Joint Confidence Regions for Rankings based on Correlated Estimates

Shaine Rosewel Paralis Matala

A Thesis proposal submitted to UP Diliman School of Statistics

In Partial Fulfillment of the Requirements for the Degree of
Master of Science in Statistics

School of Statistics
University of the Philippines
November 2025

Contents

1 Introduction

Ranked data are commonly of interest because they allow readers to compare populations based on ranked quantities. For example, top universities across the globe may be identified based on their institutional performance indicator; states may receive appropriate intervention according to their relative rank based on average travel times to work; and senatorial candidates who are likely to be granted a seat in the office can be reported by public opinion polling bodies prior to elections. Whenever data on the entire population are unavailable, the ranks are based on estimates rather than the unknown parameters. Consequently, their overall uncertainty—expressed through joint confidence intervals—should also be quantified. Marginally, these intervals provide information on the possible range of each rank while jointly, they facilitate comparing all ranks simultaneously rather than reporting them in isolation.

Several studies have addressed this concern through different techniques. Some approaches like those by Klein et al. (2020), Mohamad et al. (2019), Mogstad et al. (2024), Andersson et al. (1998), and Lyhagen & Ahlgren (2020), relied solely on the estimates and their standard errors, constructing joint confidence intervals either for the estimated quantities or directly for the ranks themselves. Others incorporate model-based uncertainty to account for dependencies inherent in the data structure. These include the works of Goldstein & Spiegelhalter (1996), who utilized conditioning through multilevel models where the ranked quantities are treated as residual effects. Hall & Miller (2009), on the other hand, developed a bootstrap algorithm that allows for the assumption of independence despite its potential violation.

Assuming independence when constructing joint confidence regions for

estimators that are, in fact, correlated may lead to overly conservative and thus wider intervals, implying greater uncertainty—contrary to what is desired. It is therefore important to account for potential dependencies, as demonstrated by Goldstein & Spiegelhalter (1996). However, in their case, the estimators were treated as latent variables. In contrast, the present work focuses on estimators that are observable or directly measurable from the data.

The more general setting, which to our knowledge has not yet been explored, allows for a certain degree of correlation among the estimates to be ranked. This thesis develops a procedure capable of handling such dependencies while maintaining coverage close to the nominal level and producing relatively narrow joint confidence intervals. The proposed methodology uses only the observed estimators and their corresponding standard errors. Although it also employs a parametric bootstrap—commonly used to estimate overall rank uncertainties (e.g., Mohamad et al. (2019), Mogstad et al. (2024), Andersson et al. (1998), Lyhagen & Ahlgren (2020))—our implementation differs from these existing approaches.

1.1 Objective

This research builds upon Klein et al. (2020)'s methodology by extending the set of joint confidence intervals used to capture uncertainty in overall rankings. In particular, it intends to:

- Develop a procedure to construct joint confidence intervals for the ranks and the ranked parameters when the estimates to be ranked may be correlated.
- Evaluate the performance of the proposed approaches under various parameter settings.
- Apply the proposed approaches to a real-life example.

1.2 Significance

In order to obtain joint confidence sets for overall ranks, Klein et al. (2020) estimates confidence intervals for the unknown parameters, with a joint coverage probability of at least $1 - \alpha$. Their goal is to produce confidence intervals that collectively produce a small difference between the upper and the lower bound to yield tighter joint uncertainty for ranks. In the same study, they used individual confidence intervals of the form $\hat{\theta}_k \pm z_{\alpha/2} \times \sigma_k$ for the population parameters $\theta_1, \theta_2, \ldots, \theta_K$, assuming that the estimates of these parameters $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K$ are independently distributed. This approach, while simple, is no longer applicable to the case where $\hat{\theta}_k$s may be correlated. In many cases, assuming independence in the presence of correlated quantities leads to conservative confidence intervals resulting in wider intervals which imply a higher uncertainty in overall ranks.

An example for which the estimates of the parameters are likely to be correlated is the case of ranking senatorial candidates in the Philippines. In such a context, the assumption of independence is limiting as it treats vote shares as statistically independent across contenders. Although senators are elected using Multiple Non-transferable Vote system (MNTV)—where candidates are voted for individually regardless of partisan membership and alliances (Ravanilla & Hicken, 2023)—David & Legara (2015) demonstrated that candidates with a name-recall advantage, such as media celebrities, incumbents, and members of dynastic families, received majority of the votes in the 2010 senatorial elections. In that year, media personalities Bong Revilla and Jinggoy Estrada secured the top spots. A similar pattern was observed in 2019, when Cynthia Villar and Grace Poe, both with prominent surnames, garnered the most votes; and again in 2022, when media figures Robin Padilla and Ramon Tulfo ranked among the top three. They also added that in weak-party systems, candidates who belong to the same political alliance or ticket commonly co-occur in ballots and hence perform with similarity.

Klein et al. (2020) also noted that although it is difficult to make gener-

alizations about strong relationships between travel times to work, certain patterns are apparent. States with large unpopulated land areas and relatively few high-density population centers tend to report shorter travel times. In contrast, longer travel times are typically observed in highly urbanized states with large populations and high population densities. Geographic location also appears to play a role—for instance, many states with shorter travel times are located in the Mountain and Central regions, whereas majority of those with longer travel times are concentrated along the East Coast. These observations suggest the presence of potential spatial structures. Accounting for these patterns allows for a more realistic assessment of uncertainty in the estimated ranks for similar use cases.

1.3   Scope and Limitations

This study proposes methodologies to construct joint confidence regions to quantify uncertainty in overall ranks, building upon the main results of Klein et al. (2020). Unlike the study of Klein et al. (2020), however, this thesis addresses the case where the parameter estimates may be correlated. Moreover, this study also proposes a methodologies to construct joint confidence regions for the ordered parameters. It applies the parametric bootstrap in implementing the procedures, with a strong emphasis on maintaining the desired coverage. However, several limitations must be acknowledged. First, the framework assumes that the data are generated from a multivariate normal distribution. Second, a limited set of correlation structures is examined to illustrate how different dependence assumptions may affect the resulting joint confidence sets.

2 Related Literature

2.1 Ranking Problem

In the problem of estimating ranks of several unknown real-valued parameters $\theta_1, \theta_2, \ldots, \theta_K$, it is desired to rank these K parameters from smallest to largest, $\theta_{(1)} < \theta_{(2)} < \ldots < \theta_{(K)}$. Let $r_1, r_2, \ldots, r_K$ be the true unknown ranks of $\theta_1, \theta_2, \ldots, \theta_K$. For example, if $\theta_{(2)} = \theta_4$, then $r_4 = 2$. Our first goal is to derive a joint confidence region for $\theta_1, \theta_2, \ldots, \theta_K$ and, using the derived confidence region, obtain joint confidence intervals for the $r_1, r_2, \ldots, r_K$ using a result from Klein et al. (2020). In deriving a joint confidence region for $\theta_1, \theta_2, \ldots, \theta_K$, we will make use of the corresponding point estimates $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K$.

