Master Thesis

Grouped Lasso Regression in Multiple System Estimation

Examining the Effect of Sample Population and Covariates

Author: Jiajian Yan

Student number: 6763294

Project supervisor: Maarten Cruyff

Second Examiner: Peter van der Heijden

Word count: 4018

Applied Data Science

Utrecht University

Date: 01/07/2024

Abstract

Multiple System Estimation (MSE) is a crucial statistical method used for population estimation in fields like human rights, ecology, and epidemiology, particularly when complete population counts are unfeasible. Traditional methods, such as log-linear analysis, often face computational challenges due to the exponential increase in model possibilities with additional registers and covariates. To address this, grouped lasso as a regularization technique has been explored, aiming at streamlining model selection by penalizing less impactful coefficients. This study evaluated various grouped lasso models through simulations, adjusting parameters like the number of registers, covariates, and population samples. Three contrast methods (treatment, sum, and polynomial) were used and evaluated against traditional log-linear regression that used the AIC and BIC criteria. Results indicated that when using the optimal lambda values, grouped lasso methods consistently achieved low medians relative to the true population, with narrow interquartile ranges and minimal outliers, demonstrating high precision but low accuracy. AIC/BIC-based model selection showed high variation and outliers, however with significantly higher precision once outliers have been removed. Results suggest for further possibilities in the exploration of grouped lasso in more simulated datasets of different parameter combinations, as well as the use of other regularization based methods such as ridge and elastic net. This thesis project is completed in collaboration with group member Yanwen Zhang (Student Number 9087605), who investigated the effect of higher sample population percentage and higher number of registers, whilst a common baseline dataset has been used between the two theses.

Contents

1 Introduction

Multiple System Estimation (MSE) is a popular statistical methodology, widely employed for population estimation, based on several incomplete data registries. This technique is commonly utilized across a range of disciplines, including social sciences, ecology, and epidemiology, particularly in contexts where precise population counts are challenging to obtain. Example of such scenarios include, in humanitarian efforts, MSE is instrumental in estimating the numbers of displaced individuals, including refugees and the homeless, thereby enhancing the efficacy of relief and human rights initiatives (Cruyff et al. 2021) . In public health, MSE serves a crucial role by measuring the speed of contagious diffusion of infectious diseases, aiding the timely formulation of intervention strategies (Hald 2005). Additionally, MSE can be supportive in wildlife conservation efforts (King and Brooks 2008), where it is often used to estimate populations of endangered species, thus informing and improving conservation policies and practices.

One of the most common methodologies employed in MSE include the log-linear analysis of a contingency table, formulated through population registers and the associated covariates. The number of potential models associated with the increases in the number of registers, covariates, and interaction terms increases exponentially. This causes the approach of going through every potential model to be inefficient and almost impossible in terms of both time and computational resources (Binette and Steorts 2022). Model selection in MSE is further complicated by the requirement on the efficient handling of complex and potentially sparse data. Instead of going through every possible unique combinations, two predominant strategies have been utilized to identify the best performing model: forward selection and backward selection. The aim of the two selection techniques is to obtain a smaller proposal set of models, where forward selection progressively adds interaction terms to the simplest model, continuing until no improvement is observed, whilst backward selection removes terms from the most complex model until no further improvement is possible.

Despite the advantages offered by forward and backward selection methods over a full search, these techniques remain resource-intensive, requiring substantial computational time, especially when dealing with multiple registers and covariates. On the other hand, Lasso regression has been a regularization technique that has commonly been used to avoid overfitting and performing efficient variable selection, where both features can be highly beneficial in the context of MSE. Lasso regression applies a penalty to the coefficients of the regression variables, shrinking those with minimal impact on the model's performance relative to the cost of their inclusion. This penalization strategy can significantly streamline the model selection process by reducing the need for sequential, step-by-step evaluation of numerous potential models, thus has the potential of offering a more computationally efficient solution (Tibshirani 1996).

In addition to the standard lasso algorithm, grouped lasso extends the use of penalization onto specific divisions of coefficients, where every coefficient within a group will be evaluated and punished equally. Grouped lasso offers significant advantages when there are natural groupings among predictors, enhances model interpretability, whilst keeping the hierarchical structures of covariates, which can be problematic for standard lasso models. Two grouping methods will be particularly explored, namely the factor grouping method and the interaction grouping method. Factor grouping keeps all levels of a covariate in an interaction term with a unique variable in the same groups for each degree interaction, whilst the interaction grouping method divides the entire degree of interaction terms into its own group.

The introduction of regularization techniques, particularly the use of the Lasso and grouped Lasso approaches, represents a potential for adaptation to the MSE framework. Most notably of which is the logilasso package developed by Dahinden (2007), originally designed to address the challenges encountered in computational biology when analyzing sparse contingency tables and complex interactions among categorical variables, such as those found in full-length cDNA libraries of alternatively spliced genes. This paper uniquely applies Dahinden's $l1$ -penalization approach, extending it to the context of MSE, aiming to provide more efficient and highly effective model selection and parameter estimation.

In order to establish the efficiency of lasso regression in MSE, different simulation data will be formulated, using the Lasso package's simdat2 function developed by Cruyff (2020). Simulated data will be formulated 10 times, separately with a 10% and 25% in-sample population rate. Additionally, to test the efficacy of grouped lasso regression models, simulation trials with higher number of covariates, increasing 2 covariates to 3 covariates, will be performed while other parameters stay the same as the reference dataset. Through adjusting the mean parameter, the generated data matrix can have a different percentage of population that are recorded by the registers.

This thesis paper aims to answer the following research question:

How effective is grouped lasso regression as a model selection tool for Multiple System Estimations (MSE), and how does it perform relative to existing methods?

