TITLE HERE

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
Month Year

Contents

1 Introduction

Ranks are typically based on estimates obtained along with their measure of uncertainty, expressed as confidence intervals. Individually, these intervals provide information on the possible range of each rank. However, in most applications, the primary interest lies in comparing all ranks simultaneously rather than reporting them in isolation.

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:

- Construct joint confidence intervals that utilize parametric bootstrap to obtain a tighter overall uncertainty for ranks.
- Establish joint confidence intervals for cases when ranks are assumed to be correlated.
- Evaluate the performance of the proposed approaches under different standard deviations, correlation structures, and dimensionalities.

1.2 Significance

In order to obtain joint confidence sets for overall ranks, the work of Klein et al. (2020) requires estimating 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. In the same paper, they considered the set of familiar $\hat{\theta} \pm z_{\alpha/2} + SE_k$ individual confidence intervals, assuming an independently distributed $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K$. This approach, while simple, disregards the idea that θ_ks may be correlated. In some cases, assuming independence in the presence of dependence leads to conservative confidence intervals resulting in wider intervals which imply a higher uncertainty in overall ranks.

For instance, in the case of ranking senatorial candidates in the Philippines, this assumption 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 candidate with 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, although not equally well. For example, in the 2025 election, several candidates from both the Marcos (Alyansa para sa Bagong Pilipinas) and Duterte (DuterTen) blocs secured seats, while others from the same groups did not. Accounting for these patterns allows for a more realistic assessment of uncertainty in the estimated ranks for similar use cases.

Klein et al. (2020) also noted that although it is difficult to make generalizations 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, the 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.

1.3 Scope and Limitations

This study focuses on presenting alternative ways to construct joint confidence regions for quantifying uncertainty of overall ranks using the main result from Klein et al. (2020). It covers the application of parametric bootstrap and consideration of potential correlation among ranks while maintaining tightness in the resulting overall uncertainty. However, certain limitations must be acknowledged. First, a constraint is introduced by assuming that the data is generated from the normal distribution. Second, a number of correlation structures are examined to demonstrate how different dependence assumptions among candidates may influence the resulting joint confidence sets. However, identifying which structure best captures the actual voting behavior in the Philippine senatorial context is beyond the scope of this study. In line with this, the third limitation is that the correlation structures are just assumed forms and not estimated from the data. Overall, these limitations suggest that the findings should be viewed mainly as methodological examples.

2 Related Literature

2.1 Joint confidence region for an overall ranking

Klein et al. (2020) proposed an approach for quantifying overall rank uncertainty following the estimation of respondents' average travel time to work in each K sampled geographical area. In their paper, rank for the kth population is defined as

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

Meanwhile, the estimated overall ranking is computed from the estimates $\hat{\theta}_1, \ldots, \hat{\theta}_K$, and expressed as $(\hat{r}_1, \ldots, \hat{r}_K)$, where

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

The true values, $\theta_1, \ldots, \theta_K$ are unknown. For this, they assumed that for each $k \in \{1, 2, \ldots, K\}$, there exists L_k and U_k such that

$$\theta_k \in (L_k, U_k) \tag{2.3}$$

and defined the following:

$$\left.\begin{aligned} I_k &= \{1, 2, \ldots, K\} - \{k\}, \\ \Lambda_{Lk} &= \{j \in I_k : U_j \le L_k\}, \\ \Lambda_{Rk} &= \{j \in I_k : U_k \le L_j\}, \\ \Lambda_{Ok} &= \{j \in I_k : U_j > L_k \text{ and } U_k > L_j\} = I_k - \{\Lambda_{Lk} \cup \Lambda_{Rk}\} \end{aligned}\right\} \tag{2.4}$$

Equation 2.4 can likewise be expressed in words as follows:

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 $\Lambda_{Lk} \cup \Lambda_{Rk} \cup \Lambda_{Ok} = I_k$

   The above implies that for each $k \in \{1, 2, \ldots, K\}$,

$$r_k \in \{|\Lambda_{Lk}| + 1, |\Lambda_{Lk}| + 2, |\Lambda_{Lk}| + 3, \ldots, |\Lambda_{Lk}| + |\Lambda_{Ok}| + 1\} \tag{2.5}$$

Equation 2.5 demonstrates that a smaller $|\Lambda_{Ok}|$ results in smaller difference between U_k and L_k. Collectively, for all k, this yields narrower confidence intervals for the overall ranks, which is desirable.