There are several real-life applications of this problem. For example, the study of Klein et al. (2020) illustrates the problem of ranking the US states in terms of travel time to work. In this context, θ_k is the true mean travel time to work for state k, and we wish to find a joint confidence statement about $r_1, r_2, \ldots, r_K$. Our second goal is to derive a joint confidence region for the ordered parameters $\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(K)}$. In the context of senatorial elections where θ_k denotes the proportion of votes earned by a candidate, confidence intervals on ordered parameters help determine if one candidate's lead is statistically significant. If the confidence intervals for two candidates' support levels do not overlap, it is highly likely (at the specified confidence level) that the candidate with the higher point estimate is the true winner. Overall uncertainty also highlights that in election polls, reported candidate rankings that seem decisive may in fact be much less stable once uncertainty is considered, and hence it is possible that these rankings do not align with

the actual election outcomes. This is particularly important when discussing the potential influence of such polls on voter behavior—commonly referred to as the bandwagon effect—because apparent leads may encourage support for candidates who are not actually ahead, underscoring the value of clearly communicating the uncertainty in poll-based rankings. As to Andersson et al. (1998), in institutions like hospitals that are ranked according to mortality rates for a particular ailment, failure to report uncertainties may create a misleading impression, potentially causing fear among patients and casting a negative light on hospitals that appear at the bottom of the ranking.

A mathematical definition of r_k is as follows:

$$r_k = \sum_{j=1}^{K} I(\theta_j \leq \theta_k) = 1 + \sum_{j:j\neq k} I(\theta_j \leq \theta_k), \qquad \text{for } k = 1, 2, \ldots, K. \quad (2.1)$$

Let $\hat{r}_k$ denote the estimated rank of the kth observation. This can be computed determined based on

$$\hat{r}_k = 1 + \sum_{j:j\neq k} I(\hat{\theta}_j \leq \hat{\theta}_k), \qquad \text{for } k = 1, 2, \ldots, K. \quad (2.2)$$

In this chapter, we review the various approaches to solve this problem of finding a joint confidence region for $r_1, r_2, \ldots, r_K$.

2.2 Klein's Joint Confidence Region for Overall Ranking Uncertainty

The study of Klein et al. (2020) assumes that $\hat{\theta}_k \sim N(\theta_k, \sigma_k^2), k = 1, 2, \ldots, K$ where θ_k is unknown but σ_k^2 is known. Their solution hinges on the fact that the uncertainty in the ranks is determined by the uncertainty in the parameters (Mogstad et al. (2024)). We now describe their methodology.

Suppose that for each $k \in \{1, 2, \ldots, K\}$ there exists values L_k and U_k such that

$$\theta_k \in (L_k, U_k), \, k = 1, 2, \ldots, K. \tag{2.3}$$

Moreover, define the quantities $I_k, \Lambda_{Lk}, \Lambda_{Rk}, \Lambda_{Ok}$ as follows:

$$\left.\begin{array}{c} I_k = \{1, 2, \ldots, K\} - \{k\}, \\ \Lambda_{Lk} = \{j \in I_k : U_j \leq L_k\}, \\ \Lambda_{Rk} = \{j \in I_k : U_k \leq L_j\}, \\ \Lambda_{Ok} = \{j \in I_k : U_j > L_k \text{ and } U_k > L_j\} = I_k - \{\Lambda_{Lk} \cup \Lambda_{Rk}\} \end{array}\right\}$$

which implies the following:

1. $j \in \Lambda_{Lk} \leftrightarrow (L_j, U_j) \cap (L_k, U_k) = \emptyset$ and (L_j, U_j) lies to the left of (L_k, U_k);
2. $j \in \Lambda_{Rk} \leftrightarrow (L_j, U_j) \cap (L_k, U_k) = \emptyset$ and (L_j, U_j) lies to the right of (L_k, U_k);
3. $j \in \Lambda_{Ok} \leftrightarrow (L_j, U_j) \cap (L_k, U_k) \neq \emptyset$
4. $\Lambda_{Lk}, \Lambda_{Rk}$, and Λ_{Ok} are mutually exclusive, and $\Lambda_{Lk} \cup \Lambda_{Rk} \cup \Lambda_{Ok} = I_k$

Let $|A|$ denote the number of elements in a set A. If the condition in (2.3) holds, the main result from Klein et al. (2020) gives a range for the value of r_k for each $k \in \{1, 2, \ldots, K\}$ as follows:

$$r_k \in \{|\Lambda_{Lk}| + 1, |\Lambda_{Lk}| + 2, |\Lambda_{Lk}| + 3, \ldots, |\Lambda_{Lk}| + |\Lambda_{Ok}| + 1\} \tag{2.4}$$

Note that the number of elements in the range given in (2.4) is $|\Lambda_{Ok}| + 1$. Since the smaller difference between U_k and L_k leads to a smaller $|\Lambda_{Ok}|$, narrower confidence intervals for $\theta_1, \theta_2, \ldots, \theta_K$ are desirable.

Suppose that for random quantities L_k and U_k the event defined in (2.3) satisfies the following probability condition:

$$P\left[\bigcap_{k=1}^{K} \{\theta_k \in (L_k, U_k)\}\right] \geq 1 - \alpha, \tag{2.5}$$

then, by the result of Klein et al. (2020), it also follows that

$$P\left[\bigcap_{k=1}^{K} \{r_k \in \{|\Lambda_{Lk}| + 1, |\Lambda_{Lk}| + 2, |\Lambda_{Lk}| + 3, \ldots, |\Lambda_{Lk}| + |\Lambda_{Ok}| + 1\}\}\right] \geq 1 - \alpha. \tag{2.6}$$

Due to the assumption of normality on $\hat{\theta}_k$ as well as the fact that σ_k is assumed known, Klein et al. (2020) set the confidence intervals (L_k, U_k) for θ_k to be of the form $\hat{\theta}_k \pm t \times \sigma_k$ for $k \in \{1, 2, \ldots, K\}$. One may use the Bonferroni approach to choose t. If such an approach is used, the choice of t that would satisfy (2.5) is $t = z_{\alpha/2K}$. Another choice of t is one that exploits the independence assumption on $\hat{\theta}_k$. Such a choice is given by $z_{\gamma/2}$ where $\gamma = 1 - (1-\alpha)^{1/K}$. To see why this is the case, note that if we define Y_k as follows:

$$Y_k = \frac{\left(\hat{\theta}_k - \theta_k\right)}{\sigma_k} \sim N(0, 1), \quad k = 1, 2, \ldots, K \tag{2.7}$$

then we get

$$\begin{aligned}
\prod_{k=1}^{K} P\left(|Y_k| \leq t\right) &= \prod_{k=1}^{K} P\left(\hat{\theta}_k - t \cdot \sigma_k \leq \theta_k \leq \hat{\theta}_k + t \cdot \sigma_k\right) \\
&= \prod_{k=1}^{K} 1 - \left[1 - (1-\alpha)^{1/K}\right] \tag{2.8} \\
&= 1 - \alpha.
\end{aligned}$$

2.3 Alternative Approaches for Ranking Uncertainty

While Klein's approach provides one framework for constructing joint confidence regions for ranks, several other studies have explored related problems using different formulations or assumptions. These alternative methods vary in whether they account for dependence structures, rely on model-based estimation, or use resampling techniques such as the bootstrap.