The remaining will be structured as follows. Section 2 explains all methods used in details, such as lasso/grouped lasso, step-wise selection and log-linear algorithm. Sector 3 will be the application of simulation and sector 4 contains the results obtained. The last section will be the discussion.

2 Methodology

2.1 Log-linear Models

Log-linear models have commonly been adopted in the field of Multiple System Estimation, with regards to analysing the relationship between categorical variables, as well as their interaction terms, through the use of contingency tables. This is based on a fundamental assumption where the logarithm of expected frequencies can be formulated into parameters bounded by a linear function.

Through formulation a two by two contingency table, two registers A and B represent the population registers that are incomplete. Such a model can be expressed using the following formula:

$$\log(\mu_{ij}) = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB},$$

where:

- $\log(\mu_{ij})$ is the expected frequency given register A at the i-th level and register B at the j-th level, for i, j belongs to 0,1.
- λ_0 is the intercept.
- λ_i^A and λ_j^B are the main effects for registers A and B.
- λ_{ij}^{AB} is the interaction term between the two registers. Note that the highest order interaction effect cannot be estimated in the cases of MSE and it will be set to 0 for all simulated data..

The log-linear models are key in analysing data with multiple covariates and interaction terms, which are commonly found in the data registers related to MSE.

2.2 Step-wise Selection

As previously noted, model selection is a common pitfall to the use of numerous complex registers with high number of covariates and possibly high number of interaction terms, as significantly more potential models have to be tested to find the optimal solution. Model selection has commonly been performed based on the principle of step-wise selection, by adding or removing a predictor at every step, based on the improvement or the loss of statistical significance to the model fit, where the most common metrics used are the Akaike information criterion (AIC) and the Baysesian information criterion (BIC). Both information criteria have been adopted in step-wise selection for measuring the increase or decrease of information that is brought by additional or removal of a predictor. Both AIC and BIC can be beneficial in finding the models with the least number of predictors, where lower AIC or BIC values represent a better performing model. BIC adds a slight twist by introducing a large penalty on using higher number of predictors. Overall, the performance of models selected through AIC or BIC have no intrinsic advantage over one another, thus a comparison and assessment is necessary for every model in every case of simulated data.

Two of the main types of step-wise selection methods include forward selection and backward selection. Whilst the two method share a common principle, their order of selection is opposite and their efficiency can be different. In forward selection, a model begins with no predictors in the model, then the model is saturated by adding an additional predictor that offers the greatest improvement in every step, until when no improvement over the set threshold can be observed. On the other hand, backward selection deploys a complete model with all potential predictors, then removing the predictor with the least importance in every step, and the selection stops when all remaining predictors have a significance higher than the set threshold. Backward selection will be the choice of method in this paper, due to some restrictions of forward selection, including the dependence of the order of predictors, as well as being prone to overfitting. Whilst both selection methods should be tested against in an ideal setting, this paper is bounded by time restrictions and thus only backward selection has been performed.

2.3 Lasso Regression & Grouped Lasso

2.3.1 Lasso

Prior to understanding how lasso could work in the setting of MSE, it is important to note the expected value of the observable counts in a contingency table can be expressed by the following:

$$\mu_i = e^{\beta_0 + \sum_j x_{ij}\beta_j}$$

where:

- n represents the number of observable cells.

- p represents the number of parameters.
- x_{ij} is the vector of predictors.
- β_0 is the intercept.
- $\boldsymbol{\beta}$ is the term in the parameter vector, which can be obtained by maximizing the following log-likelihood function:

$$\log \ell(\beta) = \sum_i \left(y_i(\beta_0 + x_{ij}\beta_j) - e^{\beta_0 + x_{ij}\beta_j} \right)$$

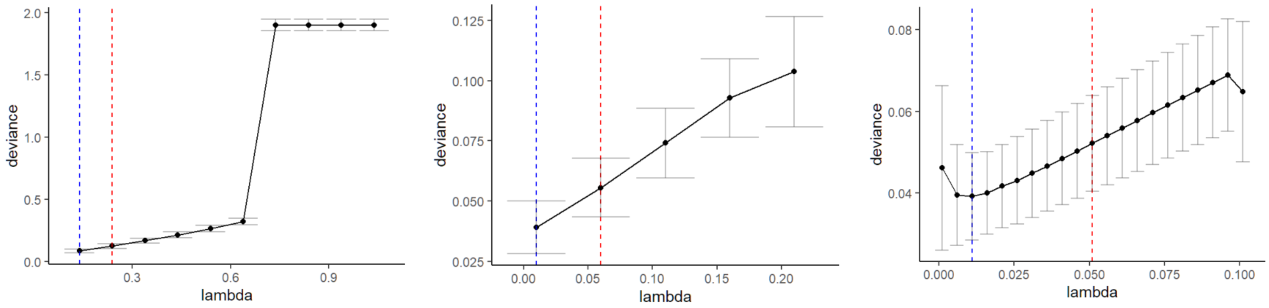
Knowing the log-likelihood function, it is easier to understand how the least absolute shrinkage and selection operator (lasso) functions. Lasso, although originally formulated for linear regressions, is a regularization technique that is able to perform variable selection through penalizing the sum of absolute values of the coefficients. This can be achieved by optimizing the lambda value of the lasso function:

$$lasso = \log \ell(\beta) - \lambda \sum_{j=1}^{p} |\beta_j|$$

where:

- $|\beta_j|$ is the coefficient of the parameter j for $j \in \{1, p\}$. Larger penalty will be assigned to larger $|\beta_j|$, which is then subtracted from the log-likelihood function.
- λ is the regularization parameter that is multiplied to the sum of the absolute value of coefficients. With a larger $\sum_{j=1}^{p} |\beta_j|$, a higher penalty will be assigned, increasing the amount of shrinkage on the coefficients. In the case of lasso, some coefficients will be shrank to zero and thus variable selection is performed. This, however, makes finding the optimal λ value crucial in the success of a lasso model, as setting a λ too high will diminish too many coefficients and cause underfitting, whilst a λ too low can cause the model to remain overfitted.

