Comparative Judgement Methods for Poverty Estimation in Dar es Salaam, Tanzania

University of Nottingham UK | CHINA | MALAYSIA

James Briant | **Rowland Seymour (Supervisor)**

School of Mathematical Sciences

Introduction

Dar es Salaam is a coastal city in Tanzania and is the largest city in East Africa with a population over 12 million. Local residents were asked to choose which region of the city they deemed to be "more deprived" in a series of pairwise comparisons. This project aims to analyse these data by comparing two variations of the Bradley-Terry model - one allowing for ties and the other ignoring ties in the data.

Bradley-Terry Model

The Bradley-Terry model (Bradley and Terry, 1952) assumes each of N regions has some underlying parameter, $\lambda_i \in \mathbb{R}$, which describes the relative level of deprivation. The probability that region i is determined to be more deprived than region j is given by

$$\pi(i > j) = \frac{\exp(\lambda_i)}{\exp(\lambda_i) + \exp(\lambda_j)}$$

In the **non-ties model** it is assumed

$$Y_{ij} \sim \text{Binomial}(n_{ij}, \pi(i > j))$$

where Y_{ij} is the random variable denoting the number of times region i beat region j and n_{ij} denotes the number of times i and j are compared. In the **ties model** another outcome is required. Davidson (1970) suggests the following

$$\pi(i > j) = \frac{\exp(\lambda_i)}{\exp(\lambda_i) + \exp(\lambda_j) + \nu\sqrt{\exp(\lambda_i + \lambda_j)}}$$

$$\pi(i = j) = \frac{\nu\sqrt{\exp(\lambda_i + \lambda_j)}}{\exp(\lambda_i) + \exp(\lambda_j) + \nu\sqrt{\exp(\lambda_i + \lambda_j)}}$$

where ν is some parameter to be estimated. This means Y_{ij} has a multinomial distribution.

Bayesian Approach

The likelihood of the multinomial distribution is given below. The non-ties model can be extracted by taking each $t_{ij} = 0$

$$\pi(\boldsymbol{y}, \boldsymbol{t} \mid \lambda_1, \ldots, \lambda_N) = \prod_{i=1}^{N} \prod_{j<i} \frac{n_{ij}!}{y_{ij}! t_{ij}! (n_{ij} - y_{ij} - t_{ij})!} \pi(i > j)^{y_{ij}} \pi(j > i)^{n_{ij} - y_{ij} - t_{ij}} \pi(i = j)^{t_{ij}}$$

Using a Bayesian approach allows for the use of a range of different prior distributions depending on how much knowledge on each of the regions is assumed. Assuming each region is independent and using Bayes' Theorem leads to the following

$$\pi(\lambda_1, \ldots, \lambda_N, \nu \mid \boldsymbol{y}, \boldsymbol{t}) \propto \pi(\boldsymbol{y}, \boldsymbol{t} \mid \lambda_1, \ldots, \lambda_N) \prod_{i=2}^{N} \pi(\lambda_i) \pi(\nu)$$