2.3.1 Calculated Measure

Other studies with similar concern include that of Andersson et al. (1998) who suggest the use of a statistic C which quantifies the number of positions ranked populations would on average change their order due to random variation. They calculate the measure using a bootstrap approach. Working on risk ratios p_k of K units, Andersson et al. (1998) draw B bootstrap proportions $\hat{p}_k^*$ from (2.9):

$$\hat{p}_{bk}^* \sim N\left(\hat{p}_k, \frac{\hat{p}_k(1 - \hat{p}_k)}{n_k}\right), \quad k = 1, 2, \ldots, K; \ b = 1, 2, \ldots, B \qquad (2.9)$$

and $\hat{p}_k$ and n_k are the sample proportion and sample size, respectively, corresponding to the kth unit. For each bootstrap iteration b, they sort the set $\hat{p}_{b1}^*, \ldots, \hat{p}_{bK}^*$ to get the corresponding ranks $r_{bk}^*, k = 1, 2, \ldots, K$. The difference d_{bk} between the original and bootstrap rank is then calculated as $d_{bk} = |\hat{r}_k - \hat{r}_{bk}^*|$, where $\hat{r}_k$ is the rank of $\hat{p}_k$ in the set $\{\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_K\}$. In turn, $\bar{d}_k$ is obtained by taking the average of d_{bk} across the bootstrap samples. Finally, the overall measure C is calculated as the average of $\bar{d}_k$ across all K units.

2.3.2 Pairwise Difference

Mohamad et al. (2019), requiring only the mean estimates and their corresponding standard errors similar to Klein et al. (2020), apply Tukey's Honest Significant Difference (HSD) to test $H_0 : \theta_j - \theta_k = 0$ for all $j \neq k \in 1, \ldots, K$ at level α. Tukey's HSD is typically used to provide simultaneous confidence statements about the differences between the means while controlling the family-wise error rate (FWER). Mohamad et al. (2019) come up with a joint confidence set for ranks expressed as

$$\left(1 + \# \left\{ j : \frac{\hat{\theta}_j - \hat{\theta}_k}{\sqrt{\sigma_j^2 + \sigma_k^2}} > q_{1-\alpha} \right\}, \ K - \# \left\{ j : \frac{\hat{\theta}_j - \hat{\theta}_k}{\sqrt{\sigma_j^2 + \sigma_k^2}} < -q_{1-\alpha} \right\} \right),$$

$$(2.10)$$

$$\text{for } k = 1, 2, \ldots, K$$

where $q_{1-\alpha}$ is the $(1 - \alpha)$ quantile of the distribution of the studentized range,

$$\max_{j,k=1,\ldots K} \frac{|\theta_j - \theta_k|}{\sqrt{\sigma_j^2 + \sigma_k^2}}.$$

In (2.10), the notation $\#\{\cdot\}$ counts the number of pairwise hypotheses that are rejected according to Tukey's HSD, which determines the lower and upper bounds of the confidence interval for the rank of $\hat{\theta}_k$. Their method has simultaneous coverage of at least $1 - \alpha$ and exactly $1 - \alpha$ when $\theta_1 = \theta_2 = \cdots = \theta_K$. However, their approach tends to be overly conservative, showing coverage levels between 0.996 and 1.0 at a 0.90 nominal level in simulations, when the θs differ. They also demonstrated that as the true differences increase from 0 to 0.5, the coverage quickly increases from the nominal level to 1. As a remedy, they proposed a rescaling technique that brings the coverage closer to the nominal level, though it remains conservative (e.g., from 1.0 to 0.978, from 0.998 to 0.961—at 0.90 confidence level).

Mogstad et al. (2024) presents another technique that closely resembles the procedure by Klein et al. (2020) and Mohamad et al. (2019). However, they define ranks in the opposite way (i.e., larger rank value for lower estimate). They construct the rectangular confidence region in (2.11), from the pairwise differences of the estimators $\hat{\theta}_1, \ldots, \hat{\theta}_K$ and an estimator of the variance of $\hat{\theta}_j - \hat{\theta}_k$. The quantities $\hat{\theta}_1, \ldots, \hat{\theta}_K$ need not be independent. Let P_k be the distribution from which $\hat{\theta}_k$ is estimated; and let $\hat{P}_k$ denote the estimate of P_k. The joint confidence intervals for $r_1, \ldots, r_K$ are obtained from the joint confidence intervals of the pairwise difference:

$$C(1-\alpha, S) = \prod_{(j,k)\in S} \left[\hat{\theta}_j - \hat{\theta}_k \pm \sqrt{\sigma_j + \sigma_k} \, L^{-1}(1-\alpha, S, \hat{P}) \right], \tag{2.11}$$

$$S \subseteq \{(j,k) \in K \times K : j \neq k\}$$

where $L^{-1}(1-\alpha, S, P)$ is the $(1-\alpha)$-quantile corresponding to the CDF:

$$L(x, S, P) = P\left\{ \max_{(j,k)\in S} \frac{|\hat{\theta}_j - \hat{\theta}_k - (\theta_j - \theta_k)|}{\sqrt{\sigma_j^2 + \sigma_k^2}} \leq x \right\}. \tag{2.12}$$

They added that if the estimators $\hat{\theta}_1, \ldots, \hat{\theta}_K$ are jointly asymptotically normally distributed, then the quantiles $L^{-1}(1-\alpha, S, \hat{P})$, can be computed from the limiting distributions of the max-statistics shown in (2.12), through resampling methods. Mogstad et al. (2024) show that their approach generally leads to a confidence set that is narrower than that of Klein et al. (2020).

2.3.3 Accounting for Data Dependencies

Some approaches explicitly account for dependencies in the data. Goldstein & Spiegelhalter (1996) use multilevel models, in the context of ranking education and health institutions (e.g., schools, hospitals, medical practitioners, etc.), to address the hierarchical nature of data structures associated with institutional performance. Rank uncertainty is presented through a visualization in which non-overlap of confidence intervals conveyed a significant difference between compared institutions (Goldstein & Healy, 1995). In an alternative approach, along with institution effect estimation through Gibbs sampling, the rank is obtained for each iteration. Their example illustrates that while the multilevel model made individual estimates more accurate, it also had the effect of making the ranks even more uncertain.

Zhang et al. (2013) analyze U.S. age-adjusted cancer incidence and mortality rates across states and counties by computing individual and overall simultaneous confidence intervals for age-adjusted health index using the

Monte Carlo method. Because many health conditions are age-dependent, they use age-adjusted rates to minimize the confounding effect of age differences when comparing different population groups. They also extend their method to handle cases where only the adjusted rates and confidence intervals are available, aligning it more closely with the approach of Klein et al. (2020). Mohamad et al. (2019) show their technique to result in joint confidence sets with very low coverage probabilities and which are only able to reach the nominal level when differences among the means are large enough.

Hall & Miller (2009) mention that in some use cases such as institutions ranking, dependencies can be accommodated through conditioning, similar to the above approaches. However, in genomics where data on expression levels of different genes from the same individual are generally not independent, they suggest using an "independent component" version of the bootstrap on the sample, where m-out-of-n bootstrap (m < n) is applied as though the ranked variables were statistically independent. They show this to perform at its best when a reasonable level of correlation is present among the variables.

