j}(P) \right|}{\sqrt{\hat{\Sigma}^n_{Y,jj}/n}} > c_{1-\alpha}(J) + h \right) \tag{40}$$

We know that, by the continuous mapping theorem, the term:

$$\max_{j \in \hat{J}_{R;n}} \frac{\left| \tilde{S}^n_{Y,j} - \mu_{Y,jj}(P) \right|}{\sqrt{\hat{\Sigma}^n_{Y,jj}/n}}$$

converges in distribution to:

$$\max_{j \in J_R(P)} \frac{|\xi_{Y,jj}|}{\sqrt{\Sigma_{Y,jj}(P)}} \tag{41}$$

where ξ_Y is drawn from a multivariate Gaussian with mean zero and variance-covariance $\Sigma_Y(P)$, by the central limit theorem, continuous mapping theorem and the fact that the set $\hat{J}_{R;n}$ converges almost surely

[10] The central limit theorem applies pointwise to P, since it applies uniformly over $\mathcal{P}$, under assumption B.1.

to $J_R(P)$. Thus, for any h:

$$\lim_{n\to\infty} Pr_P\left(\max_{\jmath\in\hat{J}_{R;n}} \frac{\left|\tilde{S}^n_{Y,\jmath} - \mu_{Y,\jmath\jmath}(P)\right|}{\sqrt{\hat{\Sigma}^n_{Y,\jmath\jmath}/n}} > c_{1-\alpha}(J) + h\right)$$
$$= Pr_P\left(\max_{j\in J_R(P)} \frac{|\xi_{Y,jj}|}{\sqrt{\Sigma_{Y,jj}(P)}} > c_{1-\alpha}(J) + h\right) \tag{42}$$

Now, since h was arbitrary, it follows that:

$$\lim_{n\to\infty} Pr_P\left(\max_{\jmath\in\hat{J}_{R;n}} \frac{\left|\tilde{S}^n_{Y,\jmath} - \mu_{Y,\jmath\jmath}(P)\right|}{\sqrt{\hat{\Sigma}^n_{Y,\jmath\jmath}/n}} > \rho_{\alpha,\beta}(L^n, U^n, \hat{\Sigma}^n/n)\right)$$
$$\geq \lim_{n\to\infty} Pr_P\left(\max_{\jmath\in\hat{J}_{R;n}} \frac{\left|\tilde{S}^n_{Y,\jmath} - \mu_{Y,\jmath\jmath}(P)\right|}{\sqrt{\hat{\Sigma}^n_{Y,\jmath\jmath}/n}} > c_{1-\alpha}(J)\right) \tag{43}$$

Ultimately implying (20) and thus proposition 5.3. ■

Now, we provide proofs of propositions A.1 and A.2, which concern the validity of both locally simultaneous approaches to inference.

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

$$\begin{aligned}
P_{\mu,\Sigma}\left(\exists\, \jmath \in \hat{J}_R;\ \mu_{Y,\jmath} \notin CS^{LS1}_{1-\alpha;\beta,\jmath}\right) &= P_{\mu,\Sigma}\left(\max_{\jmath\in\hat{J}_R} \frac{|\xi_{Y,\jmath}|}{\sqrt{\Sigma_{Y,\jmath\jmath}}} > c_{1-\alpha+\beta}\left(\hat{J}_c\right)\right)\\
&\leq P_{\mu,\Sigma}\left(\max_{\jmath\in\hat{J}_R} \frac{|\xi_{Y,\jmath}|}{\sqrt{\Sigma_{Y,\jmath\jmath}}} > c_{1-\alpha+\beta}\left(\hat{J}_c\right) \cap G \cap F\right) + P_{\mu,\Sigma}(G^c) + P_{\mu,\Sigma}(F^c)\\
&\leq P_{\mu,\Sigma}\left(\max_{\jmath\in\hat{J}_R} \frac{|\xi_{Y,\jmath}|}{\sqrt{\Sigma_{Y,\jmath\jmath}}} > c_{1-\alpha+\beta}\left(\hat{J}_c\right) \cap G \cap F\right) + \beta_1 + \beta_2\\
&\leq P_{\mu,\Sigma}\left(\max_{\jmath\in J_c} \frac{|\xi_{Y,\jmath}|}{\sqrt{\Sigma_{Y,\jmath\jmath}}} > c_{1-\alpha+\beta}(J_c) \cap G \cap F\right) + \beta_1 + \beta_2\\
&\leq P_{\mu,\Sigma}\left(\max_{\jmath\in J_c} \frac{|\xi_{Y,\jmath}|}{\sqrt{\Sigma_{Y,\jmath\jmath}}} > c_{1-\alpha+\beta}(J_c)\right) + \beta_1 + \beta_2 \leq \alpha - \beta + \beta_1 + \beta_2 = \alpha
\end{aligned}$$

Thus proving the proposition. ■

Proof. **Proof of Proposition A.2.**

Let $J_c(\mu)$ be some set such that $P_{\mu,\Sigma}(\hat{J}_R \subseteq J_c(\mu)) \geq 1 - \beta$. We notice that $J_c(\mu)$ can be taken as follows:

$$\left\{j;\ \min_{j'\in J_R(\mu)} \frac{|\mu_{X,j} - \mu_{X,j'}|}{\sqrt{\mathrm{var}_{j,j'}}} \leq d_{1-\beta}(\Sigma)\right\} \tag{44}$$

