MULTIVARIATE BAYESIAN HIERARCHICAL MODELS WITH COPULAS-BASED ASSUMPTIONS ON SPATIAL EFFECT

Mu Li, School of Demography, ANU

Supervised by

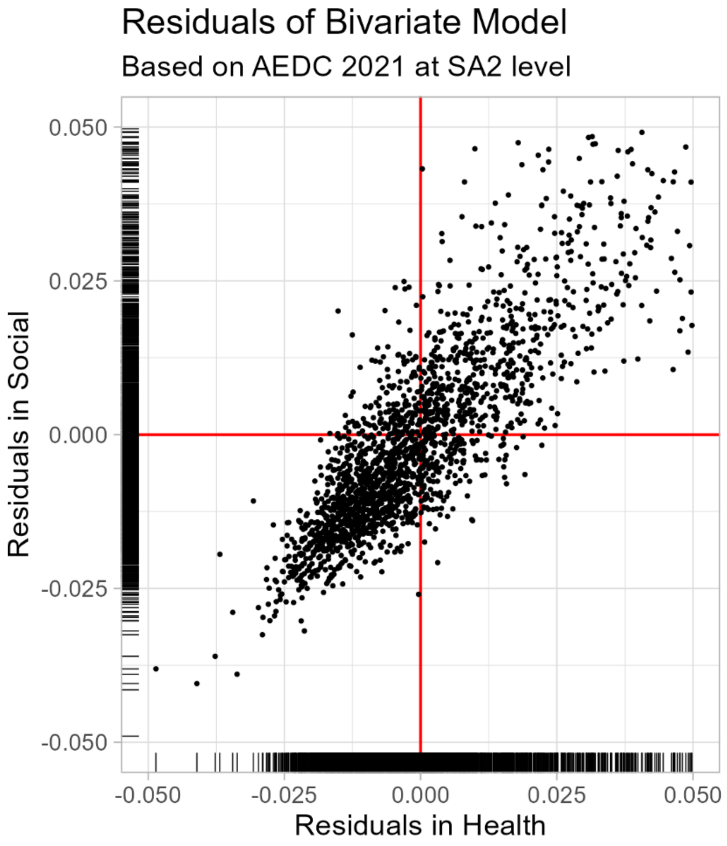
Dr. Bernard Baffour, School of Demography, ANU

Australian National University

Background

Bayesian hierarchical models

- A well-used small area estimation methods. To get a more accurate estimation when the sample size is small in each small area, the model borrow the strength from
  - Other data with much more samples (Census data for example)
  - Spatial information – area that near to each other are tend to perform the same
    - Conditional autoregressive (CAR) models [1]
- Multivariate models can capture the relations between two response together
  - Issue: Correlation between two response (or residuals of each response) may not be constant and symmetric – this would be usually happened between two risks – *"It never rains, but it pours"*
    - Key Questions: Construct a new method that can capture this non-linear correlation in small area estimation



