

Population forecasting with Bayesian hierarchical models


<https://github.com/violetacaln/SIPP>

Violeta Calian, Ómar Harðarson



USE OF R IN OFFICIAL STATISTICS 2022
10TH INTERNATIONAL
VIRTUAL CONFERENCE
Hosted by Statistics Canada
December 6–8, 2022

Content

- Production of population projections: problems and solutions
- Difficulties and new demands
- New methods and implementation in  R
- Future plans

Problem and solution

for migration, fertility, mortality (by age, time & Z- characteristics)
which give, via ccpm & stochastically: the population predictions*

- **Problem:**

- *estimating and predicting* in the presence of:
 - small area/population issues and/or rare events/ shocks
 - expert assumptions
 - uncertainty of parameters and errors in data (e.g. *deregistration*)
- while reporting *uncertainty* of estimates/predictions

- **Solution must be based on:**

- Modeling and combining (prior) information

Types of models

Recent years: functional models, ARDL & bsts models $f(\text{age}, \text{time})$

- smoothing and orthogonal expansions for Fertility and Mortality components
- econometric / decay models for short / intermediate term
- time series / assumptions for long term

New: hierarchical and/or Bayesian models of response(s)

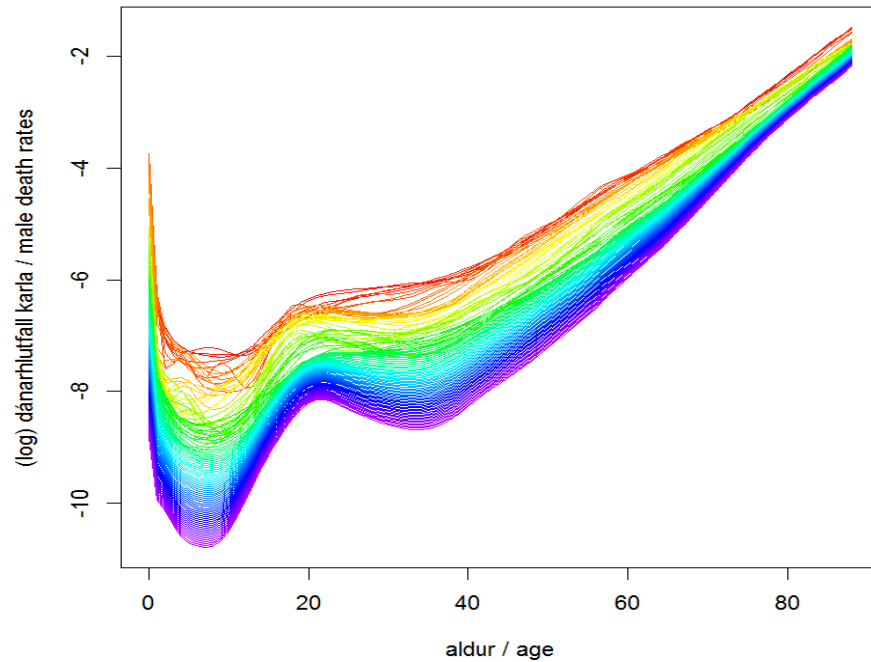
- smoothing/processes $(t, \text{space}, \text{age}, \dots)$
- complex correlation structures, spatial, demographic, social, ... characteristics
- incorporate quantitative & qualitative info: *priors*