In this paper, the lambda values used in the final model will be obtained from a set lambda path, where every lambda path is tested through cross validation in set steps. Two lambda values will be chosen, one being the most optimal value in the path, and the other adds 1 standard deviation to the optimal, which acts as a check on underfitting. The following plots show the search of the optimal lambda over the given lambda path, going from larger values to smaller values. The blue dashed line represents the position of the most optimal lambda, with the red dashed line at 1 standard deviation higher. Note that the optimal lambda has not been reached in the first two plots.



2.3.2 Grouped Lasso

The prior lasso model applied a penalty a fixed lambda but applies the punishment on every individual coefficient. In grouped lasso, the penalty is applied on specific divisions of groups, thus the name of grouped lasso. There are numerous potential benefits of using a grouped penalty, including interpretability and even in model performance.

Whilst closely reminiscent of the standard lasso equation, the grouped lasso slightly differs with the introduction of grouping, where g denotes a group, with G represents the total number of groups. The grouped lasso equation can be written as following:

$$grplasso = \log \ell(\beta) - \lambda \sum_{g=1}^{G} \sum_{j \in g} |\beta_j|$$

Within grouped lasso, there are two methods on how the groups can be divided. The two papers selected in this paper are the ”factor” and ”interaction” grouping methods.

2.3.3 Contrasts

In MSE, contrasts play a crucial row in how categorical variables within the contingency tables are encoded, thus effectively changes how the tables are formulated and interpreted. This paper adopts three principal types of contrasts, namely treatment contrast, polynomial (poly) contrasts and sum contrasts.

Treatment contrasts, commonly referred to as dummy encoding, selects one level in the categorical variable as a baseline, which is then used to measure the effects of the other levels. In a design matrix, the baseline level is presented as a row of zeros. Treatment contrasts is highly advantageous compared to the other two contrasts in its interpretability.

On the other hand, polynomial (poly) contrasts are especially useful when dealing with data across a naturally ordered spectrum. Poly contrasts assess trends and patterns by fitting polynomial terms to the levels of a categorical variable, thus capturing linear, quadratic, or higher-order relationships. This can be crucial for identifying and modeling potential nonlinear relationships that might otherwise be ignored in purely linear contrasts.

Finally, sum contrasts , also known as deviation encoding, computes the deviations of the means per category against the overall mean, and thus eliminates the necessity of assigning a single category as the baseline, forming a symmetrical analysis. In MSE, sum contrasts are particularly valuable when the goal is to evaluate the deviation of each category from the overall population estimate.

2.4 Standardization

In lasso and grouped lasso regression, the magnitude of the coefficients is usually linked to the variance of the respective variables, thereby influencing the regularization term effectively applied to each coefficient. Standardization adjusts each variable to have zero mean and a variance of one, thus normalizing the influence of each variable's scale on the analysis. This process is critical as it ensures that the Lasso penalty is uniformly applied across all variables, emphasizes the selection of variables based on their statistical significance and contribution to the predictive power of the model rather than their scale.

Variables exhibiting a considerable variance tend to be associated with smaller coefficients, as a minimal increment in such a variable could significantly influence the dependent variable. On the other hand, these coefficients ($|\beta_i|$) are inherently smaller, leading to a lower penalty in the lasso/grouped lasso model. Therefore, variables with a greater variance are less likely to be eliminated by the lasso penalty, given ceteris paribus. Similarly, variables with smaller variance are more susceptible to larger penalties, increasing their likelihood of exclusion from the model.

Overall, testing the implementation of grouped lasso in MSE with and without standardization would be an interesting approach, and discover any possible performance discrepancies between the two scenarios.

3 Simulation / Model Setup

The simulation of various setups is generated using the simdat2 function, which generates a sequence of randomly generated observed frequencies form the same set of randomly generated log linear parameters. It has a range of arguments that can be specified to form the desired data structure, some of which include:

- *regs* number of registers.
- *xlevs* vector with the number of levels of the covariates. If `NULL`, no covariates are made.
- N population size.
- k number of simulated data sets.
- $mean, sd$ arguments of the function `rnorm` used for drawing the log linear parameters.
- *contrast* the contrast of the design matrix.
- *standardize* logical for standardization of the design matrix.

- $print_data$ logical to print the contingency table and first 5 simulated frequencies along with some summary statistics to the screen.
- *seed* optional seed for reproducibility.

In the base setup, 3 registers, 2 covariates, 3 levels per covariate, and 25% of the population in the sample were used, with standardization. For the various setups, the default total population of 1000 is used, with k fixed at 10, generating a total of 10 data sets for each setup.

Lower percentage of population captured in sample:

In this configuration, the same setup of 3 registers, 2 covariates, and 3 levels per covariate is maintained, but only 10% of the population is included in the sample. This reduction is manipulated through lowering the mean parameter in the simdat2 function. The reduction in the sample captured population allows for an analysis of the impact of smaller population in sample on the model's performance and the stability of the estimates. Standardization of the design matrix is applied as in the base setup.

Increasing the number of covariates:

For this variation, the setup includes 3 registers and 3 levels per covariate with 25% of the population size in the sample, but the number of covariates is increased from 2 to 3. This adjustment enables the investigation of the effect of additional covariates on the model's complexity and the interaction terms. The total population remains at 1000, and standardization of the design matrix continues to be implemented to ensure consistency across setups.