By author

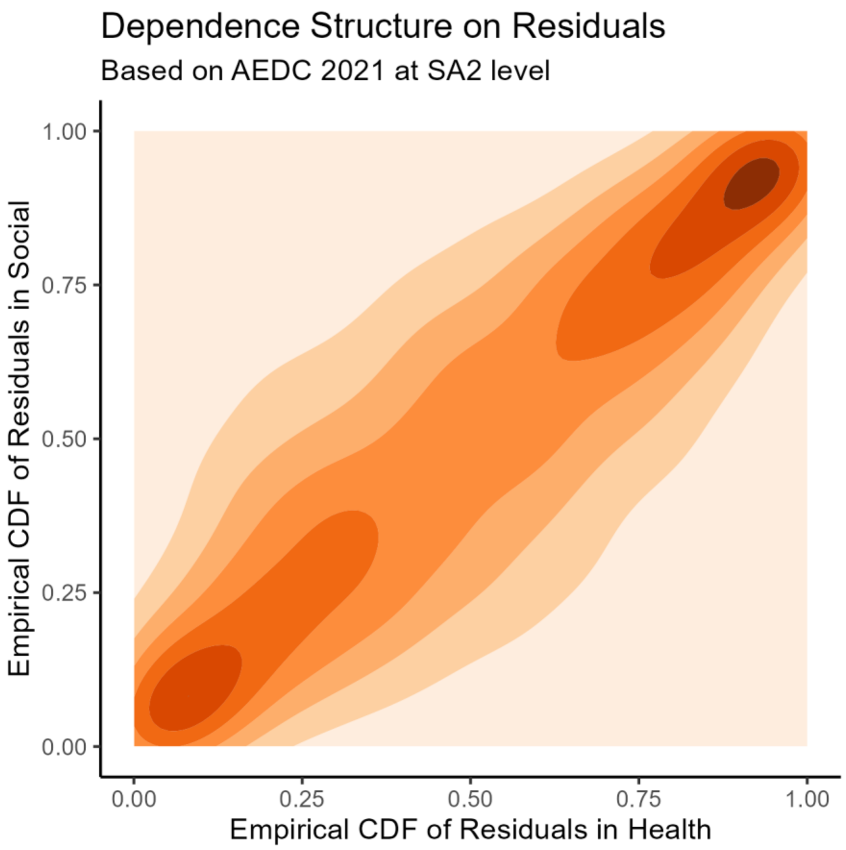


Image credit here

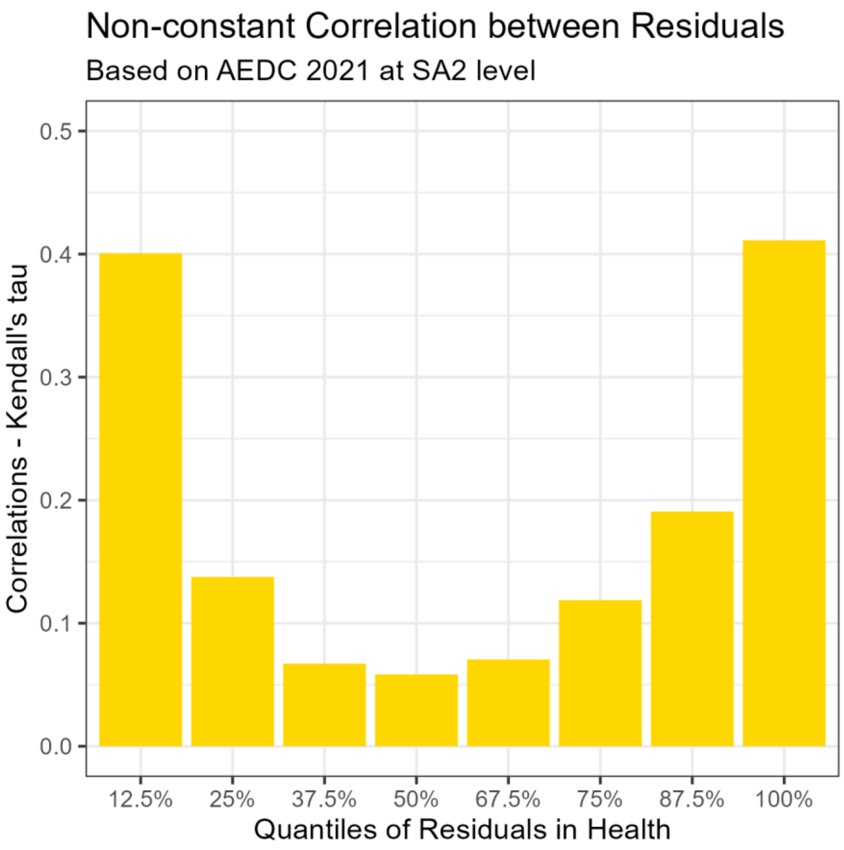


Image credit here

Non-symmetric Correlation

Where does the non-symmetric correlation come from?

Simulations for Non-symmetric Correlation

We would like to investigate the possible theoretical source of non-symmetric correlation by simulation in a simple way.

Unobservable, Catched in Res

- Generate bivariate Y

$$Y_1 = X_1 \boxed{+ X_2 + \epsilon_1}$$

$$Y_2 = X_1 \boxed{+ X_2 + X_3 + 0.3X_2X_3 + \epsilon_2}$$

  – Where we just simply assume that the fixed effect X_1, X_2, X_3 and random effects ϵ_1 and ϵ_2 are all generated from standard normal distributions

  – We also assume that when run simple linear regression models, the only observed value would be X_1, i.e. $Y \sim X_1$, where the rest of the fixed value are "latent variables" that can only be captured by residuals.

- What would residuals $Res(Y_1)$ and $Res(Y_2)$ consist of?

  – X_2 both contained in $Res(Y_1)$ and $Res(Y_2)$, which is the source of correlation level

  – ϵ_1 and ϵ_2 are independent

  – $Res(Y_2)$ contains additional values $X_3 + \frac{1}{4}X_2X_3$ which represent that effects from the hidden variable X_3 and the interaction terms between X_2 and X_3