They also assumed a conservative confidence region whose joint coverage probability is at least as large as the nominal level, $1 - \alpha$, as shown in Equation 2.6.

$$P\left[\bigcap_{k=1}^{K} \{\theta_k \in (L_k, U_k)\}\right] \ge 1 - \alpha \tag{2.6}$$

This yields the joint confidence set for the overall ranking, as defined in Equation 2.7, which they showed to also have a joint probability of at least $1 - \alpha$.

$$\{(r_1,\ldots,r_K): r_k \in \{|\Lambda_{Lk}|+1, |\Lambda_{Lk}|+2, |\Lambda_{Lk}|+3,\ldots,|\Lambda_{Lk}|+| \text{ for } k=1,2,\ldots,K\}\} \tag{2.7}$$

In line with this, they presented a proof demonstrating that if $(L_1,U_1),\ldots,(L_K,U_K)$ are constructed such that the estimator $\hat{\theta}_k \in (L_k,U_k)\ \forall k \in \{1,2,\ldots,K\}$, then the estimated ranking $(\hat{r}_1,\hat{r}_2,\ldots,\hat{r}_K)$ lies within the joint confidence region defined in Equation 2.7 with probability 1.

They also noted that the joint confidence region in 2.7 contains more than one possible overall ranking unless the values of θ_k differ from each other such that $(L_k,U_k)\cap(L_{k'},U_{k'}) = \emptyset,\ \forall\ k \neq k'$. This implies that the unique overall ranking arises only from the narrowest attainable joint confidence region and it is the estimated ranking, $(\hat{r}_1,\hat{r}_2,\ldots,\hat{r}_K)$.

2.2 Joint confidence intervals for θ_ks

Klein et al. (2020) used individual confidence intervals of the form $\hat{\theta}_k \pm z_{\alpha/2}SE_k^2$, with $\hat{\theta}_k \sim N(\theta_k, SE_k)$ for $k \in \{1,2,\ldots,K\}$, where $\theta_1,\theta_2,\ldots\theta_k$ are unknown and $SE_1,SE_2,\ldots,SE_k$ are known. It was noted that $MOE_k = z_{\alpha/2} \times SE_k$ where $SE_k = \frac{\sigma_k}{\sqrt{n}}$.

The first one can be traced from Theorem 1 of Šidák (1967) which states that for a vector of random variables of dimension K, $\mathbf{X} = (X_1,X_2,\ldots,X_K)$, with $\mathbf{X} \sim N_K(\mathbf{0},\Sigma)$ and having an arbitrary correlation matrix $\mathbf{R} = \{\rho_{kk'}\}_{k,k'=1}^{K}$,

$$\begin{aligned} P(|X_1| \le c_1,\ \ldots,|X_K| \le c_K) \ge \\ & P(|X_1| \le c_1) \times P(|X_2| \le c_2,\ldots,|X_K| \le c_K), \\ & \text{for any positive numbers } c_1,c_2,\ldots,c_K \end{aligned} \tag{2.8}$$

He showed by induction that under the assumptions of Theorem 1,

$$P(|X_1| \le c_1,\ \ldots,|X_K| \le c_K) \ge \prod_{k=1}^{K} P\left(|X_k| \le c_k\right) \tag{2.9}$$

In words, this means that the smallest confidence level that can be attained will always be $1-\alpha$ and that in cases of presence of dependence when independence is assumed, coverage will always be more than $1-\alpha$.