4 Results & Analysis

To evaluate the results across the three selected contrast methods and their corresponding AIC/BIC baselines, box plots will be used for each parameter combination. These box plots illustrate the median, interquartile ranges, and outliers for every method. A red horizontal line indicates the chosen true population of 1000. The method that provides the closest median to this red line alongside the narrowest interquartile ranges demonstrates the highest accuracy and precision. In the following box plots, each contrast have two results denoted by 1 and 2, where 2 represents the use of the optimal lambda and 1 represents the use of the larger lambda value. Finally, any values higher than 3000 are deemed as extreme outliers and removed from the box plots for visualization and interpretability of the plots.

4.1 Baseline Data

The results displayed from Figures 1 to 4 are rather surprising, as the AIC and BIC methods provided results that contain the true population within their interquartile ranges, with all grouped lasso methods significantly undershooting the 1000 mark. When excluding its five extreme outliers, AIC has the best performing median value for its predictions. On the other hand, BIC has overshot the true population, albeit with a narrower interquartile range.

The grouped lasso methods, however, have all completely undershot the true population by significant margins. The polynomial contrasts and sum contrasts have performed similarly, both with similar interquartile ranges. On the other hand, despite extremely narrow interquartile ranges, the treatment contrasts obtained even lower median predictions compared to the other contrasts methods. For all three contrast methods, the results obtained with the optimal lambda value provided predictions closer to the true population.

Comparing the use of standardization and interaction/factor grouping, the overall differences have been minor. Figure 1 with standardization and interaction grouping displayed a wider interquartile range for polynomial and sum contrasts, in comparison with the other three figures. A similar narrowing of interquartile ranges is spotted for polynomial and sum contrasts when no standardization is performed, whilst the opposite effect seems true for treatment contrasts.

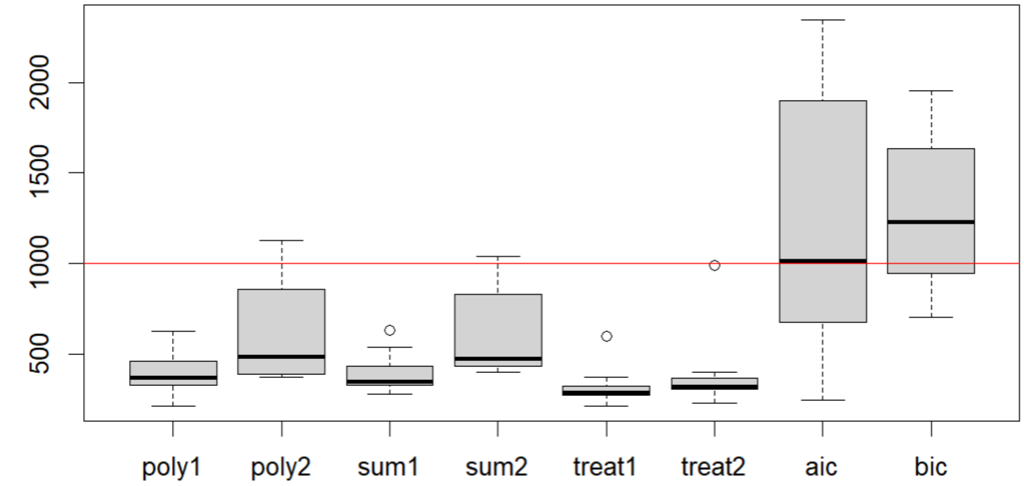


Figure 1: With standardization, interaction grouping, 50% of outliers were removed for AIC

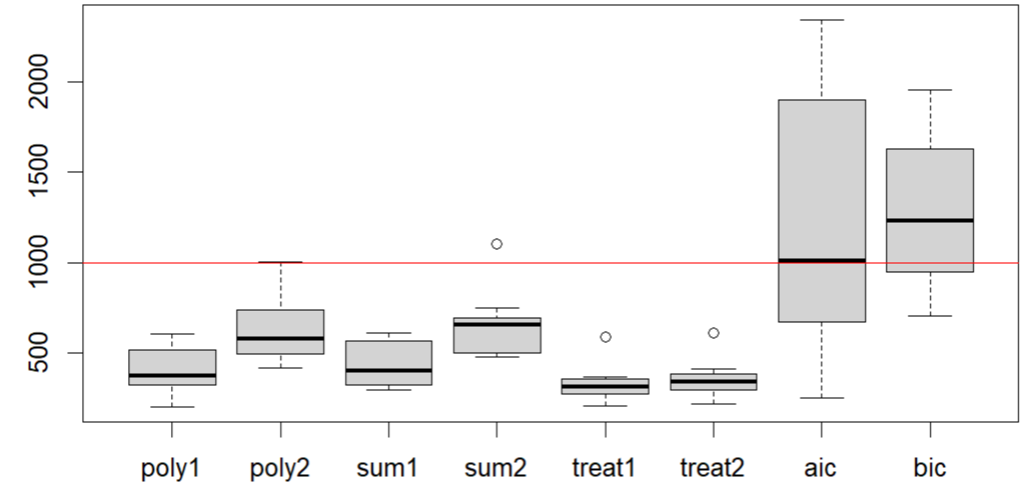


Figure 2: With standardization, factor grouping, 50% of outliers were removed for AIC