Bazylik et al. (2025), in their recent study, tackle the ranking of political candidates or parties using the estimated share of support each one receives in surveys. They use the multinomial distribution to develop confidence sets for finite samples and explore the use of the bootstrap in the case of approximately large samples. They address the dependence attributed to the success probabilities of different categories by using their proposed bootstrap algorithm. Their simulations show that bootstrap-based confidence sets may have coverage probability below the nominal level despite their being excessively wide. In contrast, the finite-sample confidence sets have coverage probability at least as large as expected and may even be relatively shorter.

3 Methodology

This section introduces the proposed methodologies to obtain joint confidence intervals that can be used to quantify uncertainty for the unknown overall true ranking. It addresses the scenario when the estimates of the parameters being ranked are correlated to certain degrees. Section 3.1 develops the procedure used to compute the joint confidence region for $\theta_1, \theta_2, \ldots, \theta_K$. These include a non-rank and rank-based methods. Section 3.2, on the other hand, develops the procedure to compute the joint confidence region for $\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(K)}$. Section 3.3 discusses some possible correlation structures to be considered for the estimates. Sections 3.4 and 3.5 discuss how to evaluate the proposed methodology and the simulation settings to be used.

3.1 Proposed methodology to compute the joint confidence region for the unordered parameters

In this section, we adopt the notations defined in Section 2.1 and 2.2. Moreover, we define $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K)'$ and assume that $\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_K)'$ is unknown and $\boldsymbol{\Sigma}$ is a known $K \times K$ positive definite matrix. The diagonal elements of $\boldsymbol{\Sigma}$ are $\sigma_1^2, \ldots, \sigma_K^2$. We note that in the literature on inferences on the ranks, it is customary to assume that the variances are known. As previously mentioned, the setup in this thesis allows the estimates to be correlated, which is in contrast with that of Klein et al. (2020), which assumes independence among the estimates. In other words, in this thesis, the covariance matrix $\boldsymbol{\Sigma}$ need not be a diagonal matrix. Similar to the study of Klein et al. (2020), we shall first derive simultaneous confidence intervals for $\theta_1, \theta_2, \ldots, \theta_K$ of the form

$$\mathfrak{R}_1 = [\hat{\theta}_1 \pm t \times \sigma_1] \times [\hat{\theta}_2 \pm t \times \sigma_2] \times \cdots \times [\hat{\theta}_K \pm t \times \sigma_K]. \qquad (3.1)$$

We want the joint confidence region in (3.1) to satisfy the following probability condition:

$$P\left(\hat{\theta}_k - t \cdot \sigma_k \leq \theta_k \leq \hat{\theta}_k + t \cdot \sigma_k, \ \forall \ k = 1, 2, \ldots, K\right) = 1 - \alpha. \qquad (3.2)$$

Equivalently, we require

$$P\left(\max_{k=1,2,\ldots,K} \left|\frac{\hat{\theta}_k - \theta_k}{\sigma_k}\right| \leq t\right) = 1 - \alpha. \qquad (3.3)$$

We use the parametric bootstrap to estimate the quantile t that will be used to construct the confidence intervals while controlling the coverage to be around the nominal level. In implementing the parametric bootstrap, we sample from the multivariate normal distribution with $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\Sigma}$ as mean and variance, respectively. The idea is also conceptually similar to that of Mogstad et al. (2024) who use resampling to obtain the quantile in a different context. Algorithm 1 describes the procedure to compute the joint confidence region based on unordered estimates.

Once the confidence intervals in (3.1) have been obtained, we can then use the result of Klein et al. (2020) in (2.6) to get the lower and upper bounds on the ranks $r_k, k = 1, 2, \ldots, K$. That is, we also get ajoint confidence region for $r_1, r_2, \ldots, r_K$.

Algorithm 1 Computing the joint confidence region for the unordered parameters

Let the data be represented by $\hat{\boldsymbol{\theta}} = \left(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K\right)'$ and suppose that $\boldsymbol{\Sigma}$ is known

1: **for** $b = 1, 2, \ldots, B$ **do**

2:     Generate $\hat{\boldsymbol{\theta}}_b^* \sim N_K\left(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}\right)$ and write $\hat{\boldsymbol{\theta}}_b^* = \left(\hat{\theta}_{b1}^*, \hat{\theta}_{b2}^*, \ldots, \hat{\theta}_{bK}^*\right)'$

3:     Compute

$$t_b^* = \max_{1 \le k \le K} \left|\frac{\hat{\theta}_{bk}^* - \hat{\theta}_k}{\sigma_k}\right|$$

4: **end for**

5: Compute the $(1 - \alpha)$-sample quantile of $t_1^*, t_2^*, \ldots, t_B^*$, call this $\hat{t}$.

6: The joint confidence region for $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_K)'$ is given by

$$\mathfrak{R}_1 = \left[\hat{\theta}_1 \pm \hat{t} \times \sigma_1\right] \times \left[\hat{\theta}_2 \pm \hat{t} \times \sigma_2\right] \times \cdots \times \left[\hat{\theta}_K \pm \hat{t} \times \sigma_K\right].$$

3.2 Proposed methodology to compute a joint confidence region for the ordered parameters

We shall also be proposing an approach to compute simultaneous confidence intervals for the ordered parameters. Suppose that for the parameters $\theta_1, \ldots, \theta_K$, the corresponding ordered values are $\theta_{(1)}, \ldots, \theta_{(K)}$. Similarly, let $\hat{\theta}_{(1)}, \ldots, \hat{\theta}_{(K)}$ be the ordered estimates. We now derive a joint confidence region for these ordered values, and set the region to be of the following form:

$$\mathfrak{R}_2 = [\hat{\theta}_{(1)} \pm t \times \sigma_{(1)}] \times [\hat{\theta}_{(2)} \pm t \times \sigma_{(2)}] \times \cdots \times [\hat{\theta}_{(K)} \pm t \times \sigma_{(K)}], \quad (3.4)$$

where $\sigma_{(k)} = \sqrt{V\left(\hat{\theta}_{(k)}\right)}$. The confidence region in (3.4) is computed such that the following condition holds:

$$P\left(\hat{\theta}_{(k)} - t \cdot \sigma_{(k)} \leq \theta_{(k)} \leq \hat{\theta}_{(k)} + t \cdot \sigma_{(k)}, \ \forall k = 1, 2, \ldots, K\right) = 1 - \alpha,$$
$$(3.5)$$

which is equivalent to

$$P\left(\max_{k=1,2,\ldots,K} \left|\frac{\hat{\theta}_{(k)} - \theta_{(k)}}{\sigma_{(k)}}\right| \leq t\right) = 1 - \alpha, \tag{3.6}$$

and we estimate t, which in (3.6) is the $(1 - \alpha)$-quantile of the distribution of $\max_k \left|\dfrac{\hat{\theta}_{(k)} - \theta_{(k)}}{\sigma_{(k)}}\right|$, via a suitable parametric bootstrap. Since $\sigma_{(k)}$ is not known, we estimate its value. We present two approaches to estimate $\sigma_{(k)}$. The first uses results from Chen (1976) and Dudewicz (1972) to obtain an expression of the asymptotic variance of $\hat{\theta}_{(k)}$ as follows:

$$V(\hat{\theta}_{(k)}) = \text{kth ordered value among } \left\{\theta_1^2 + \sigma_1^2, \theta_2^2 + \sigma_2^2, \ldots, \theta_K^2 + \sigma_K^2\right\} - \theta_{(k)}^2.$$
$$(3.7)$$

Consequently, we get an estimate of the asymptotic variance of $\hat{\theta}_{(k)}$, given by:

$$\widehat{V(\hat{\theta}_{(k)})} = \text{kth ordered value among } \left\{\hat{\theta}_1^2 + \sigma_1^2, \hat{\theta}_2^2 + \sigma_2^2, \ldots, \hat{\theta}_K^2 + \sigma_K^2\right\} - \hat{\theta}_{(k)}^2.$$
$$(3.8)$$

Algorithm 2 describes the procedure to compute the joint confidence region in (3.4) using this approach.