For the simultaneous confidence intervals used by Klein, Šidák (1967) considered a random sample of n vectors of $\mathbf{Y}_i = (Y_{i1},Y_{i2},\ldots,Y_{iK})'$, $i=1,\ldots,n$ where $Y_{ik} \sim N(\mu_k,\sigma_k^2)$ with unknown μ_k and known σ_k^2 and stated that

$$X_k = \frac{\left(\hat{\theta}_k - \mu_k\right)}{\sigma_k \big/ \sqrt{n}} \sim N(0,1), \quad k=1,\ldots,K \tag{2.10}$$

where

$$\hat{\theta}_k = \bar{Y}_k = n^{-1} \sum_{i=1}^{n} Y_{ik} \tag{2.11}$$

satisfies the requirements of Theorem 1. Hence, when constructing a simultaneous confidence interval for $\theta_k = \mu_k,\ \forall k \in \{1, 2, \ldots, K\}$ with $(1-\alpha)$ confidence level, it follows from 2.9 and 2.10 that,

$$\begin{aligned}\prod_{k=1}^{K} P\left(|X_k| \leq c_k\right) &= \prod_{k=1}^{K} P\left(\hat{\theta}_k - c_k \cdot \frac{\sigma}{\sqrt{n}} \leq \theta_k \leq \hat{\theta}_k + c_k \cdot \frac{\sigma}{\sqrt{n}}\right) \\ &= \prod_{k=1}^{K} P\left(\hat{\theta}_k - c_k \cdot SE_k \leq \theta_k \leq \hat{\theta}_k + c_k \cdot SE_k\right) \\ &= 1 - \alpha\end{aligned} \tag{2.12}$$

As a result, this will always yield a confidence level for $\left(\hat{\theta}_k - c_k \cdot SE_k, \hat{\theta}_k + c_k \cdot SE_k\right)$ that is least as large as $1-\alpha$ - being equal when independence holds and larger than $1-\alpha$ when dependence is actually present.

For the choice of c_k, Šidák advised to assume independence with $c_1 = \cdots = c_K = c_\gamma$ where γ is the individual significance level so that

$$\prod_{k=1}^{K} P\left(|X_k| \leq c_k\right) = \prod_{k=1}^{K} (1-\gamma) = (1-\gamma)^K = 1 - \alpha$$

and deriving γ returns $1 - (1-\alpha)^{1/K}$. Under this condition, the two-sided $100(1-\alpha)\%$ confidence interval for the parameter $\theta_k = \mu_k$ is simultaneously given for each $k \in \{1, \ldots, K\}$ by

$$I_{k(ind)} = \left(\hat{\theta}_k - z_{\gamma/2} SE_k,\ \hat{\theta}_k + z_{\gamma/2} SE_k\right), \qquad \text{for } k \in \{1, 2, \ldots, K\} \tag{2.13}$$

where

$$z_{\gamma/2} = \Phi^{-1}\left(1 - \frac{\gamma}{2}\right) = \Phi^{-1}\left(1 - \frac{1-(1-\alpha)^{1/K}}{2}\right) \tag{2.14}$$

Šidák also suggested the use of Bonferroni inqeuality for the case when variances are unknown and unequal.This was demonstrated by Dunn (1958) as follows:

$$P(|X_1| \leq c_1,\ \ldots, |X_K| \leq c_K) \geq 1 - 2K\left[1 - \Phi(c_\alpha)\right] = 1 - \alpha \tag{2.15}$$

where solving for $c_\alpha = z_{\frac{\alpha}{2K}}$ gives $\Phi^{-1}\left(1 - \frac{\alpha}{2K}\right)$ resulting in a conservative joint coverage for $\theta_1, \theta_2, \ldots, \theta_K$ of at least $1-\alpha$. The corresponding two-sided $100(1-\alpha)\%$ confidence intervals are as defined in 2.16.