    » *Let's check the residuals!*

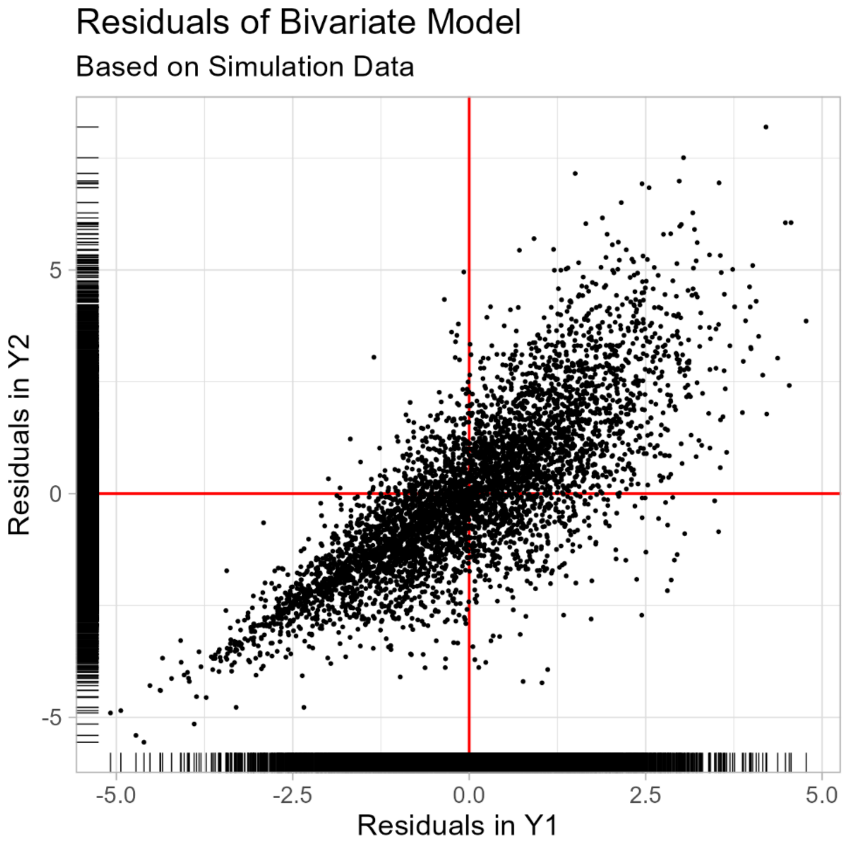


Image credit here

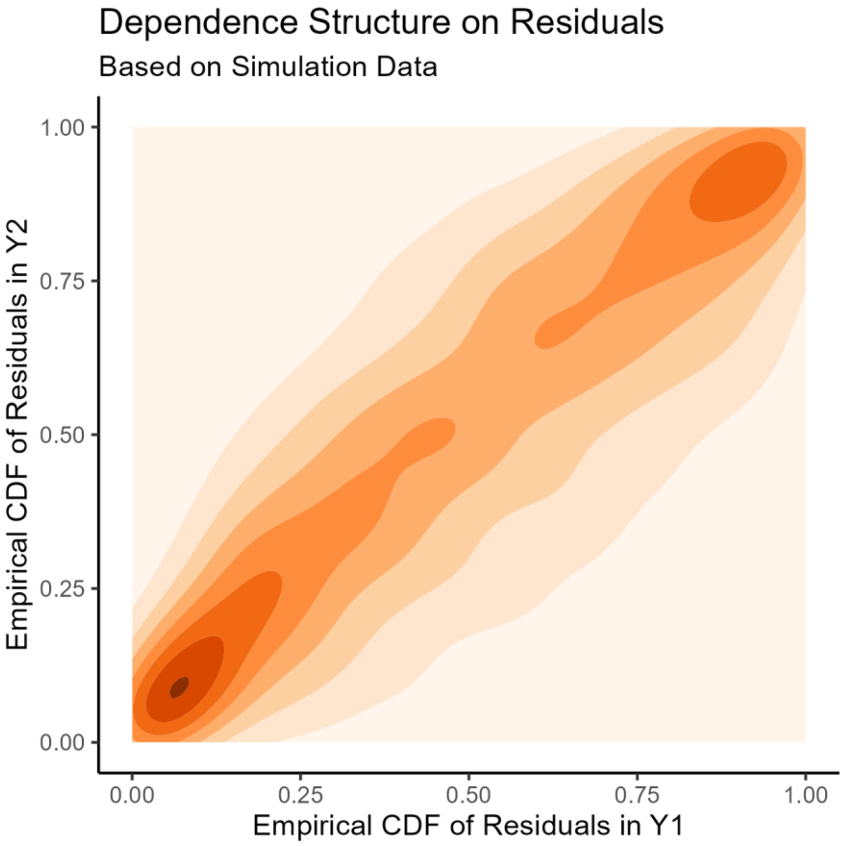


Image credit here

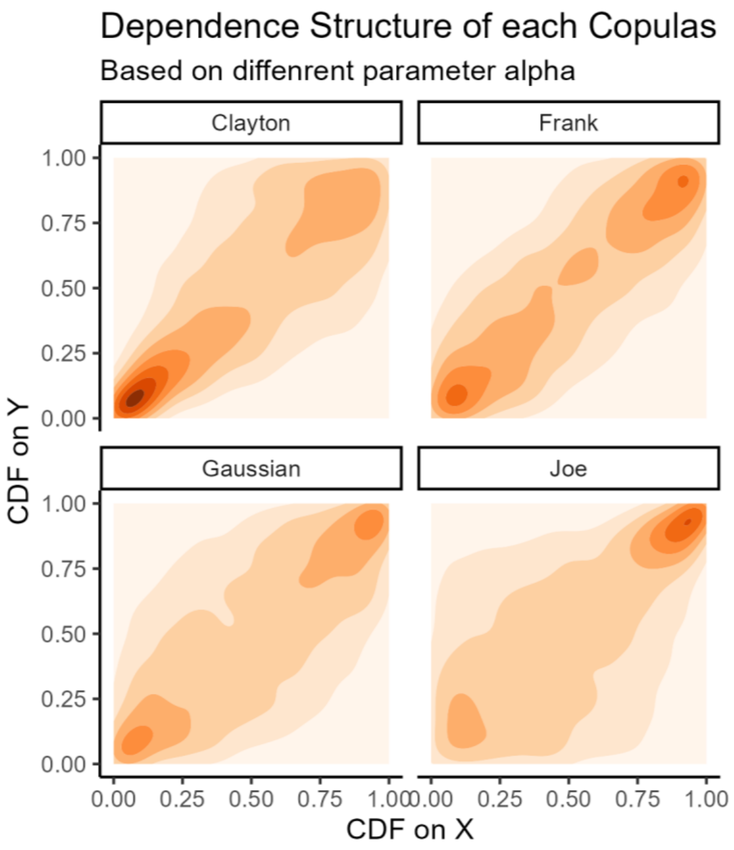


Image credit here

Methods

Idea of Copulas

A powerful tool for decompose the joint distribution into marginal distribution and the dependence structure [2].

$$F_{XY}(x,y) = C(F_X^{-1}(x), F_Y^{-1}(y))$$

$C(.,.)$ denotes for a copula function that could capture the dependence structure

- We could choose a copula function that could give a non-linear correlation performance shown by data
  - Clayton copula
    - High correlation in one-side tail, with only one parameter to adjust the shape of the dependence structure

$$C(u,v) = (u^{-\alpha} + v^{-\alpha} - 1)^{-\frac{1}{\alpha}}$$

Methods

Idea of Conditional Autoregressive Model

Areas that near to each others tends to perform in the same pattern

- Could be tested by Local Moran or Local Indicators Spatial Associations (LISA)
- The mean of the random effects were conditionally determined by the value of random effects of the neighbours; And the variance of random effects were conditionally determined by the number of the neighbours – It is a Gaussian Field
- Value of ρ represents the strength of spatial effects

$$(\psi_i|\psi_{-i}, W) \sim N(\frac{\rho\sum_{j=1}^{R} w_{ij}\psi_j}{1+\rho\sum_{j=1}^{R} w_{ij}-\rho}, \frac{\tau}{1+\rho\sum_{j=1}^{R} w_{ij}-\rho})$$