****** *Model averaging options*

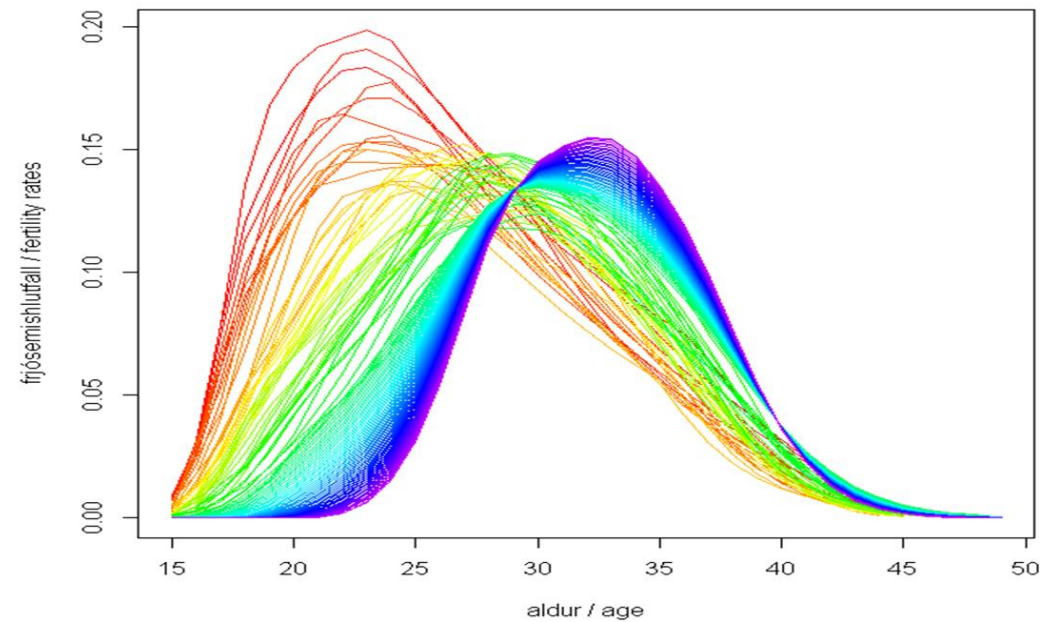
Typical result of functional models

<https://cran.r-project.org/web/packages/demography/index.html>

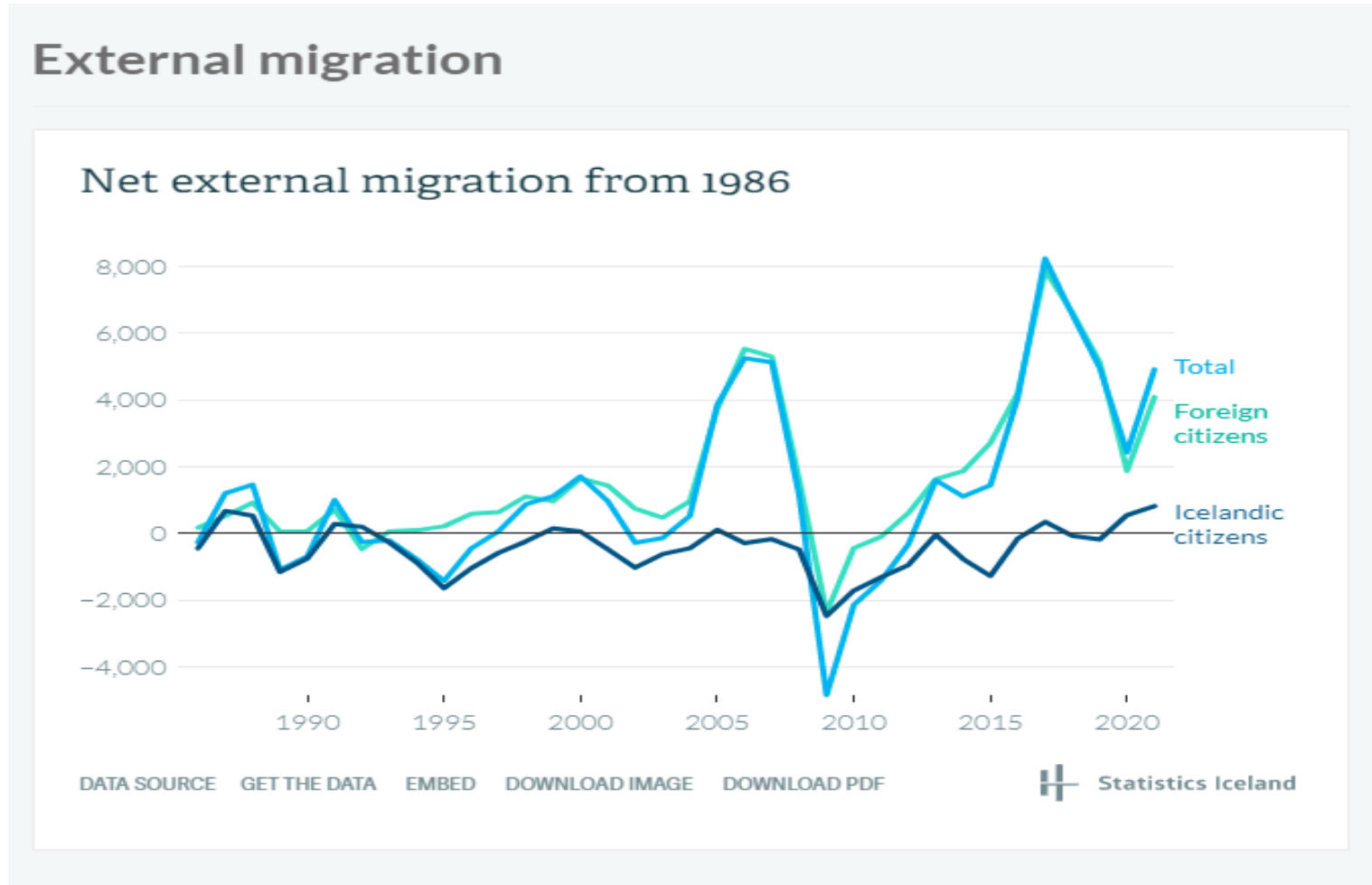
Mortality rates



Fertility rates



Migration issues and options for modeling



Predicting future values of migration/fertility/mortality:

- projections versus forecast
- advantages of Bayesian approach

Past

- Disjoint, long and short term migration predictions
- Unique models for predicting mortality and fertility, but *only* as functions of time, age (and gender)

New

*& allowing for **local** predictions*

- *Difficulties and new user demands* → Strengths of the new approach
- *Exceptions:* shocks (economical, political, natural, ...)