Algorithm 2 Computing the joint confidence region for the ordered parameters using the estimated asymptotic variance

1: **for** $b = 1, 2, \ldots, B$ **do**

2:     Generate $\hat{\boldsymbol{\theta}}_b^* = \left(\hat{\theta}_{b1}^*, \hat{\theta}_{b2}^*, \ldots, \hat{\theta}_{bK}^* \right)' \sim N_K \left(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma} \right)$ and let $\hat{\theta}_{b(1)}^*, \hat{\theta}_{b(2)}^*, \ldots, \hat{\theta}_{b(K)}^*$ be the corresponding ordered values

3:     Compute
$$\hat{\sigma}_{b(k)}^* = \sqrt{\left[\text{kth ordered value among } \left\{ \hat{\theta}_{b1}^{*2} + \sigma_1^2, \hat{\theta}_{b2}^{*2} + \sigma_2^2, \ldots, \hat{\theta}_{bK}^{*2} + \sigma_K^2 \right\} \right] - \hat{\theta}_{(k)}^{*2}}$$

4:     Compute $t_b^* = \max_{1 \leq k \leq K} \left| \dfrac{\hat{\theta}_{b(k)}^* - \hat{\theta}_k^*}{\hat{\sigma}_{b(k)}^*} \right|$

5: **end for**

6: Compute the $(1 - \alpha)$-sample quantile of $t_1^*, t_2^*, \ldots, t_B^*$, call this $\hat{t}$.

7: The joint confidence region of $\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(K)}$ is given by
$$\mathfrak{R}_2 = \left[\hat{\theta}_{(1)} \pm \hat{t} \times \hat{\sigma}_{(1)} \right] \times \left[\hat{\theta}_{(2)} \pm \hat{t} \times \hat{\sigma}_{(2)} \right] \times \cdots \times \left[\hat{\theta}_{(K)} \pm \hat{t} \times \hat{\sigma}_{(K)} \right]$$
where $\hat{\sigma}_{(k)}$ is computed as
$$\hat{\sigma}_{(k)} = \sqrt{\text{kth ordered value among } \left\{ \hat{\theta}_1^2 + \sigma_1^2, \hat{\theta}_2^2 + \sigma_2^2, \ldots, \hat{\theta}_K^2 + \sigma_K^2 \right\} - \hat{\theta}_{(k)}^2}$$

The expression for the asymptotic variance of $\hat{\theta}_{(k)}$ has actually been derived in a setting where the $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K$ are independent. It is not clear if the expression in (3.7) holds whenever $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K$ are correlated. Since this thesis allows for the $\hat{\theta}_k$s to be correlated, it is possible that the expression in (3.8) for the estimate of the variance of $\hat{\theta}_{(k)}$ lacks optimality. As an alternative to using the asymptotic variance, we shall also consider using the bootstrap to estimate the variance of $\hat{\theta}_{(k)}$. To this end, a second-level bootstrap can be employed to estimate the variance. Algorithm 3 illustrates the procedure to compute the joint confidence region for the ordered parameters with the use of the bootstrap to estimate the variance of the ordered estimates. As may be expected, this approach is computationally intensive.

Algorithm 3 Computing the joint confidence region for the ordered parameters using the bootstrap estimate of the variance

1: **for** $b = 1, 2, \ldots, B$ **do**

2:      Generate $\hat{\boldsymbol{\theta}}_b^* = \left(\hat{\theta}_{b1}^*, \hat{\theta}_{b2}^*, \ldots, \hat{\theta}_{bK}^* \right)' \sim N_K \left(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma} \right)$ and let $\hat{\theta}_{b(1)}^*, \hat{\theta}_{b(2)}^*, \ldots, \hat{\theta}_{b(K)}^*$ be the corresponding ordered values of $\hat{\theta}_{b1}^*, \hat{\theta}_{b2}^*, \ldots, \hat{\theta}_{bK}^*$

3:      **for** $c = 1, 2, \ldots, C$ **do**

4:          Generate $\hat{\boldsymbol{\theta}}_{bc}^{**} = \left(\hat{\theta}_{bc1}^{**}, \hat{\theta}_{bc2}^{**}, \ldots, \hat{\theta}_{bcK}^{**} \right) \sim N_K \left(\hat{\boldsymbol{\theta}}_b^*, \boldsymbol{\Sigma} \right)$ and let $\hat{\theta}_{bc(1)}^{**}, \hat{\theta}_{bc(2)}^{**}, \ldots, \hat{\theta}_{bc(K)}^{**}$ be the corresponding ordered values of $\hat{\theta}_{bc1}^{**}, \hat{\theta}_{bc2}^{**}, \ldots, \hat{\theta}_{bcK}^{**}$

5:          Compute $\hat{\sigma}_{b(k)}^* = \dfrac{\Sigma_{c=1}^{C} \left(\hat{\theta}_{bc(k)}^{**} - \bar{\hat{\theta}}_{b\cdot(k)}^{**} \right)^2}{C-1}, \quad \bar{\hat{\theta}}_{b\cdot(k)}^{**} = \dfrac{1}{C} \sum_{c=1}^{C} \hat{\theta}_{bc(k)}^{**}$

6:      **end for**

7:      Compute $t_b^* = \max\limits_{1 \leq k < K} \left| \dfrac{\hat{\theta}_{b(k)}^* - \hat{\theta}_{(k)}}{\hat{\sigma}_{b(k)}^*} \right|$

8: **end for**

9: Compute the $(1-\alpha)$-sample quantile of $t_1^*, t_2^*, \ldots, t_B^*$, call this $\hat{t}$.