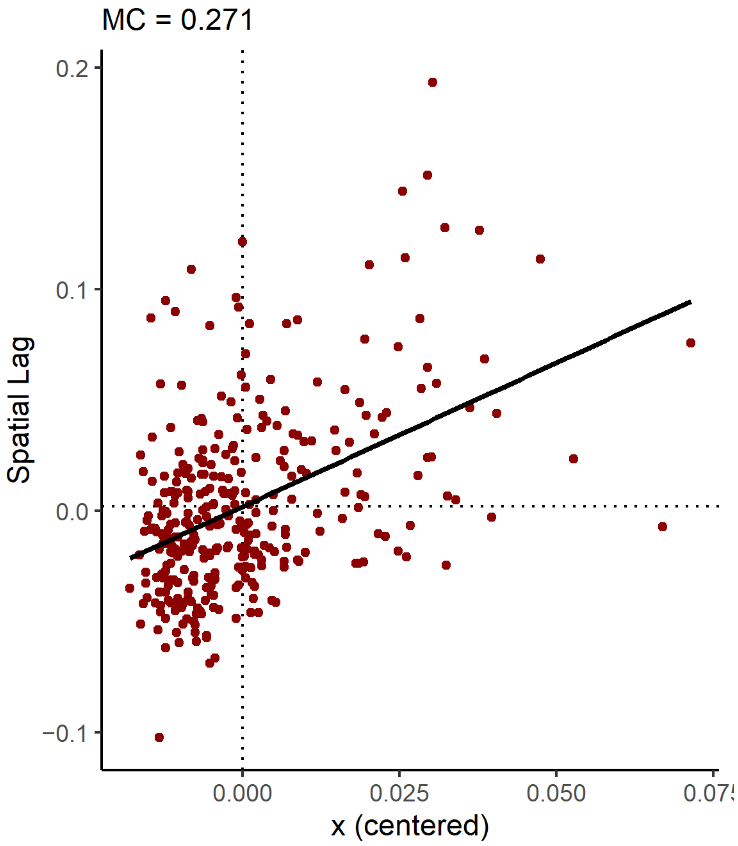


Image credit here

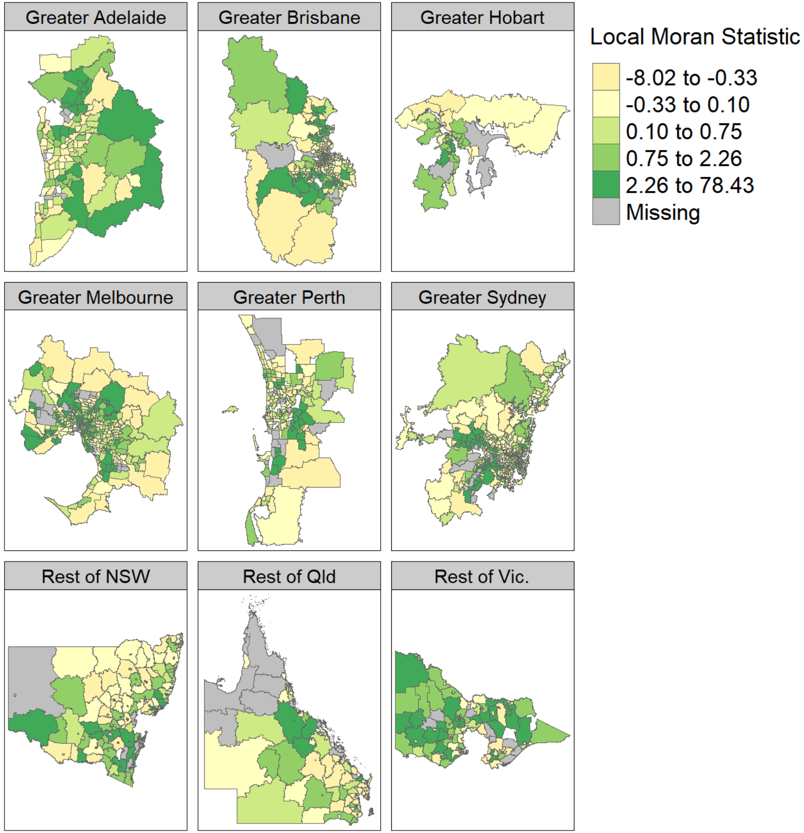


Image credit here

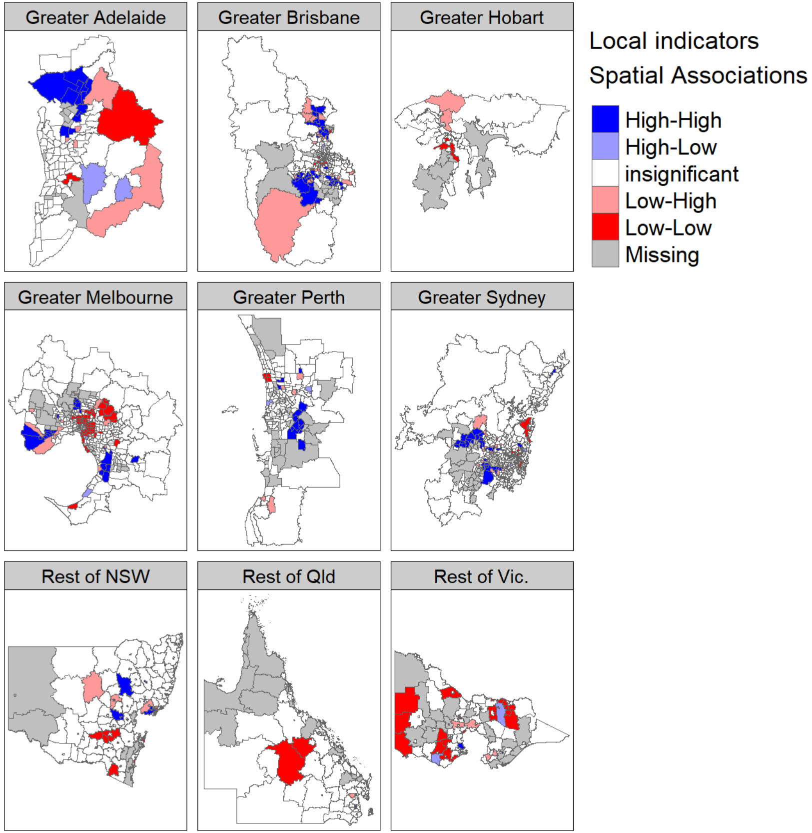


Image credit here

Methods

Copula-based CAR

Combine Copula with multivariate CAR

Now, we apply the CAR models on marginal distributions and generate the joint distribution with the Copulas when fitting the Bayesian hierarchical model

$$y_{ir} \sim Binomial(N_{ir}, p_{ir})$$

$$\log\left(\frac{p_{ir}}{1-p_{ir}}\right) = \boldsymbol{X\beta_i} + \boldsymbol{\psi_{ir}}$$

$$\boldsymbol{\psi_r} \sim C_\alpha\ (\psi_{1r}, \psi_{2r})$$

- Specifications of the Binomial Case:
  - y_{ir} denotes the responses of ith domain in rth area;
  - N_{ir} denotes size of total valid observations of ith domain in rth area;
  - $\boldsymbol{\psi_r}$ denotes the bivariate random effect that follows the joint distribution constructed by copula function $C(.,.)$ with parameter α, and the marginal distribution of the random effects from the first and the second domain ψ_{1r} and ψ_{2r}
  - ψ_{1r} and ψ_{2r} are generated from the CAR model
    - » **This capture both spatial effects and the non-symmetric correlations**