Consequently, we notice that $J_c(\mu)$ depends on μ only through $J_R(\mu)$. By inverting 44, we can take a level $1-\beta$ confidence region for $J_R(\mu)$ as follows:

$$\bar{J}_R(\mu) := \left\{ j;\ \min_{j' \in \hat{J}_R} \frac{|X_j - X_{j'}|}{\sqrt{\text{var}_{j,j'}}} \leq d_{1-\beta}(\Sigma) \right\} \tag{45}$$

We notice that the $\hat{J}_c$ defined in 26 is simply the union of forecast intervals in 44 over all plausible $J_R(\mu)$ as collected in $\bar{J}_R(\mu)$. Thus, we have that $P_{\mu,\Sigma}(J_c \subseteq \hat{J}_c) \geq 1-\beta$. Let G be the event that $\max_{j,j' \in R} \frac{|\xi_{X,j} - \xi_{X,j'}|}{\sqrt{\text{var}_{j,j'}}} \leq d_{1-\beta}(\Sigma)$. Moreover, we notice that the event G implies that $J_R(\mu) \subseteq \bar{J}_R(\mu)$, that $J_c(\mu) \subseteq \hat{J}_c$ and that $\hat{J}_R \subseteq J_c(\mu)$. The first implication is a simple consequence of the triangle inequality. The second implication is a consequence of the first. For the final implication, we notice that $\max_{j,j' \in R} \frac{|\xi_{X,j} - \xi_{X,j'}|}{\sqrt{\text{var}_{j,j'}}} \leq d_{1-\beta}(\Sigma)$ implies that:

$$\begin{aligned} &\forall\ j \in J_R(\mu),\ \exists\ j' \in \hat{J}_R \quad \text{s.t.} \quad \frac{|X_j - X_{j'}|}{\sqrt{\text{var}_{j,j'}}} \leq d_{1-\beta}(\Sigma) \\ \implies &\forall\ j \in \hat{J}_R,\ \exists\ j' \in J_R(\mu) \quad \text{s.t.} \quad \frac{|X_j - X_{j'}|}{\sqrt{\text{var}_{j,j'}}} \leq d_{1-\beta}(\Sigma) \end{aligned}$$

Indeed, we have on G that $\hat{J}_R \subseteq J_c \subseteq \hat{J}_c$. Finally, we have:

$$\begin{aligned} P_{\mu,\Sigma}\left(\exists\ \jmath \in \hat{J}_R;\ \mu_{Y,\jmath} \notin CS^{LS2}_{1-\alpha;\beta,\jmath}\right) &= P_{\mu,\Sigma}\left(\max_{\jmath \in \hat{J}_R} \frac{|\xi_{Y,\jmath}|}{\sqrt{\Sigma_{Y,\jmath\jmath}}} > c_{1-\alpha+\beta}\left(\hat{J}_c\right)\right) \\ &\leq P_{\mu,\Sigma}\left(\max_{\jmath \in \hat{J}_R} \frac{|\xi_{Y,\jmath}|}{\sqrt{\Sigma_{Y,\jmath\jmath}}} > c_{1-\alpha+\beta}\left(\hat{J}_c\right) \cap G\right) + P_{\mu,\Sigma}(G^c) \\ &\leq P_{\mu,\Sigma}\left(\max_{\jmath \in J_c} \frac{|\xi_{Y,\jmath}|}{\sqrt{\Sigma_{Y,\jmath\jmath}}} > c_{1-\alpha+\beta}\left(J_c\right)\right) + P_{\mu,\Sigma}(G^c) = \alpha - \beta + \beta = \alpha \end{aligned}$$

Thus proving the proposition. ■

D Inference on Multiple Winners in Microcredit

In this section, we present a stylized setting, based on the microcredit literature, wherein the inference on multiple winners problem becomes particularly relevant in assessing treatment effect heterogeneity across studies. We consider studies on microcredit treatment effects from Bosnia and Herzegovina (Augsburg et al. (2015)), Ethiopia (Tarozzi et al. (2015)), India (Banerjee et al. (2015a)), Mexico (Angelucci et al. (2015)), Mongolia (Attanasio et al. (2015)), and Morocco (Crépon et al. (2015)). All six studies are reviewed in Banerjee et al. (2015b). Depending on the outcome (either business profits, household earnings, or consumption), some of the above studies identify statistically significant treatment effects, while others do not. Pritchett and Sandefur (2015) suggest that this heterogeneity across studies is due to contextual differences between study sites. An alternative explanation is that those studies selected as statistically significant suffer from a winners' curse, and that statistical significance arises due to a version of the multiple testing problem. We list the observational treatment effects of microcredit on household consumption from each of the six studies below:

Study	Treatment Effect
Bosnia (Augsburg et al. (2015))	0.193** (0.0793)
Ethiopia (Tarozzi et al. (2015))	-0.0280 (0.0502)
India (Banerjee et al. (2015a))	0.0415 (0.0295)
Mexico (Angelucci et al. (2015))	0.0845** (0.0429)
Mongolia (Attanasio et al. (2015))	-0.0317*** (0.00764)
Morocco (Crépon et al. (2015))	0.194** (0.0895)

Table 7: Observational Microcredit Treatment Effects on Household Income per Pritchett and Sandefur (2015)

We notice that, of the six treatment effects reported in Pritchett and Sandefur (2015), four are statistically significant at the 5% level, namely Augsburg et al. (2015), Angelucci et al. (2015), Attanasio et al. (2015), and Crépon et al. (2015). Empirically, we find that after correcting for cutoff-based selection, treatment effects remain statistically significant for all of these studies excluding Angelucci et al. (2015) and Attanasio et al. (2015). We demonstrate, via simulation, that a naive multiple testing correction may undercover following the selection described above, motivating a need for post-selection inference procedures. In these simulations, we compare three methods for inference: a naive bonferroni correction on only the selected programs, projection inference (as developed in Andrews et al. (2023), and our novel two-step inference procedure.