Markov Chain Monte Carlo Algorithm

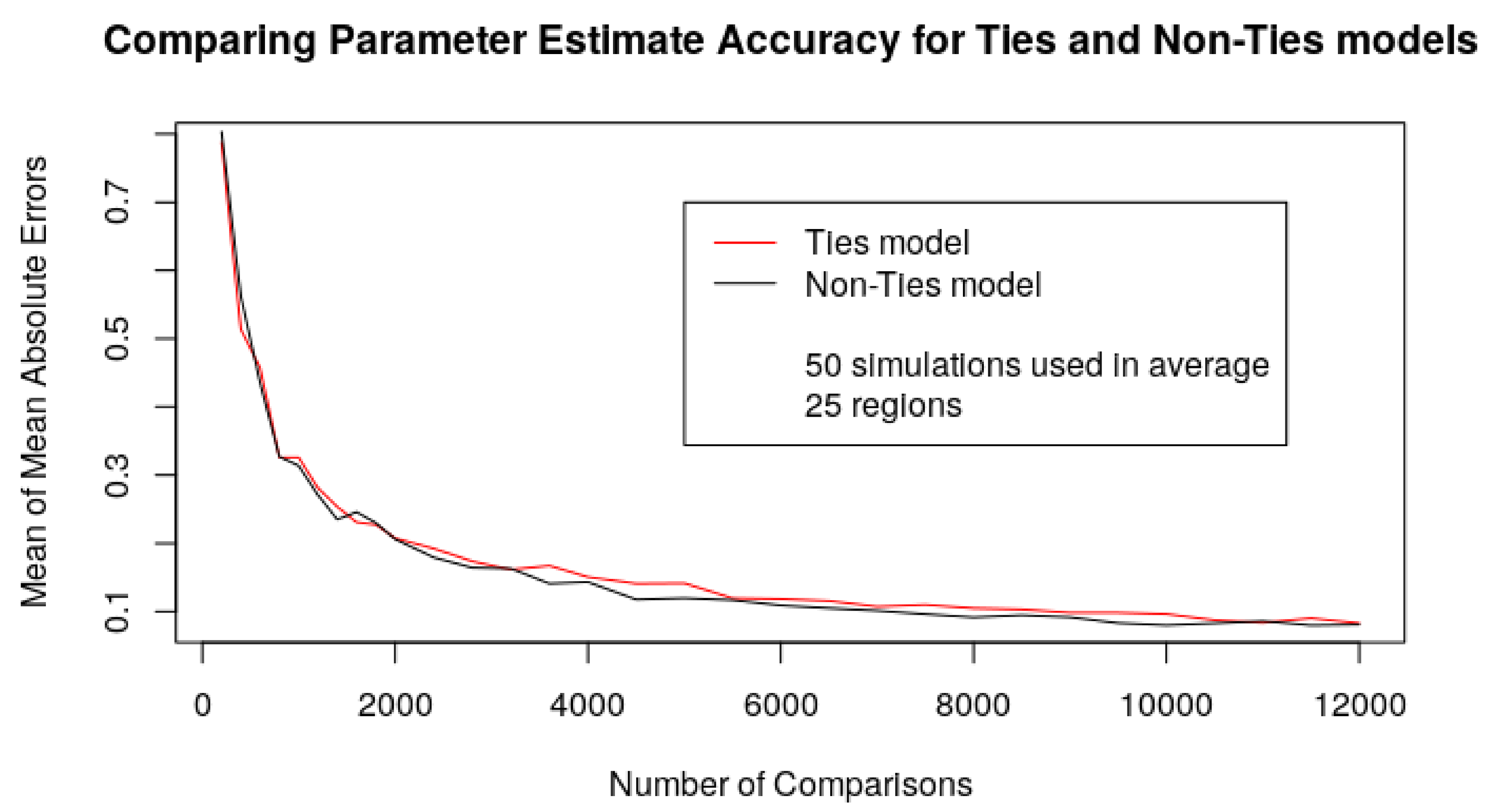
We develop an MCMC algorithm to infer the model parameters.

Algorithm 1 MCMC Algorithm for Estimating Model Parameters

Initialise the chain with values $\boldsymbol{\lambda}^{(0)}$ and $\nu^{(0)}$
for iteration i of the MCMC algorithm
    for all λ_j in $\boldsymbol{\lambda}$
        Update λ_i using a Metropolis-Hastings step
    Update ν using a Metropolis-Hastings step