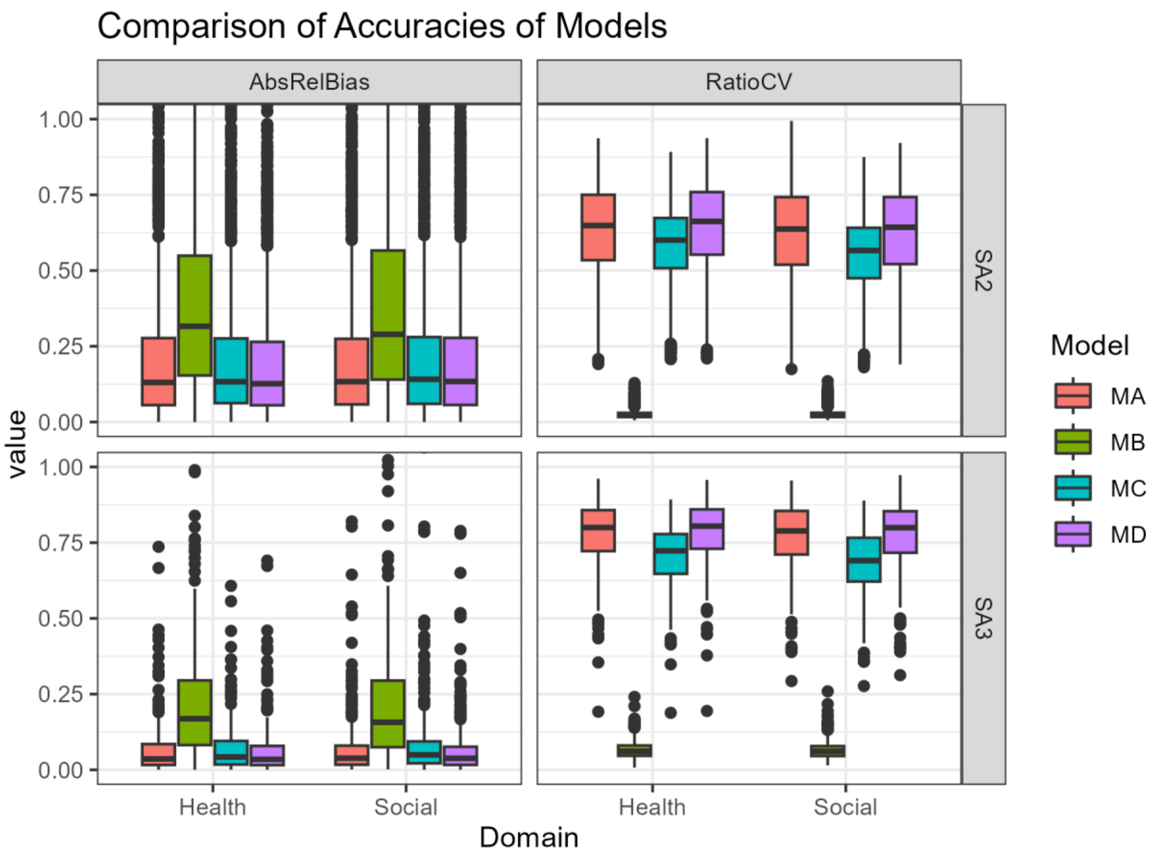
Results

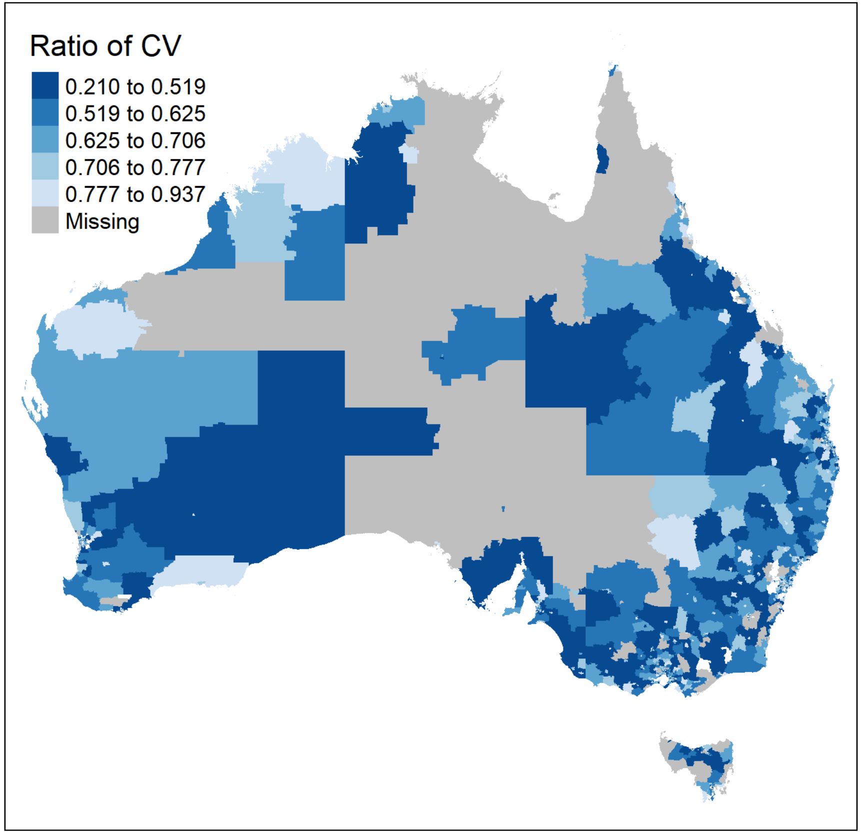
Models Performances

Model Specifications:

- Model A: univariate CAR model for each domain separately
- Model B: univariate GLM model for each domain separately
- Model C: Bivariate CAR model
- **Model D: Bivariate CAR model with Clayton Inverse Copula assumptions**

- Ratio of Coefficients of Variation (CV)