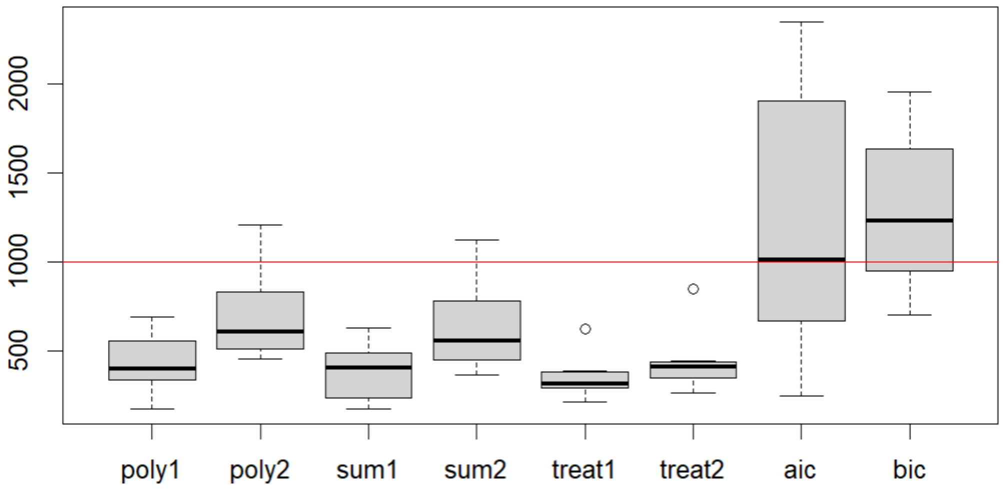


Figure 3: Non standardization, interaction grouping, 50% of outliers were removed for AIC

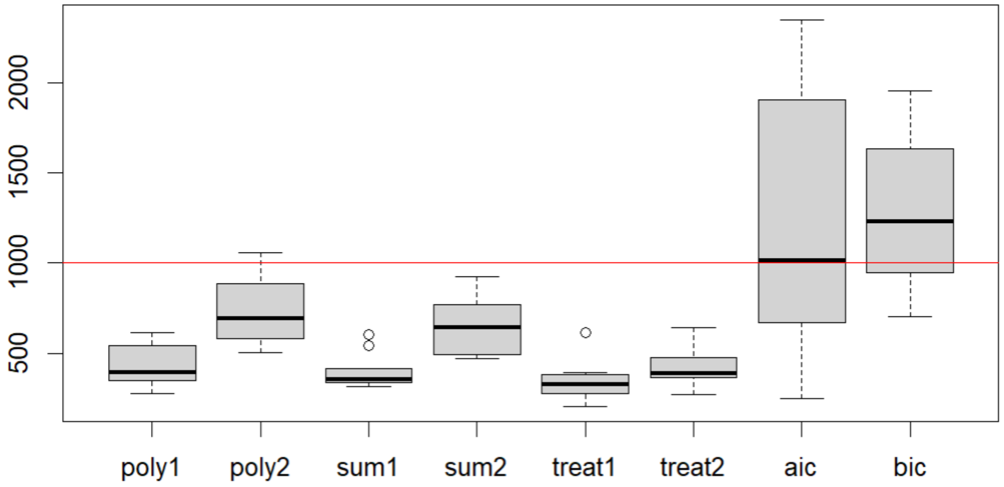


Figure 4: Non standardization, factor grouping, 50% of outliers were removed for AIC

4.2 Lower Percentage of Population in Samples

Due to the lack of significant variation in the trends of results between standardization, non-standardization and the interaction versus factor grouping, only one combination, namely interaction grouping with standardization, will be displayed in this result section as reference and the remaining plots are available in the Appendix section.

Comparing Figure 5 to Figure 1, which has the same setting of standardization and interaction grouping, it can be immediately seen that the interquartile ranges across all methods have been significantly narrowed. Additionally, it should be noted that the extreme outliers removed in Figure 5 has increased, where AIC had 7 extreme outliers and BIC had 3. This makes more than half of all predictions by AIC outliers, suggesting for highly unstable predictions of the method, despite the remaining results scoring close to the true population. BIC has now overtaken AIC as the best performer, despite having more outliers, likely the result of reduced population that is available in the sample.

The reduction in availability of population has also damaged the accuracy of grouped lasso methods, as the medians of all grouped lasso contrasts have further lowered in Figure 5. The trend of extremely narrow interquartile ranges have continues, so does the trend of higher performance observed for polynomial dn sum contrasts. Treatment contrasts have remain the lowest predictions, albeit with the narrowest interquartile ranges as well.

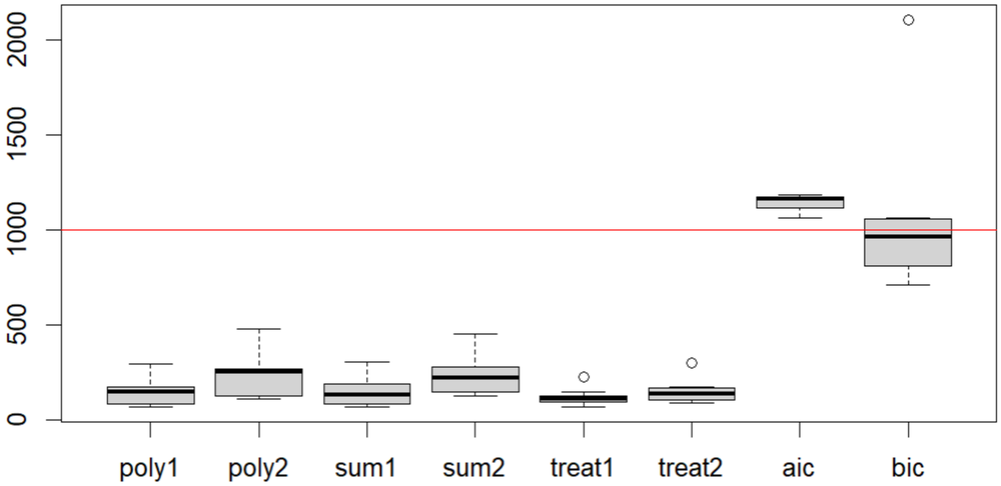


Figure 5: With standardization, interaction grouping, 7 outliers were removed for AIC and 3 outliers for BIC

4.3 Higher Number of Covariates

Finally, Figure 6 provides the results when the number of covariates increases from 2 to 3. Similar to the prior section, only one combination, namely interaction grouping with standardization, will be displayed in this result section as reference and the remaining plots are available in the Appendix section.