10: The joint confidence region of $\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(K)}$ is

$$\mathfrak{R}_2 = \left[\hat{\theta}_{(1)} \pm \hat{t} \times \hat{\sigma}_{(1)} \right] \times \left[\hat{\theta}_{(2)} \pm \hat{t} \times \hat{\sigma}_{(2)} \right] \times \cdots \times \left[\hat{\theta}_{(K)} \pm \hat{t} \times \hat{\sigma}_{(K)} \right]$$

where $\hat{\sigma}_{(k)}$ is computed as

$$\hat{\sigma}_{(k)} = \frac{\Sigma_{b=1}^{B} \left(\hat{\theta}_{b(k)}^* - \bar{\hat{\theta}}_{\cdot(k)}^* \right)^2}{B-1}, \quad \bar{\hat{\theta}}_{\cdot(k)}^* = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}_{b(k)}^*$$

3.3 Correlation structures

The proposed methodologies assume that $V(\hat{\boldsymbol{\theta}}) = \boldsymbol{\Sigma}$ is known. Note that we can express $\boldsymbol{\Sigma}$ as in (3.9), where $\mathbf{R}$ is the population correlation matrix.

$$\boldsymbol{\Sigma} = \boldsymbol{\Delta}^{1/2} \mathbf{R} \boldsymbol{\Delta}^{1/2}; \quad \boldsymbol{\Delta} = \text{diag} \left\{ \sigma_1^2, \sigma_2^2, \ldots, \sigma_K^2 \right\} \tag{3.9}$$

The diagonal elements of $\mathbf{\Sigma}$, which are $\sigma_k^2 = V(\hat{\theta}_k)$ for $k = 1, 2, \ldots, K$, are treated as known quantities in practice. With large sample sizes, the variance estimates are stable enough that they are treated as the actual values. However, there is limited information to use as basis for the correlations among the stimates. In this thesis, we shall be assuming certain correlation structures among the $\hat{\theta}$s.

One structure that may be used is an equicorrelation matrix shown in (3.10). This assumes that the k variables are equally correlated, i.e., that $\rho_{jk} = \rho$ where $\rho \in [-1, 1]$ for $j \neq k \in \{1, \ldots, K\}$.

$$\mathbf{R}_{\text{eq}} = (1 - \rho)\, \mathbf{I}_K + \rho \mathbf{1}_K \mathbf{1}'_K = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}_{K \times K} \tag{3.10}$$

In a block correlation matrix $\mathbf{R}_{block}$ with G blocks, as represented by Archakova & Hansen (2020), the correlation between any two variables is determined by the block to which the two variables belong. Each diagonal block represents an equicorrelation structure within group g, denoted by

$$\mathbf{R}_{\text{eq,g}} = (1 - \rho_g)\, \mathbf{I}_{n_g} + \rho_g \mathbf{1}_{n_g} \mathbf{1}'_{n_g}$$

where ρ_g is the within-block correlation and n_g is the number of variables in block g such that $\sum_{g=1}^{G} n_g = K$. The off-diagonal blocks capture between-block correlations, represented by

$$\mathbf{C}_{g'g} = \mathbf{C}_{gg'} = \rho_{gg'} \mathbf{1}_{n_g} \mathbf{1}'_{n_g}$$
$$\text{where } g \neq g' \in \{1, \ldots, G\}$$

Thus, the full block correlation matrix can be expressed as in (3.11).

$$\mathbf{R}_{\text{block}} = \begin{bmatrix} \mathbf{R}_{eq,1} & \mathbf{C}_{12} & \cdots & \mathbf{C}_{1G} \\ \mathbf{C}_{21} & \mathbf{R}_{eq,2} & \cdots & \mathbf{C}_{2G} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}_{G1} & \mathbf{C}_{G2} & \cdots & \mathbf{R}_{eq,G} \end{bmatrix}_{K \times K} \tag{3.11}$$

In the context of pre-election surveys, each block may represent correlations induced by party or ticket membership, reflecting stronger associations within parties and weaker associations between them.

Correlation structures that account for spatial proximity can be borrowed from geostatistics. This is particularly relevant in light of Klein's observation that states located within certain regions exhibit similar travel time characteristics. In such cases, spatial dependence can be modeled using a stationary (i.e., no directional dependence) Matérn correlation function, which for two locations $\mathbf{s}_i$ and $\mathbf{s}_j$ is expressed as in (3.12).

$$\rho_{\text{matern}} = \frac{2^{1-\nu}}{\Gamma(\nu)}(\kappa \parallel \mathbf{s}_i - \mathbf{s}_j \parallel)^{\nu} K_{\nu}(\kappa \parallel \mathbf{s}_i - \mathbf{s}_j \parallel) \qquad (3.12)$$

where $\parallel \cdot \parallel$ denotes the Euclidean distance and K_{ν} is the second kind of the modified Bessel function. It has a scale parameter $\kappa > 0$ and a smoothness parameter $\nu > 0$. ρ_{matern} reduces to the exponential correlation when $\nu = 0.5$ and to Gaussian correlation function when $\nu = \infty$. In this paper, the R package "BayesNGSP" (Turek & Risser (2022)), is used to construct the $\mathbf{R}_{\text{matern}}$.

3.4 Evaluation

Algorithm 4 estimates the coverage probabilities associated with the proposed methodologies. The coverage probabilities correspond to the proportion of replications in which the true parameter values are contained within the confidence intervals for all K simultaneously. Moreover, the tightness of the joint confidence region that results from Algorithm 1 is assessed using three summary measures: the arithmetic mean (T_1), geometric mean (T_2), and the metric T_3 introduced by Wright (2025), as presented in Equations 3.13–3.15.

$$T_1 = \frac{1}{K} \sum_{k=1}^{K} \left| \Lambda_{Ok} \right| \qquad (3.13)$$

$$T_2 = \prod_{k=1}^{K} \left| \Lambda_{Ok} \right| \tag{3.14}$$

$$T_3 = 1 - \frac{OP}{K^2} \tag{3.15}$$

In equation 3.15, $OP = K + \Sigma_{k=1}^{K} \left| \Lambda_{Ok} \right|$ denotes the total number of occupied positions in a joint confidence region out of the total number of positions K^2; or the sum of the differences between the upper and lower bound of the simultaneous rank intervals added by 1, for each population k. Higher values of T_1 and T_2 indicate wider confidence intervals and are therefore less desirable, whereas higher values of T_3 are preferable. T_3 can range from 0, indicating no tightness, to $\frac{K-1}{K}$, implying the confidence region only contains the estimated ranking which is likely the true ranking.

Algorithm 4 Computing the coverage probability and tightness measures

---

For given values of $\boldsymbol{\Sigma}$ and $\theta_1, \theta_2, \ldots, \theta_K$ (with corresponding $\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(K)}$ for rank-based methods)

1: **for** replications $= 1, 2, \ldots, 5000$ **do**

2:     Generate $\hat{\boldsymbol{\theta}} \sim N_K(\boldsymbol{\theta}, \boldsymbol{\Sigma})$

3:     Compute the confidence region $\mathfrak{R}_1$ for the unordered parameters using Algorithm 1 and the confidence region for the ordered parameters $\mathfrak{R}_2$ using Algorithms 2 and 3.

4:     For the unordered parameters, check if $(\theta_1, \theta_2, \ldots, \theta_K) \in \mathfrak{R}_1$ and compute $T_1, T_2,$ and T_3. For the ordered parameters, check if $\left(\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(K)}\right) \in \mathfrak{R}_2$

5: **end for**

6: Compute the proportion of times that the condition in line 4 is satisfied and the average of $T_1, T_2,$ and T_3.

3.5 Simulation settings

For the simulation settings to be used in evaluating the performance of the proposed methodologies, we shall be varying the settings for the population mean $\boldsymbol{\theta}$, population variances $\sigma_1^2, \ldots, \sigma_K^2$, population correlation matrix $\mathbf{R}$, and number of populations being ranked K. In carrying out the simulations, a nominal level of $1 - \alpha = 0.95$ will be used. The table below summarizes the simulations settings.