D.1 Empirical Results

We apply both our novel two-step confidence set and the naive projection confidence set to the four statistically significant studies identified in Pritchett and Sandefur (2015). In particular, we observe the following confidence intervals for the treatment effect in each selected study:

Study	Treatment Effect	Two-Step CI	Projection CI
Augsburg et al. (2015)	0.193	(0.005, 0.381)	(0, 0.385)
Crépon et al. (2015)	0.194	(0.006, 0.382)	(0.002, 0.386)
Angelucci et al. (2015)	0.084	(-0.103, 0.272)	(-0.108, 0.277)
Attanasio et al. (2015)	-0.0317	(-0.219, 0.156)	(-0.225, 0.161)

Table 8: Revised Observational Treatment Effects Confidence Sets for Statistically Significant Treatment Effects from Pritchett and Sandefur (2015)

Even after making the appropriate post-selection inference correction, we continue to observe statistically significant treatment effects in two out of the four selected studies, namely Augsburg et al. (2015) and Crépon et al. (2015). These findings suggest that, in part, the heterogeneity in treatment effects observed by Pritchett and Sandefur (2015) cannot be explained by the selection of statistically significant studies. Rather, treatment effect heterogeneity may indeed be due to variation in treatment design and setting.

D.2 Simulation Results

In this section, we develop a range of simulations based on the empirical setting introduced above. In particular, we consider three approaches to inference: projection accounting for selection, projection not accounting for selection, and the two-step approach. We provide further discussion of the first and third approaches in sections 4 and 5, respectively. The first approach is similar in spirit to a Bonferroni correction over all indices J. The second approach is similar in spirit to a Bonferroni correction over only the selected indices. Thus, the second approach can be thought of as correcting for the multiple testing component of the inference on multiple winners problem, but not the post-selection component. We find that projection accounting for selection provides wide marginal confidence intervals and overcovers, even asymptotically. On the other hand, projection not accounting for selection provides short marginal confidence intervals and undercovers in the finite sample, but provides valid coverage asymptotically. Finally, our novel two-step approach displays similar properties to projection accounting for selection, but does not overcover asymptotically. This is a key feature of the two-step approach, as discussed in proposition 5.3.

In the simulations that follow, we fix treatment effects normalized by standard errors, and select the studies for which this normalized value exceeds the critical value 1.96. We vary the variance covariance Σ that results from the transformation described above, scaling by $1/n$ for a given sample size n. The figure below plots coverage probability for the three approaches described above with n varying across the x-axis:

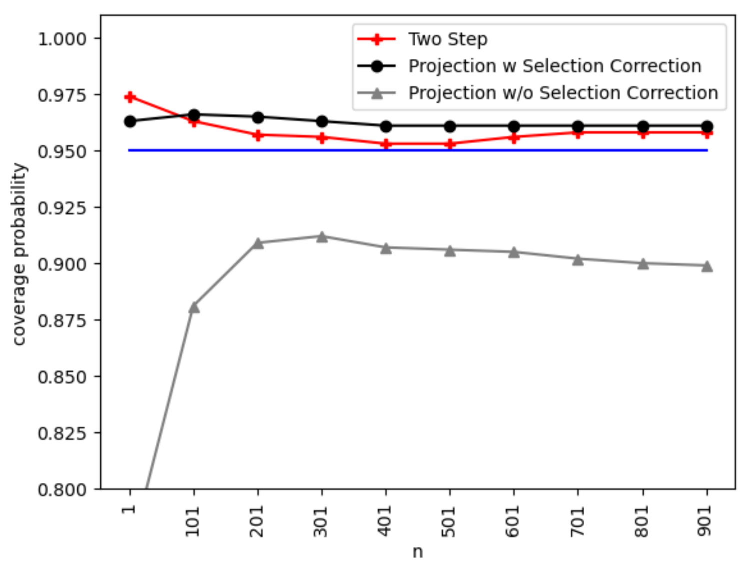


Figure 5: Coverage Probability in the Microcredit Application for Various Inference Procedures

We see reductions in over-coverage error by up to 50%. We see similar reductions in confidence set length. We consider the mean volume of our confidence sets across the three inference procedures described, finding that volume decreases with n, and that confidence set volume converges for the three methods asymptotically. In general, the two-step approach to inference provides shorter confidence sets than projection with a selection correction, while projection without a selection correction generally provides the smallest confidence sets in the finite sample.