Simulation Study

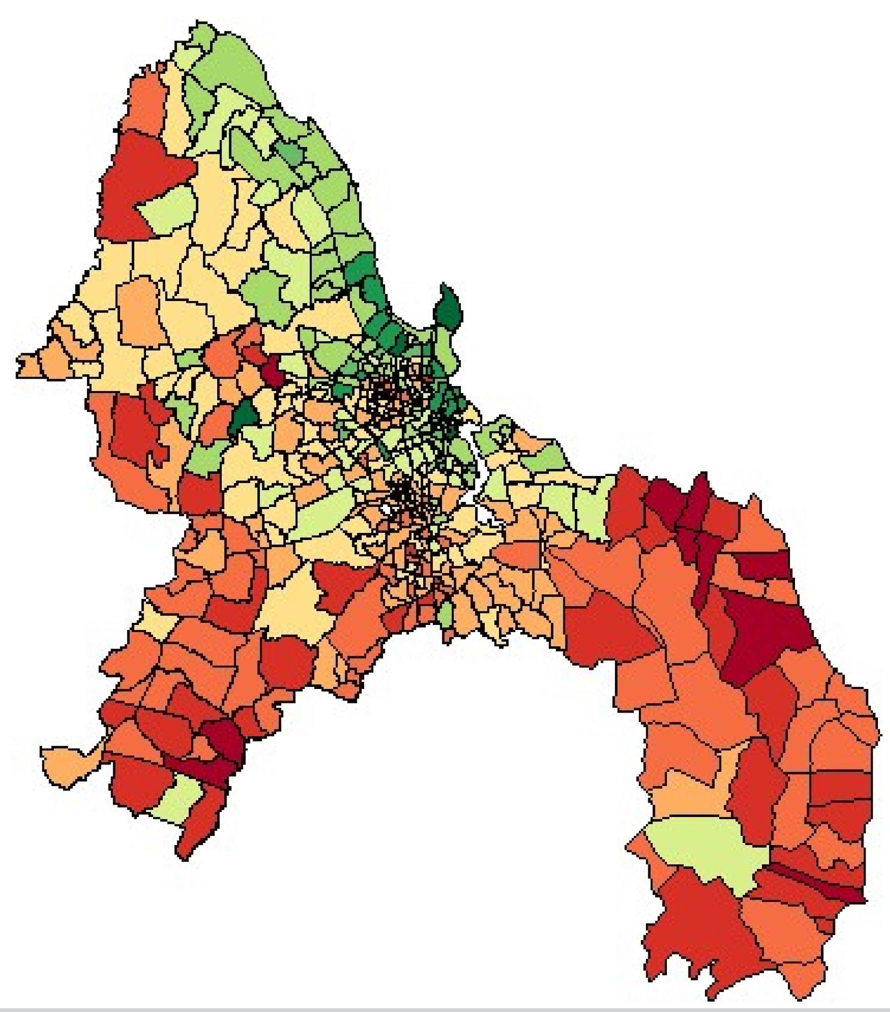
By taking an average over 50 simulations, the mean of the mean absolute error of the deprivation parameters $\boldsymbol{\lambda}$ provides a good indication of the accuracy of the models for different amounts of data.



The plots above suggests the ties and non-ties models perform similarly well. This indicates that despite a tie carrying less information than a non-tie, the ties model successfully extracts the same amount of information as the non-ties model through inference of ν.

Model Results

The map below gives a visual representation of the deprivation in each of the regions of Dar es Salaam. The colour corresponds to the value of λ_i associated with the region as estimated using the Bradley-Terry model.



The greener regions represent the more affluent regions as determined by the comparative judgement data. These regions typically have many tourist attractions and hotels. The red regions are described as the more deprived areas. This includes the more rural and suburban areas of Dar es Salaam.

Conclusions

The Bayesian approach accurately estimates the Bradley-Terry model parameters, especially so for large amounts of data. However, there is no benefit to including ties data in the model. It is therefore recommended to force users to make a decision when collecting data in the field.

References

Bradley, R. A. and Terry, M. E. (1952). Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345.

Davidson, R. R. (1970). On extending the bradley-terry model to accommodate ties in paired comparison experiments. *Journal of the American Statistical Association*, 65(329):317–328.

UKRI Engineering and Physical Sciences Research Council

This work is supported the UK Engineering and Physical Sciences Research Council and the Big East African Data Science research group at the University of Nottingham.