  - Compare the CV with the direct estimation method, each method reduce the CV obviously
  - **Copula-based model give an acceptable CV improvement**
- Absolute value of Relative Bias (RB)
  - Compare the estimates with the direct estimation
  - Considering Spatial effects reduce the relative bias
  - **Copula-based model give the lowest Relative Bias**





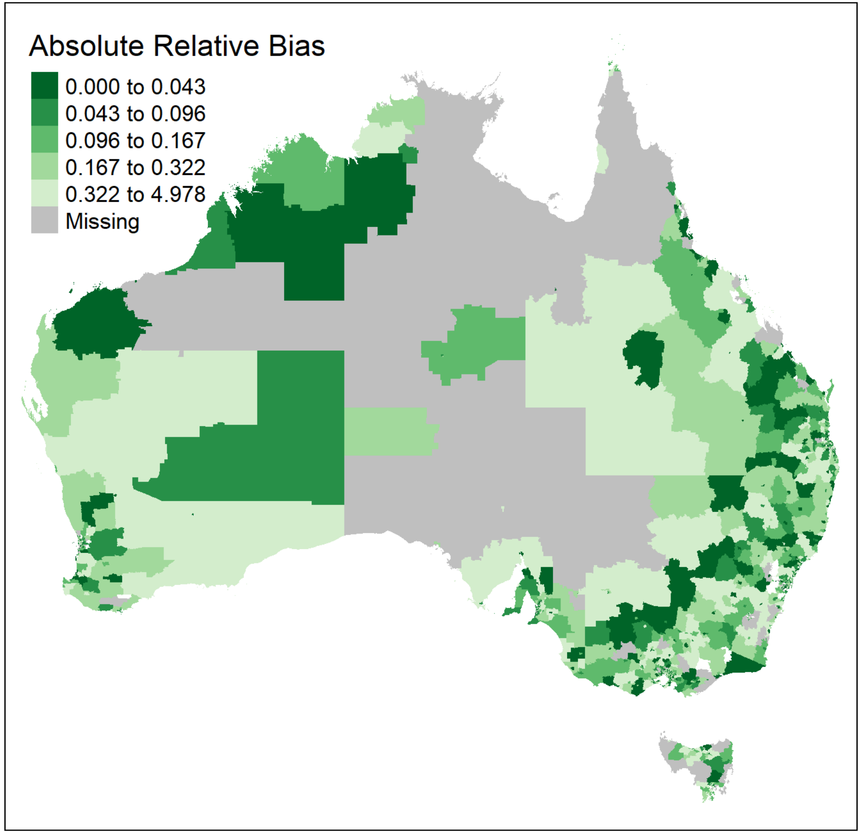
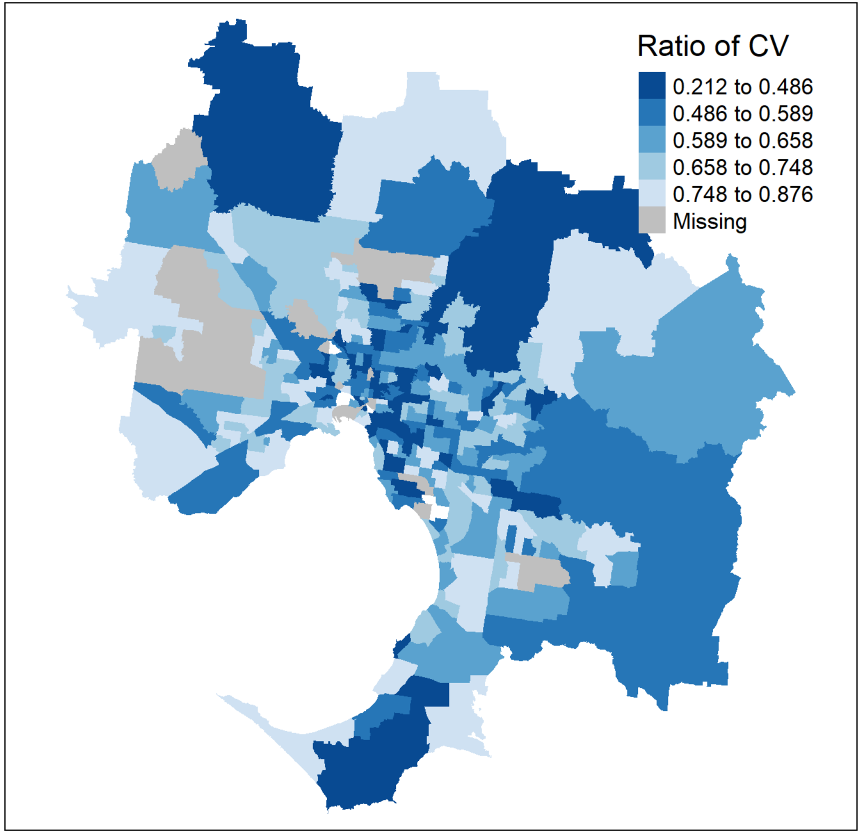