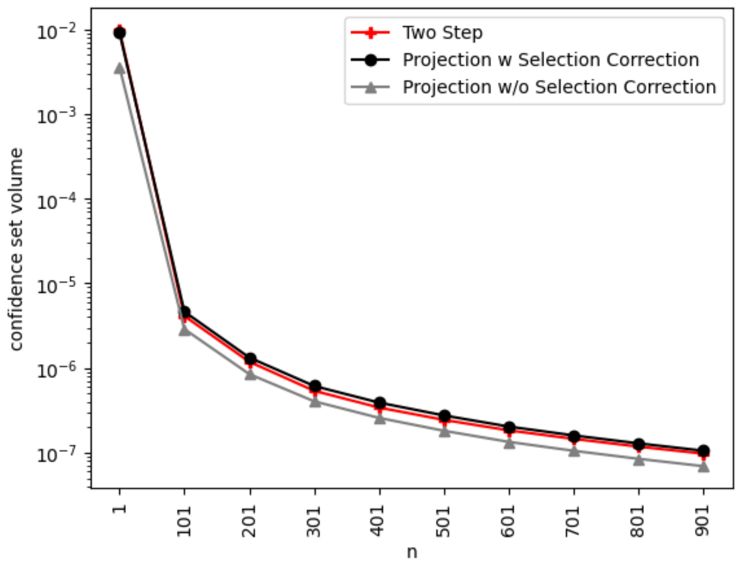


Figure 6: Confidence Set Volume in the Microcredit Application for Various Inference Procedures

E 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 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.

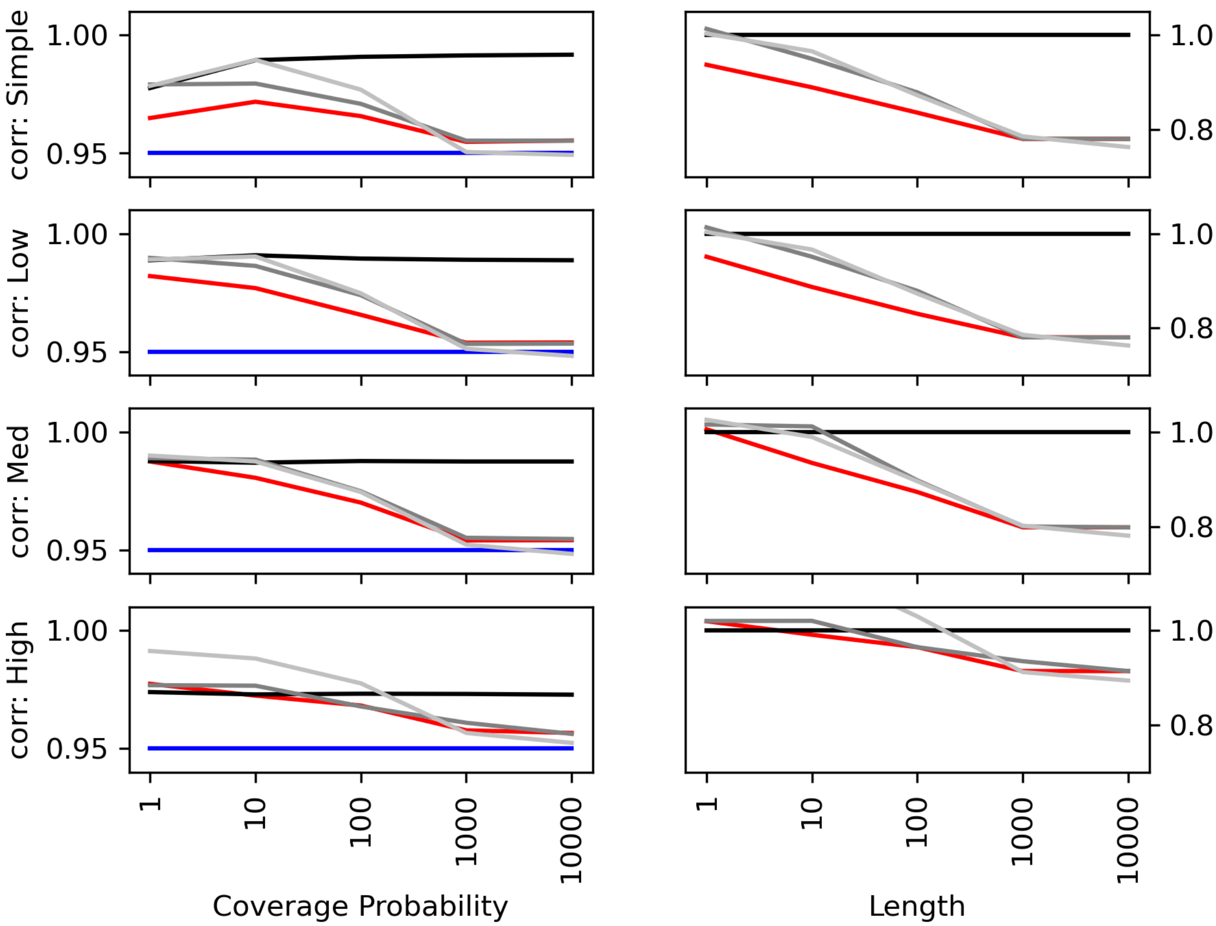


Figure 7: Coverage probability and length in design 1. CI lengths are presented as fractions of the projection CI length.