New Models (parametric or not!) & implementation in

$P(Y/a) = F1(X)$ where $P(a/b) = F2(c, Z)$ where...

- **Type: Hierarchical / Multilevel**, Bayesian / frequentist
 - data used: *individual* or *aggregated* response (Y) versions
 - what: **processes**, since *time series* data $P(Y_i, \dots, Y_n | a)$, versus $P(Y_i)$: account for (auto-) correlations
- **Why:**
 - {components/prior information/characteristics/local&total} & **uncertainty** report
- **R-Tools:**

several *Stan*-running R packages like *brms*, *arm*, *bayestestR*, ...

and the faster: *lme4*, *mgcv*, ... for frequentist / as initializations / not only (see Laplace approx of *mgcv*)

Structure, novelty and advantages

$x=(\text{year, age, gender, citizenship, municipality, education, ...})$

- V1. Counts/rates

- $N(x, \dots) \sim \text{Poisson}(e^{f(x, \dots)} N_0(x, \dots))$ (or: neg. binom, over-disperssion issues, etc)
- $f_m(\{x, \dots\}) \sim \text{GP}(m(\{x, \dots\}), C(\{x, \dots\}, \{x', \dots\})) + \dots$

- V2. Microdata

- $N(x, \dots) \sim B(e^{f(x, \dots)} N_0(x, \dots))$
- $f_m(\{x, \dots\}) \sim \text{GP}(m(\{x, \dots\}), C(\{x, \dots\}, \{x', \dots\})) + \dots$
- f_m – denotes here: sums of smooths/re/t2 and/or by {grouping factors} OR unknown functions with Gaussian Processes as priors (with m -> long term behaviour, C -> correl between any two response values)

Tests, Correlations, Dimensions

- Variability

- of:

fertility, mortality, migration

- through (and with auto-correl):

Time & age (*s(t)*, *gp(t, age, ...)*, *t2(t, age)*, ...)

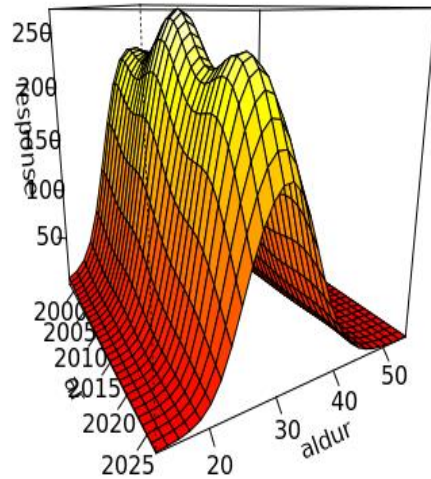
- by:

- **Municipality**, Region – significant for: migration but not/less for mortality, fertility
- Other characteristics: education, family related, municipality attributes, ...
- More grouping characteristics (ut/is, im/e, f/m)