Here, the results are shocking as all AIC predictions were higher than 3000 and are excluded from the plot, whilst one outlier has been removed in the case of BIC. It is possible that AIC had overfitted due to increased number of covariates, and the tendency of BIC to penalize models with a large number of parameters more severely than AIC has likely avoided the overfitting. That said, the prediction of BIC has worsened slightly compared to the baseline, so even BIC is not completely immune to the increase in number of covariates.

Similarly, the median of predictions for the grouped lasso methods are lower than the baseline counterparts. The top and bottom fences of all three contrasts, on the other hand, are relatively close with one another, despite treatment contrasts scoring lower in the overall median value. In the setting of increased number of covariates, using a larger value of lambda narrowed their performances relative to using the optimal lambda values, when compared to the rather significant differences that were observed with the baseline data, suggesting again that the higher punishment with a higher lambda value may have been the key when narrowing the performance gap.

5 Discussion

Multiple System Estimation (MSE) has been a widely used statistical method for population estimation in various fields, such as social sciences, ecology, and epidemiology. It is especially useful in scenarios where full population counts are impossible or infeasible to obtain. Traditional MSE methodologies, like log-linear analysis of contingency tables, can be computationally intensive in model selection, due to the exponential increase in potential models introduced from more registers and covariates. In effort to remedy the model selection prob-

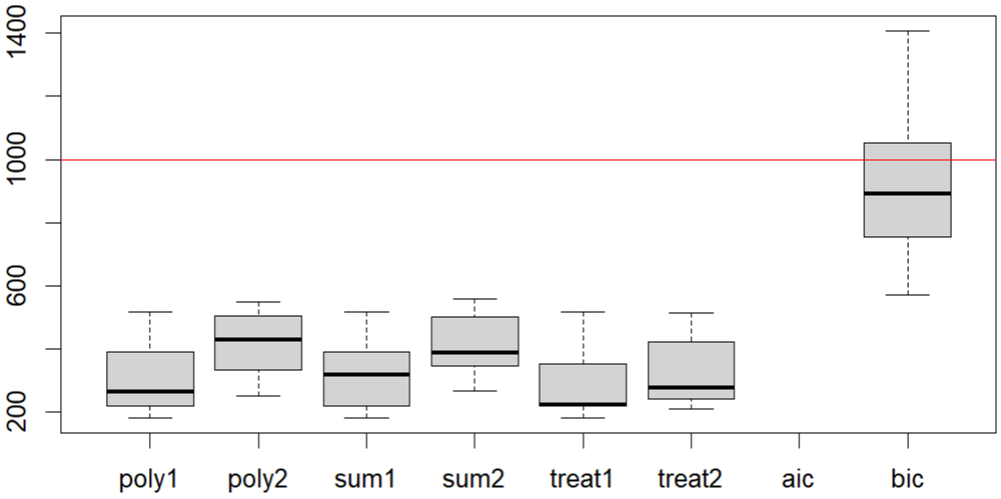


Figure 6: With standardization, interaction grouping, all AIC were removed and 1 outlier for BIC

lem, regularization techniques like Lasso and grouped lasso regression were explored, aimed at offering an efficient approach by penalizing and removing the less impactful coefficients, thus streamlining the model selection process.

In this study, the performance of different Grouped Lasso regression models was evaluated through multiple simulations of data by adjusting various parameters. That includes setting a baseline data with 3 registers, 2 covariates each with 3 levels and 25 % of population captured in sample. Three contrast methods are analysed for any data, namely the treatment, sum and polynomial contrasts, alongside traditional log-linear regression modelling that are evaluated based on AIC and BIC. Additionally, the effect of interaction versus factor grouping and the effect of standardization is explored using the baseline data. Population parameter is then adjusted, lowering the percentage of population captured in sample to 10%, followed by adjusting the number covariates from 2 to 3.

The results from these simulations showed that grouped Lasso methods generally produced low medians relative to the true population, with narrow interquartile ranges and no significant outliers. Performances with the optimal lambda value were superior to using larger lambda values, which is unsurprising given the optimal lambda is selected using a long lambda path with many steps. Step-wise model selection using AIC/BIC provided accurate predictions but exhibited high variation and many outliers, which was particularly the case with

AIC. This highlighted a probable use case of group lasso methods, to serve as a lower bound baseline, especially when AIC/BIC methods result in severe outliers.

The study has faced many factors which placed limitations on the study designs and the actual simulations. Due to the high runtime requirements of the algorithms, the exploration of a broader range of parameter combinations were restricted. Many of potential parameters from the Simdat2 function remain untested, including the number of levels for covariates, as well as other population settings. Frequent errors from the Lasso package also necessitated manual interventions, reducing automation and extending runtime.

Future research should focus on testing a wider range of parameter combinations, including adjustments to registers, covariates, categorical variable levels, and population settings. Additionally, exploring the bias-variance trade-off that is inherent in regularization methods could provide further insights, as these methods tend to produce predictions with higher bias and lower variance, as observed in the results of this paper. Experimenting with alternative regularization techniques, including the likes of ridge regression and elastic net could also be beneficial.

In conclusion, while Grouped Lasso methods showed decent results at optimal lambda values, with the lack of significant outliers, they were effective only within a narrow range of parameters over existing AIC/BIC methods. Addressing the computational and algorithmic limitations and expanding the scope of parameter testing could further examine the conditions under which Grouped Lasso methods can be most beneficial.

6 References

Binette, O., & Steorts, R. C. (2022). On the Reliability of Multiple Systems Estimation for the Quantification of Modern Slavery. Series a. Statistics in Society/Journal of the Royal Statistical Society, 185(2), 640–676. https://doi.org/10.1111/rssa.12803

Cruyff, M., Overstall, A., Papathomas, M., & McCrea, R. (2021). Multiple System Estimation of Victims of Human Trafficking: Model Assessment and Selection. Crime & Delinquency, 67(13-14), 2237-2253. https://doi.org/10.1177/0011128720981908

Hald, A. (2005). A history of probability and statistics and their applications before 1750. John Wiley & Sons.