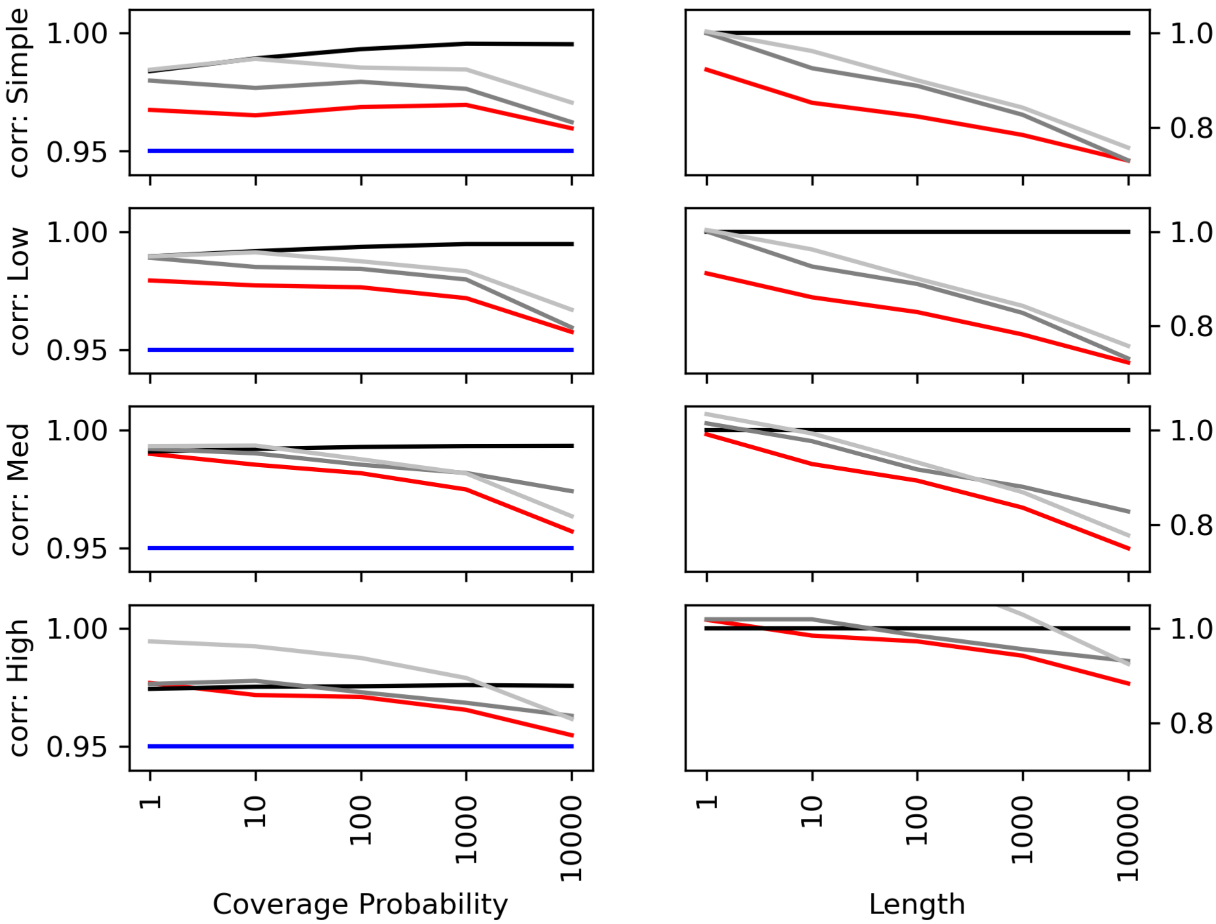


Figure 8: Coverage probability and length in design 2. CI lengths are presented as fractions of the projection CI length.