$$I_{k(bonf)} = \left(\hat{\theta}_k - z_{(\alpha/K)/2}SE_k,\ \hat{\theta}_k + z_{(\alpha/K)/2}SE_k\right), \qquad \text{for } k = 1, 2, \ldots, K \tag{2.16}$$

Other studies with similar concern include that of Mohamad et al. (2019) where using the same assumptions, Tukey's pairwise comparison procedure is used to come up with their $(1-\alpha)$ joint confidence intervals for ranks defined in Equation 2.17. They showed this to yield uniformly narrower intervals than that of Klein's approach, for the case when SEs are equal.

$$\left(1 + \#\left\{j : y_i - y_j - q_{1-\alpha}\sqrt{SE_i^2 + \sigma_j^2} > 0\right\},\ n - \#\left\{j : y_i - y_j + q_{1-\alpha}\sqrt{SE_i^2 + \sigma_j^2} < 0\right\}\right), \tag{2.17}$$

where $\gamma = 1 - (1-\alpha)^{\frac{1}{K}}$.

Zhang et al. (2013) analyzed 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 used age-adjusted rates to minimize the confounding effect of age differences when comparing different population groups. They also extended their method to handle cases where only the rates and confidence intervals are available, aligning it more closely with the approaches of Klein et al. (2020) and Mohamad et al. (2019).

3 Methodology

This section introduces the proposed methodologies to obtain confidence regions for the unknown overall true ranking. It adds to the list of approaches from Klein et al. (2020) which uses the Bonferroni correction and the independence assumption, by addressing the case when items ranked are assumed correlated to certain degrees. Section 3.1 lists the algorithms employed to compute the joint confidence regions. This includes a non-rank and rank-based methods. It also has a subsection that discusses correlation structures suitable to the intended use cases. The calculated joint confidence regions are then assessed on the basis of coverage and metrics that measure the tightness of estimated confidence regions. These are tackled in section 3.2.

3.1 Parametric bootstrap approaches for constructing joint confidence intervals for correlated $\theta_1, \ldots, \theta_K$

The proposed approaches follow those of Klein et al. (2020) and Mohamad et al. (2019), which do not require knowledge of the sampling design and estimation methodology for each population. We primarily use parametric bootstrap to approximate the quantile used to calculate the confidence region. This is constructed to account for assumed correlation among items being ranked. In line with this, various correlation structures $\boldsymbol{\rho}$ are listed in section 3.1.3, to be later examined in a simulation study. The correlation matrix is used in the calculation of the covariance matrix as show in Equation 3.1.).

$$\boldsymbol{\Sigma} = \boldsymbol{\Delta}^{1/2}\boldsymbol{\rho}\boldsymbol{\Delta}^{1/2}; \quad \boldsymbol{\Delta} = \operatorname{diag}\left\{\sigma_1^2, \sigma_2^2, \ldots, \sigma_K^2\right\} \tag{3.1}$$

with known σ_k's and $\boldsymbol{\rho}$. This form of $\boldsymbol{\Sigma}$ will be used in sections 3.1.1 and 3.1.2.

3.1.1 Nonrank-based method

Algorithm 1 Computation of Joint Confidence Region

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 of $\theta_1, \theta_2, \ldots, \theta_K$ is given by

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

3.1.2 Rank-based methods

Algorithm 2 Computation of Joint Confidence Region

Let the data consist of $\hat{\boldsymbol{\theta}} = \left(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K\right)$ and suppose $\boldsymbol{\Sigma}$

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

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

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

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

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

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

3.1.2.1 Asymptotic variance

Algorithm 3 Computation of Joint Confidence Region

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

  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 $\hat{\theta}_{b1}^*, \hat{\theta}_{b2}^*, \ldots, \hat{\theta}_{bK}^*$