R. King, S. P. Brooks, On the Bayesian Estimation of a Closed Population Size in the Presence of Heterogeneity and Model Uncertainty, Biometrics, Volume 64, Issue 3, September 2008, Pages 816–824, https://doi.org/10.1111/j.1541-0420.2007.00938.x

Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. J. R. Statist. Soc. B, 58, 267–288.

7 Appendix

7.1 Codes

The codes to this thesis topic can be accessed via: `https://drive.google.com/drive/folders/1avuJiyzCn0lRGSQGqj4q6gMSlEJu_nFD?usp=drive_link`

The file names follow the following conventions:

323/333 represents the number of registers, number of covariates and the number of levels per covariate. k10 represents 10 sets of simulations per parameter combination. 10/25 represent the percentage of in-sample population. Finally, simdat2 represents the data simulation function of simdat2 that has been used.

7.2 Lower Percentage of Population in Samples

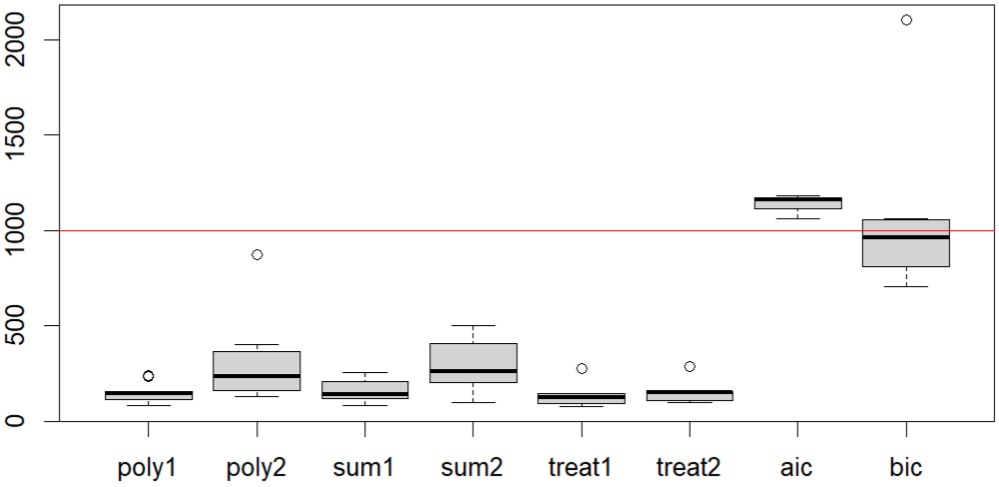


Figure 7: With standardization, factor grouping, 7 outliers were removed for AIC and 3 outliers for BIC

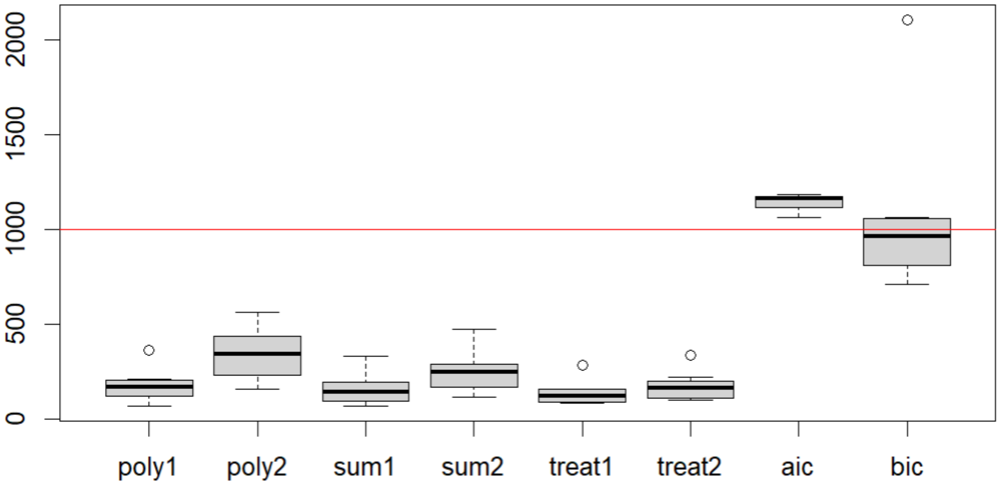


Figure 8: Non standardization, interaction grouping, 7 outliers were removed for AIC and 3 outliers for BIC

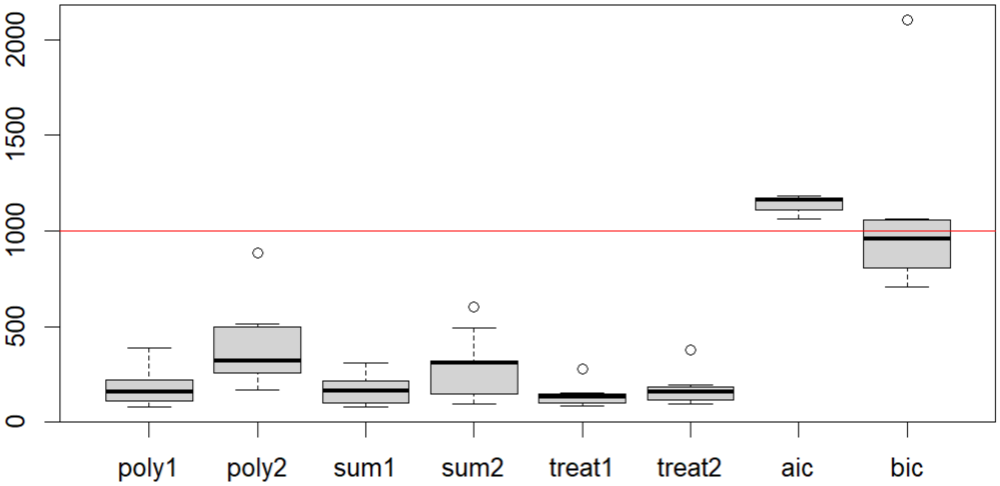


Figure 9: Non standardization, factor grouping, 7 outliers were removed for AIC and 3 outliers for BIC