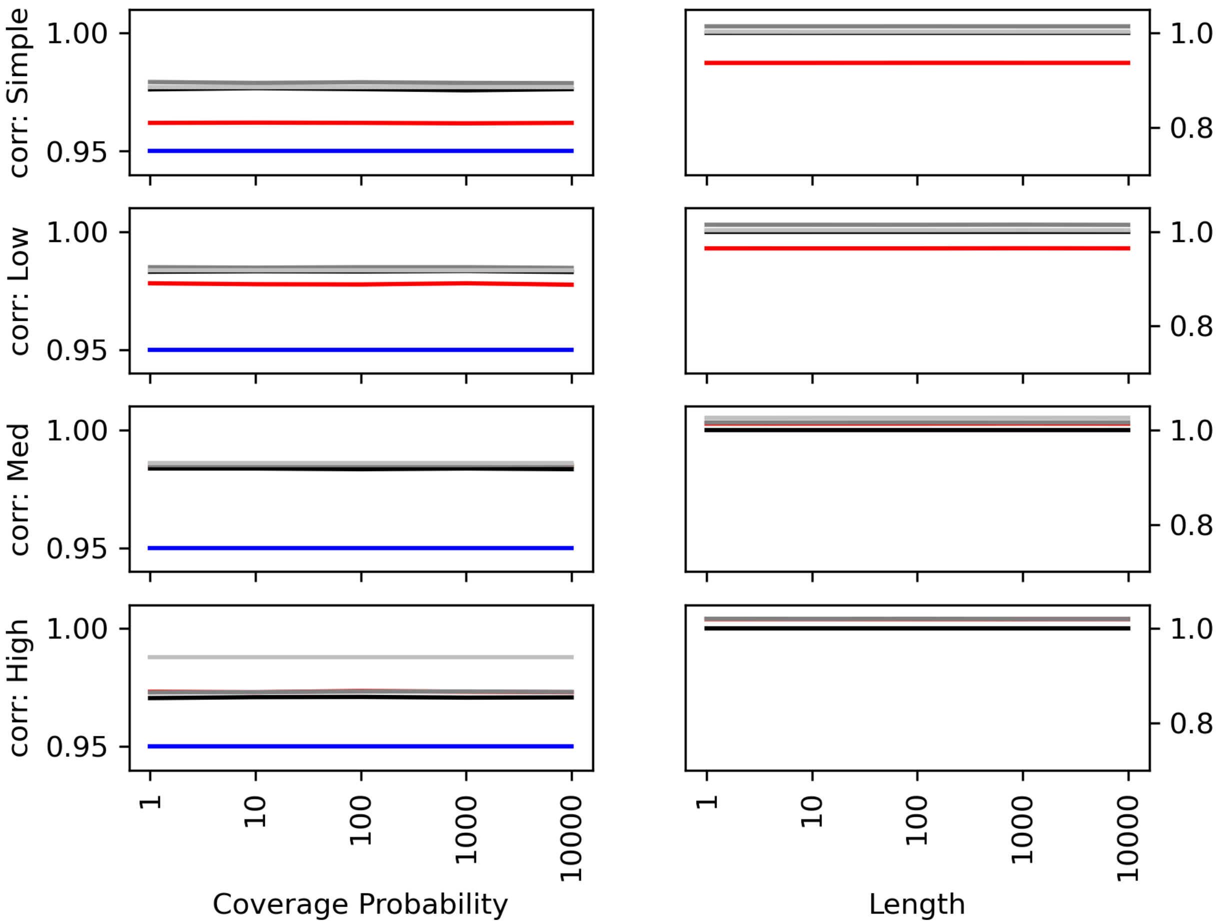


Figure 9: Coverage probability and length in design 2. CI lengths are presented as fractions of the projection CI length.

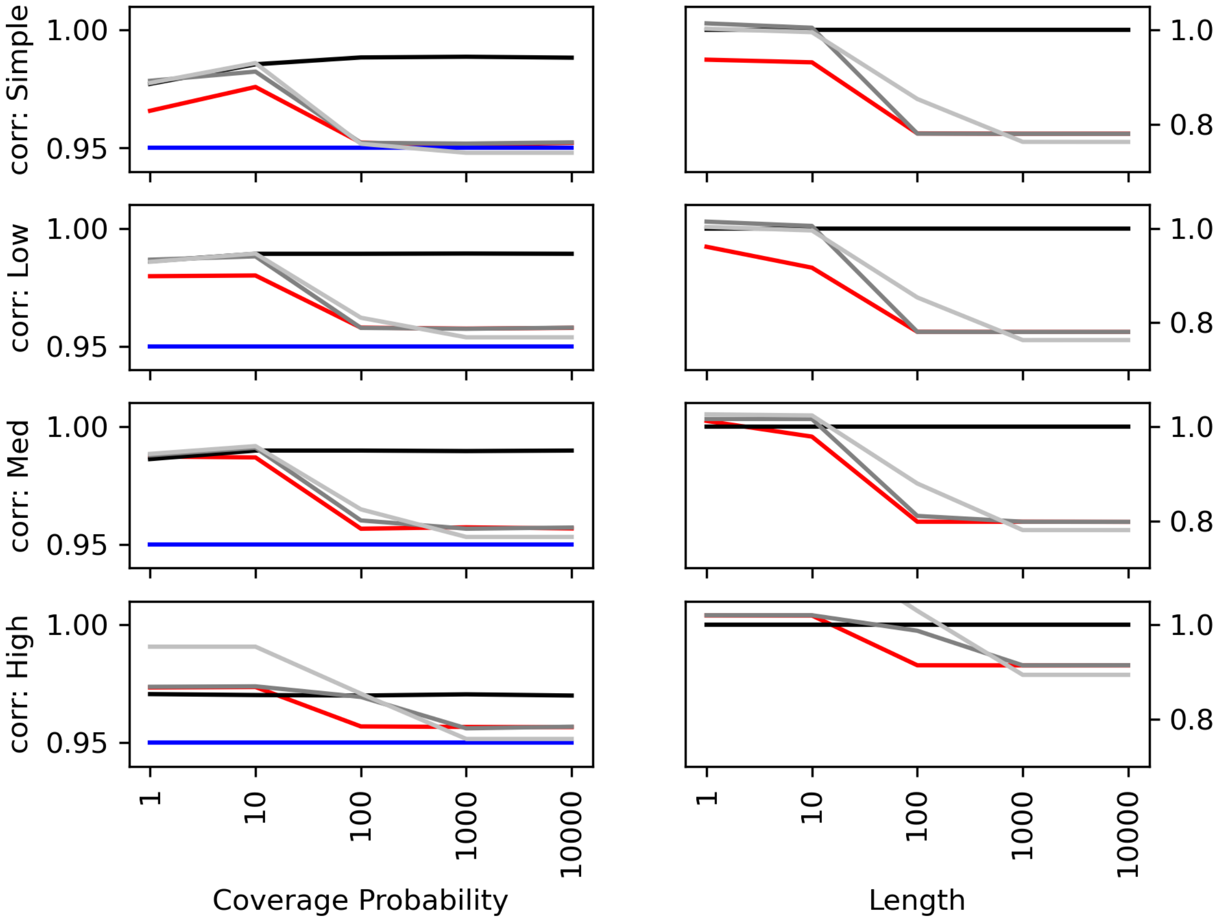


Figure 10: Coverage probability and length in design 2. CI lengths are presented as fractions of the projection CI length.