$\boldsymbol{\theta}$	$\mathbf{R}$
K = 5	
$(22.5, 24.9, 27.5, 24.3, 24.8)'$	$\mathbf{R}_{\text{eq}}, \rho = 0.1, 0.5, 0.9$ $\mathbf{R}_{\text{block}}$
$(21.5, 25.7, 30.5, 24.7, 25.5)'$	$\mathbf{R}_{\text{eq}}, \rho = 0.1, 0.5, 0.9$ $\mathbf{R}_{\text{block}}$
$(19.9, 27, 34.9, 25.3, 26.7)'$	$\mathbf{R}_{\text{eq}}, \rho = 0.1, 0.5, 0.9$ $\mathbf{R}_{\text{block}}$
K = 10	
$(22.5, 24.9, 27.5, 24.3, 24.8, 23.2, 23, 21.6, 25.1, 19)'$	$\mathbf{R}_{\text{eq}}, \rho = 0.1, 0.5, 0.9$ $\mathbf{R}_{\text{block,10}}$
$(21.5, 25.7, 30.5, 24.7, 25.5, 22.6, 22.3, 19.9, 26.1, 15.2)'$	$\mathbf{R}_{\text{eq}}, \rho = 0.1, 0.5, 0.9$ $\mathbf{R}_{\text{block,10}}$
$(19.9, 27, 34.9, 25.3, 26.7, 21.9, 21.3, 17.3, 27.7, 9.4)'$	$\mathbf{R}_{\text{eq}}, \rho = 0.1, 0.5, 0.9$ $\mathbf{R}_{\text{block,10}}$
K = 20	
$(22.5, 24.9, 27.5, 24.3, 24.8, 23.2, 23, 21.6, 25.1, 19, 21.3,$ $24.1, 26.5, 26.9, 23.2, 25.8, 21.8, 21.6, 25.6, 21.2)'$	$\mathbf{R}_{\text{eq}}, \rho = 0.1, 0.5, 0.9$ $\mathbf{R}_{\text{block}}$
$(21.5, 25.7, 30.5, 24.7, 25.5, 22.6, 22.3, 19.9, 26.1, 15.2,$ $19.3, 24.3, 28.7, 29.3, 22.7, 27.5, 20.1, 19.9, 27.1, 19.1)'$	$\mathbf{R}_{\text{eq}}, \rho = 0.1, 0.5, 0.9$ $\mathbf{R}_{\text{block}}$
$(19.9, 27, 34.9, 25.3, 26.7, 21.9, 21.3, 17.3, 27.7, 9.4, 16.2,$ $24.6, 31.9, 33, 21.9, 29.9, 17.7, 17.3, 29.3, 16)'$	$\mathbf{R}_{\text{eq}}, \rho = 0.1, 0.5, 0.9$ $\mathbf{R}_{\text{block}}$
K = 30	
$(22.5, 24.9, 27.5, 24.3, 24.8, 23.2, 23, 21.6, 25.1, 19, 21.3,$ $24.1, 26.5, 26.9, 23.2, 25.8, 21.8, 21.6, 25.6, 21.2, 25.7,$ $22.2, 22.7, 22.6, 21.9, 24.2, 23.6, 22, 24.5, 25)'$	$\mathbf{R}_{\text{eq}}, \rho = 0.1, 0.5, 0.9$ $\mathbf{R}_{\text{block}}$
$(21.5, 25.7, 30.5, 24.7, 25.5, 22.6, 22.3, 19.9, 26.1, 15.2,$ $19.3, 24.3, 28.7, 29.3, 22.7, 27.5, 20.1, 19.9, 27.1, 19.1,$ $27.2, 20.9, 21.8, 21.7, 20.4, 24.5, 23.5, 20.5, 25.1, 26)'$	$\mathbf{R}_{\text{eq}}, \rho = 0.1, 0.5, 0.9$ $\mathbf{R}_{\text{block}}$

$\boldsymbol{\theta}$	$\mathbf{R}$
K = 30	
(19.9, 27, 34.9, 25.3, 26.7, 21.9, 21.3, 17.3, 27.7, 9.4, 16.2, 24.6, 31.9, 33, 21.9, 29.9, 17.7, 17.3, 29.3, 16, 29.4, 19, 20.4, 20.3, 18.2, 25, 23.3, 18.3, 26, 27.4)$'$	$\mathbf{R}_{\text{eq}}, \rho = 0.1, 0.5, 0.9$ $\mathbf{R}_{\text{block}}$
K = 40	
(22.5, 24.9, 27.5, 24.3, 24.8, 23.2, 23, 21.6, 25.1, 19, 21.3, 24.1, 26.5, 26.9, 23.2, 25.8, 21.8, 21.6, 25.6, 21.2, 25.7, 22.2, 22.7, 22.6, 21.9, 24.2, 23.6, 22, 24.5, 25, 26.5, 25.1, 25.2, 27.1, 25.3, 22.8, 20.6, 21.1, 26, 26)$'$	$\mathbf{R}_{\text{eq}}, \rho = 0.1, 0.5, 0.9$ $\mathbf{R}_{\text{block}}$
(21.5, 25.7, 30.5, 24.7, 25.5, 22.6, 22.3, 19.9, 26.1, 15.2, 19.3, 24.3, 28.7, 29.3, 22.7, 27.5, 20.1, 19.9, 27.1, 19.1, 27.2, 20.9, 21.8, 21.7, 20.4, 24.5, 23.5, 20.5, 25.1, 26, 28.6, 26.1, 26.4, 29.7, 26.5, 22.1, 18, 18.9, 27.8, 27.8)$'$	$\mathbf{R}_{\text{eq}}, \rho = 0.1, 0.5, 0.9$ $\mathbf{R}_{\text{block}}$
(19.9, 27, 34.9, 25.3, 26.7, 21.9, 21.3, 17.3, 27.7, 9.4, 16.2, 24.6, 31.9, 33, 21.9, 29.9, 17.7, 17.3, 29.3, 16, 29.4, 19, 20.4, 20.3, 18.2, 25, 23.3, 18.3, 26, 27.4, 31.9, 27.7, 28.1, 33.6, 28.2, 20.9, 14.1, 15.6, 30.5, 30.5)$'$	$\mathbf{R}_{\text{eq}}, \rho = 0.1, 0.5, 0.9$ $\mathbf{R}_{\text{block}}$
K = 50	
(22.5, 24.9, 27.5, 24.3, 24.8, 23.2, 23, 21.6, 25.1, 19, 21.3, 24.1, 26.5, 26.9, 23.2, 25.8, 21.8, 21.6, 25.6, 21.2, 25.7, 22.2, 22.7, 22.6, 21.9, 24.2, 23.6, 22, 24.5, 25, 26.5, 25.1, 25.2, 27.1, 25.3, 22.8, 20.6, 21.1, 26, 26, 22.1, 21.4, 23.2, 22.3, 22.4, 23.1, 21.4, 24.8, 27, 23.4)$'$	$\mathbf{R}_{\text{eq}}, \rho = 0.1, 0.5, 0.9$ $\mathbf{R}_{\text{block}}$
(21.5, 25.7, 30.5, 24.7, 25.5, 22.6, 22.3, 19.9, 26.1, 15.2, 19.3, 24.3, 28.7, 29.3, 22.7, 27.5, 20.1, 19.9, 27.1, 19.1, 27.2, 20.9, 21.8, 21.7, 20.4, 24.5, 23.5, 20.5, 25.1, 26, 28.6, 26.1, 26.4, 29.7, 26.5, 22.1, 18, 18.9, 27.8, 27.8, 20.8, 19.5, 22.8, 21.2, 21.3, 22.5, 19.5, 25.5, 29.6, 23)$'$	$\mathbf{R}_{\text{eq}}, \rho = 0.1, 0.5, 0.9$ $\mathbf{R}_{\text{block}}$
(19.9, 27, 34.9, 25.3, 26.7, 21.9, 21.3, 17.3, 27.7, 9.4, 16.2, 24.6, 31.9, 33, 21.9, 29.9, 17.7, 17.3, 29.3, 16, 29.4, 19, 20.4, 20.3, 18.2, 25, 23.3, 18.3, 26, 27.4, 31.9, 27.7, 28.1, 33.6, 28.2, 20.9, 14.1, 15.6, 30.5, 30.5, 18.8, 16.6, 22.1, 19.4, 19.6, 21.6, 16.7, 26.7, 33.5, 22.5)$'$	$\mathbf{R}_{\text{eq}}, \rho = 0.1, 0.5, 0.9$ $\mathbf{R}_{\text{block}}$