7.3 Higher Number of Covariates

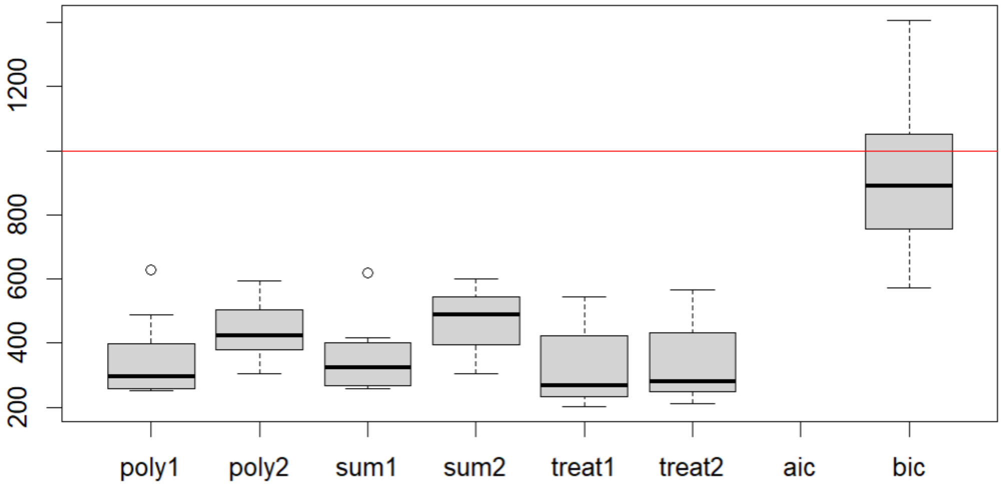


Figure 10: With standardization, factor grouping, all AIC were removed and 1 outlier was removed for BIC

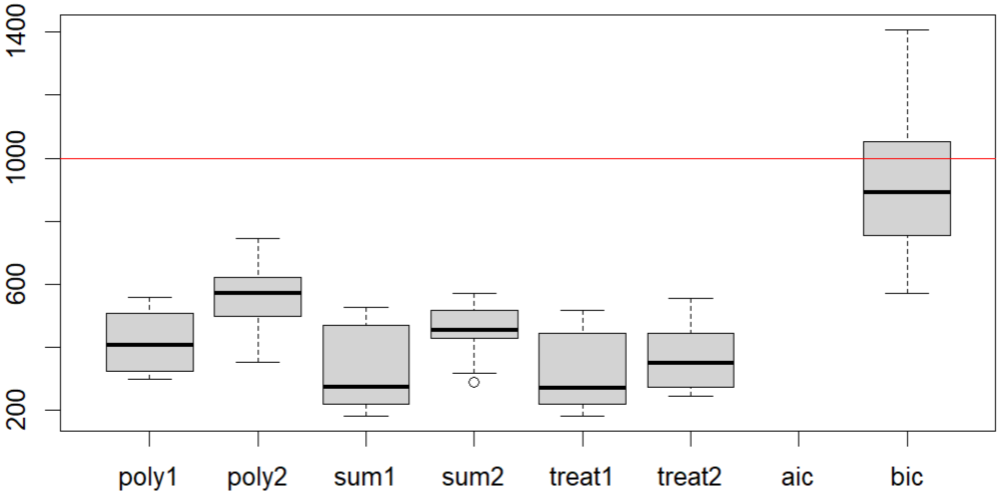


Figure 11: Non standardization, interaction grouping, all AIC were removed and 1 outlier was removed for BIC

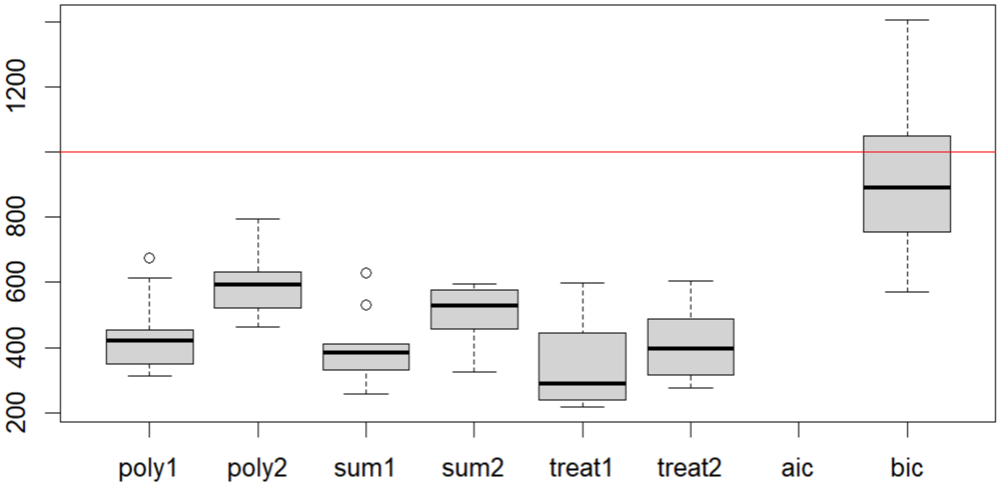


Figure 12: Non standardization, factor grouping, all AIC were removed and 1 outlier was removed for BIC