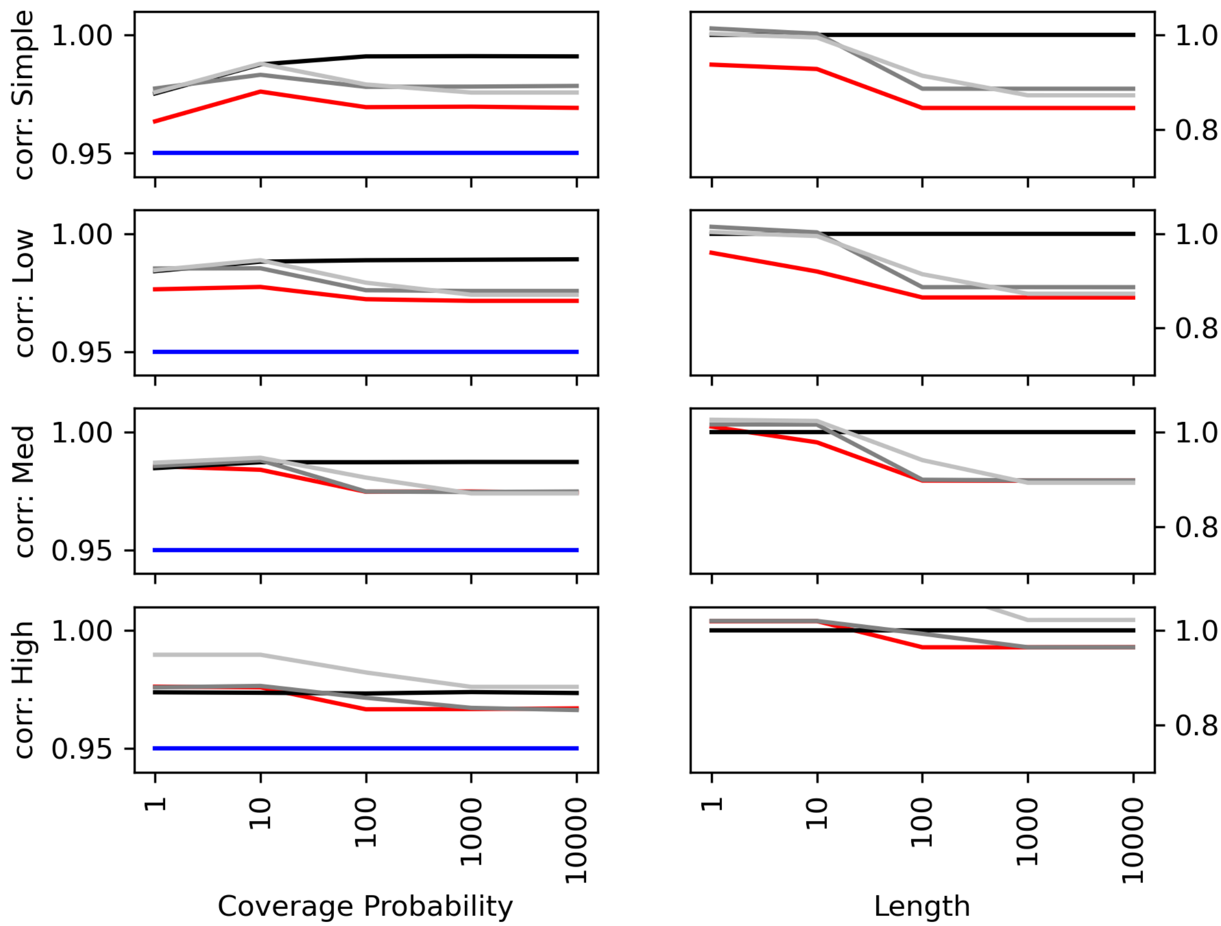


Figure 11: Coverage probability and length in design 2. CI lengths are presented as fractions of the projection CI length.