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

    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 $\hat{\theta}_{b1}^*, \hat{\theta}_{b2}^*, \ldots, \hat{\theta}_{bK}^*$

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

  end for

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

end for

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

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

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

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

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

3.1.2.2 Variance from second-level bootstrap

3.1.3 Correlation structures

This section discusses the correlation structures considered in the simulation. For simplicity, an equicorrelation matrix is included. This assumes that the k variables are equally correlated, such that $\rho_{kk'} = r$ where $r \in [-1, 1]$. In matrix form,

$$\mathbf{R}_{eq} = (1-r)\,\mathbf{I}_K + r\mathbf{1}_K\mathbf{1}_K' = \begin{bmatrix} 1 & r & \cdots & r \\ r & 1 & \cdots & r \\ \vdots & \vdots & \ddots & \vdots \\ r & r & \cdots & 1 \end{bmatrix}_{K\times K} \tag{3.2}$$

In a block correlation matrix, the correlation between any two variables is solely determined by the block to which the two variables belong. This results in a matrix with a common correlation coefficient within each block. It can be thought of as

$$\mathbf{R}_{block} = \begin{bmatrix} \mathbf{R}_{eq,\rho_{11}} & \rho_{12}\mathbf{1}_{n_1\times n_2} & \cdots & \rho_{1G}\mathbf{1}_{n_1\times n_G} \\ \rho_{12}\mathbf{1}_{n_1\times n_2} & \mathbf{R}_{eq,\rho_{22}} & \cdots & \rho_{2G}\mathbf{1}_{n_2\times n_G} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1G}\mathbf{1}_{n_1\times n_G} & \rho_{2G}\mathbf{1}_{n_2\times n_G} & \cdots & \mathbf{R}_{eq,\rho_{GG}} \end{bmatrix}_{K\times K} \tag{3.3}$$

3.2 Evaluation

Algorithm 5 is employed to estimate the coverage, which corresponds to the proportion of replications in which the true parameter values are contained within the confidence intervals for all K simultaneously. Likewise,the tightness of the joint confidence region is 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 3.4–3.6.

Algorithm 4 is similar to Algorithm 5 but computes for the coverage and average T_1, T_2, and T_3 for the nonrank-based method.

Algorithm 4 Computation of Coverage Probability for Nonrank-based Method

For given values of $\theta_1, \theta_2, \ldots, \theta_K$ and $\boldsymbol{\Sigma}$

1: **for** replications $= 1, 2, \ldots, 5000$ **do**
2: Generate $\hat{\boldsymbol{\theta}} \sim N_K(\boldsymbol{\theta}, \boldsymbol{\Sigma})$
3: Compute the rectangular confidence region $\mathfrak{R}$ using Algorithm 1.
4: Check if $(\theta_1, \theta_2, \ldots, \theta_K) \in \mathfrak{R}$ and compute T_1, T_2, and T_3.
5: **end for**
6: Compute the proportion of times that the condition in step 4 is satisfied and the average of T_1, T_2, and T_3.

$$T_1 = \frac{1}{K} \sum_{k=1}^{K} \left| \Lambda_{Ok} \right| \tag{3.4}$$

$$T_2 = \prod_{k=1}^{K} \left| \Lambda_{Ok} \right| \tag{3.5}$$

$$T_3 = 1 - \frac{K + \sum_{k=1}^{K} \left| \Lambda_{Ok} \right|}{K^2} \tag{3.6}$$

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.

Algorithm 5 Computation of Coverage Probability for Parametric Bootstrap

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

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

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

3: Compute the rectangular confidence region $\mathfrak{R}$ using Algorithm 2 and 3.

4: Check if $\left(\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(K)}\right) \in \mathfrak{R}$ and compute T_1, T_2, and T_3

5: **end for**

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

Bibliography

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

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

Klein, M., Wright, T., & Wieczorek, J. (2020). *A joint confidence region for an overall ranking of populations.*

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

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.*

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.*