K	$\mathbf{R}_{\text{block,K}}$
5	
10	$\mathbf{R}_{\text{eq},n=3}$ with $\rho = 0.1$, $\mathbf{R}_{\text{eq},n=3}$ with $\rho = 0.5$, $\mathbf{R}_{\text{eq},n=4}$ with $\rho = 0.9$ with $\rho = 0.1$ in the off-diagonal blocks
20	
30	
40	
50	

Each k shares the same population variances $\sigma_1^2, \ldots, \sigma_K^2$:

K	σ
5	$(0.14, 0.33, 0.15, 0.23, 0.07)'$
10	$(0.14, 0.33, 0.15, 0.23, 0.07, 0.19, 0.19, 0.37, 0.32, 0.11)'$
20	$(0.14, 0.33, 0.15, 0.23, 0.07, 0.19, 0.19, 0.37, 0.32, 0.11, 0.17, 0.27, 0.24,$ $0.11, 0.11, 0.13, 0.16, 0.15, 0.15, 0.25)'$
30	$(0.14, 0.33, 0.15, 0.23, 0.07, 0.19, 0.19, 0.37, 0.32, 0.11, 0.17, 0.27, 0.24,$ $0.11, 0.11, 0.13, 0.16, 0.15, 0.15, 0.25, 0.15, 0.13, 0.1, 0.1, 0.24, 0.13, 0.32,$ $0.19, 0.27, 0.3)'$
40	$(0.14, 0.33, 0.15, 0.23, 0.07, 0.19, 0.19, 0.37, 0.32, 0.11, 0.17, 0.27, 0.24,$ $0.11, 0.11, 0.13, 0.16, 0.15, 0.15, 0.25, 0.15, 0.13, 0.1, 0.1, 0.24, 0.13, 0.32,$ $0.19, 0.27, 0.3, 0.12, 0.27, 0.09, 0.12, 0.36, 0.09, 0.15, 0.16, 0.09, 0.29)'$
50	$(0.14, 0.33, 0.15, 0.23, 0.07, 0.19, 0.19, 0.37, 0.32, 0.11, 0.17, 0.27, 0.24,$ $0.11, 0.11, 0.13, 0.16, 0.15, 0.15, 0.25, 0.15, 0.13, 0.1, 0.1, 0.24, 0.13, 0.32,$ $0.19, 0.27, 0.3, 0.12, 0.27, 0.09, 0.12, 0.36, 0.09, 0.15, 0.16, 0.09, 0.29,$ $0.16, 0.28, 0.14, 0.07, 0.2, 0.31, 0.13, 0.14, 0.31, 0.11)'$

Bibliography

Andersson, J., Carling, K., & Mattson, S. (1998). Random ranking of hospitals is unsound. *CHANCE*, *11*(3), 33–39. https://doi.org/doi: 10.1080/09332480.1998.10542106

Archakova, I., & Hansen, P. R. (2020). *A canonical representation of block matrices with applications to covariance and correlation matrices.*

Bazylik, S., Mogstad, M., Romano, J. P., Shaikh, A. M., & Wilhelm, D. (2025). Simultaneous confidence regions for ranks. *Journal of Econometrics.* https://doi.org/https://doi.org/10.1016/j.jeconom.2025.106010

Chen, H. J. (1976). Strong consistency and asymptotic unbiasedness of a natural estimator for a ranked parameter. *Sankhyā: The Indian Journal of Statistics*, *38*(1), 92–94.

David, C., & Legara, E. F. (2015). *How voters combine candidates on the ballot: The case of the philippine senatorial elections.*

Dudewicz, E. J. (1972). Point estimation of ordered parameters: The general location parameter case. *Tamkang Journal of Mathematics*, *3*(2), 101–114.

Dunn, O. J. (1958). *Estimation of the means of dependent variables.*

Goldstein, H., & Healy, M. J. R. (1995). The graphical presentation of a collection of means. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *158*(1), 175–177.

Goldstein, H., & Spiegelhalter, D. J. (1996). League tables and their limitations: Statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *159*(3), 385–443.

Hall, P., & Miller, H. (2009). Using the bootstrap to quantify the authority

of an empirical ranking. *The Annals of Statistics*, *37*(6B), 3929–3959.

Klein, M., Wright, T., & Wieczorek, J. (2020). A joint confidence region for an overall ranking of populations. *Journal of the Royal Statistical Society*, 589–606.

Krainski, E. T., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., & Rue, H. (2019). *Advanced spatial modeling with stochastic partial differential equations using r and INLA*. Chapman & Hall/CRC Press.

Lyhagen, J., & Ahlgren, P. (2020). Uncertainty and the ranking of economics journals. *Scientometrics*, *125*, 2545–2560. https://doi.org/https://doi.org/10.1007/s11192-020-03681-5

Mogstad, M., Romano, J. P., Shaikh, A. M., & Wilhelm, D. (2024). Inference for ranks with applications to mobility across neighbourhoods and academic achievement across countries. *The Review of Economic Studies*, *91*(1), 476–518.

Mohamad, D. A., Zwet, E. W. van, & Goeman, J. J. (2019). Simultaneous confidence intervals for ranks with application to ranking institutions. *Journal of the International Biometric Society*.

Ravanilla, N., & Hicken, A. (2023). *When legislators don't bring home the pork: The case of philippine senators*.

Šidák, Z. (1967). *Rectangular confidence regions for the means of multivariate normal distributions*.

Turek, D., & Risser, M. (2022). *bayesNSGP: Bayesian analysis of nonstationary gaussian process models*. https://cran.stat.auckland.ac.nz/web/packages/BayesNSGP/BayesNSGP.pdf

Wright, T. (2025). *Optimal tightening of the KWW joint confidence region for a ranking*.

Zhang, S., Luo, J., Zhu, L., Stinchcomb, D. G., Campbell, D., Carter, G., Gilkesone, S., & Feuerc, E. J. (2013). *Confidence intervals for ranks of age-adjusted rates across states or counties*.