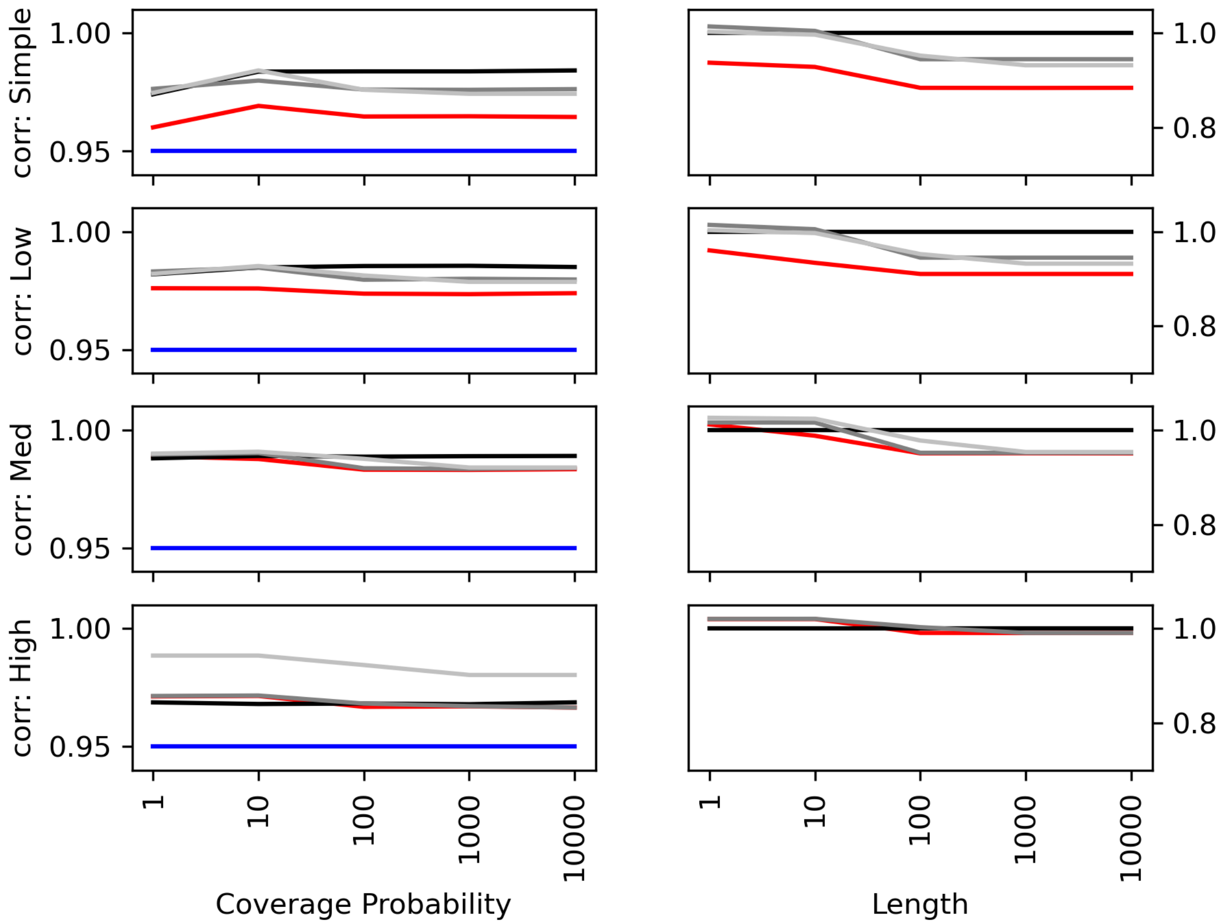


Figure 12: Coverage probability and length in design 2. CI lengths are presented as fractions of the projection CI length.

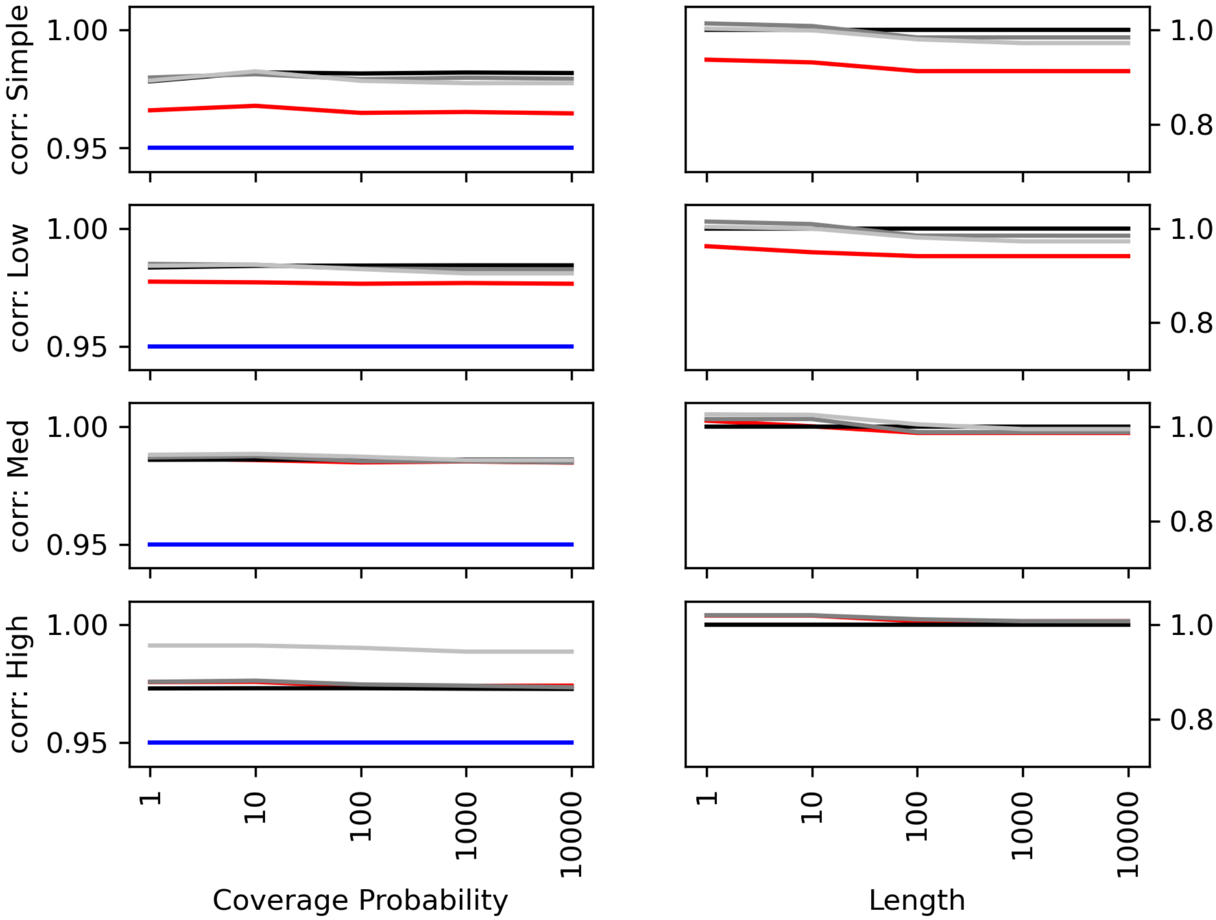


Figure 13: Coverage probability and length in design 2. CI lengths are presented as fractions of the projection CI length.