Image credit here



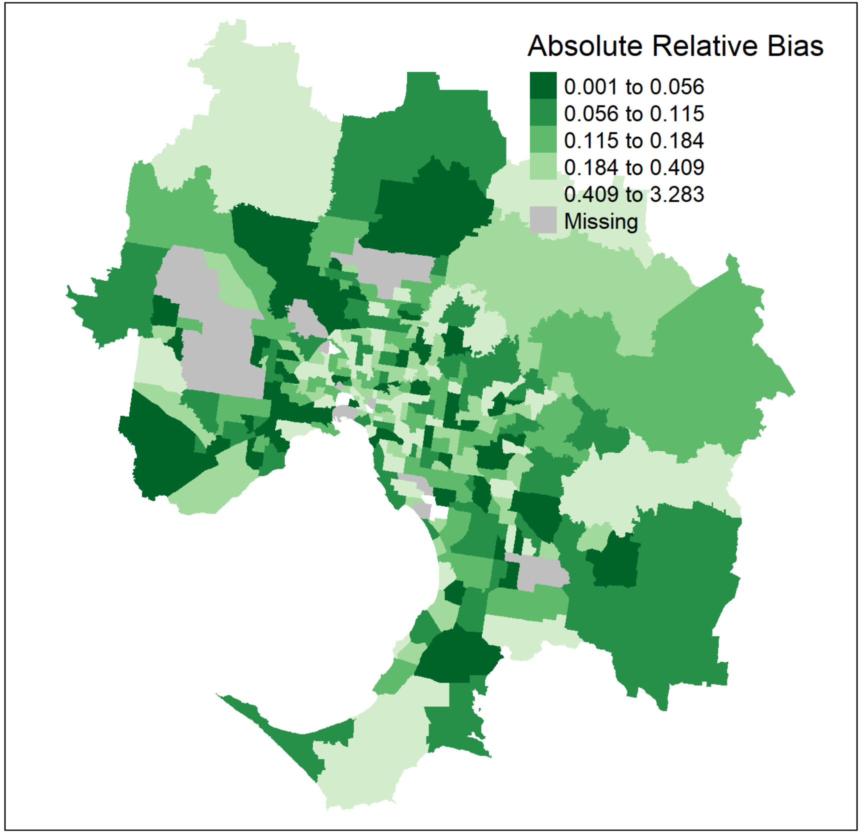


Image credit here

THANK YOU

Contact Us

Mu Li
School of Demography, ANU
Email: mu.li@anu.edu.au
R Package: https://github.com/ArmyNoBusiness/CARBayesCopulas.git

Reference List:

[1] Lee, D. (2011). A comparison of conditional autoregressive models used in Bayesian disease mapping. Spatial and spatio-temporal epidemiology, 2(2), 79-89.
[2] Sklar, A. (1973). Random variables, joint distribution functions, and copulas. Kybernetika, 9(6), 449-460.

Australian National University

THANKS!

DON'T FORGET TO COME AND JOIN US IN BRISBANE 13-18 JULY 2025!

Register here for more info

30TH INTERNATIONAL POPULATION CONFERENCE

13–18 JULY 2025

BRISBANE CONVENTION & EXHIBITION CENTRE AUSTRALIA

Australian National University