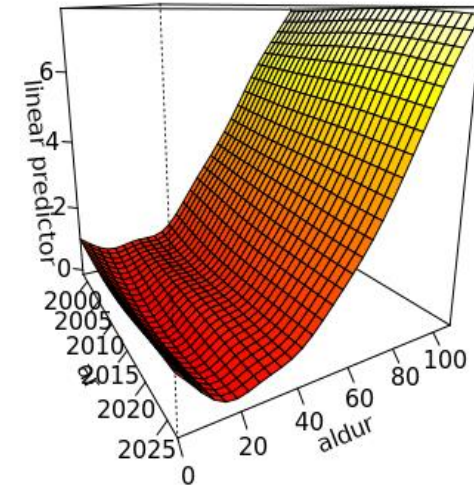
V1. Count data input

<https://cran.r-project.org/web/packages/mgcv/index.html>

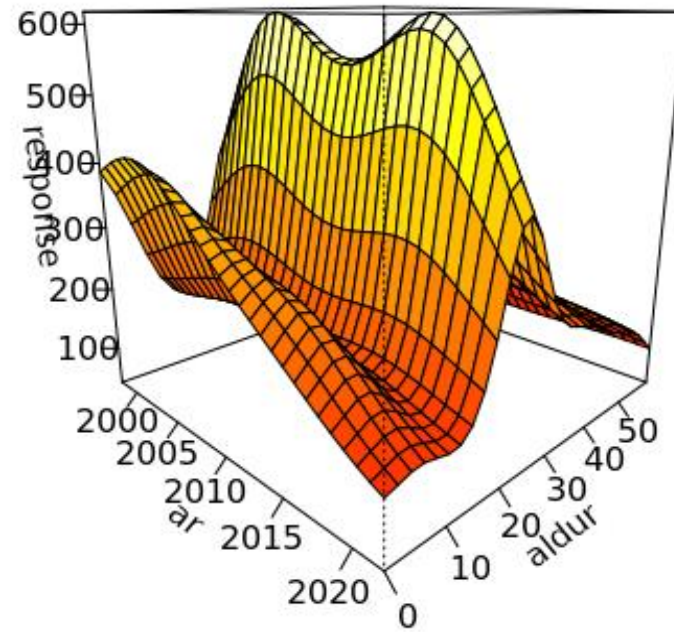
Births



Death rates (log)

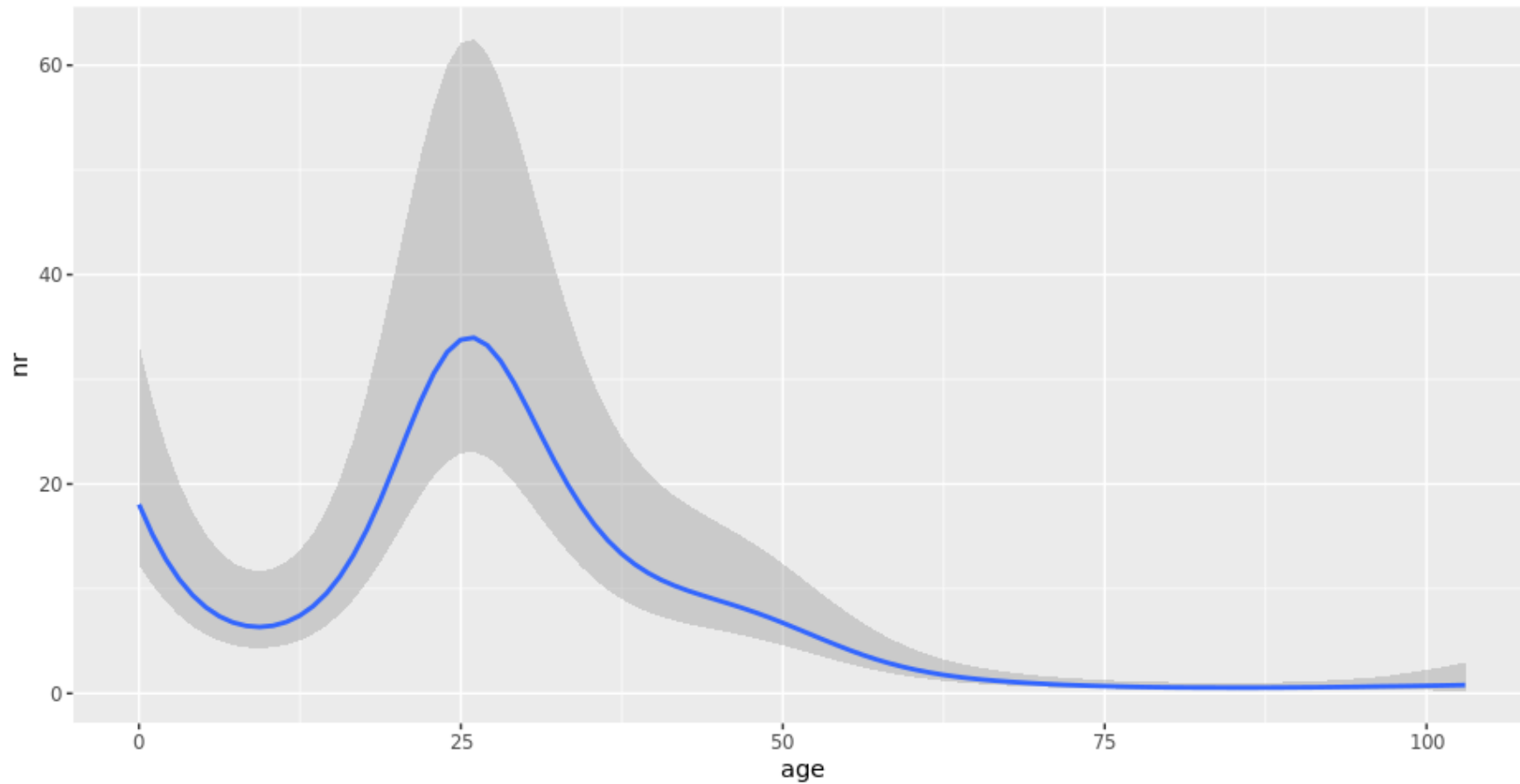


Migration

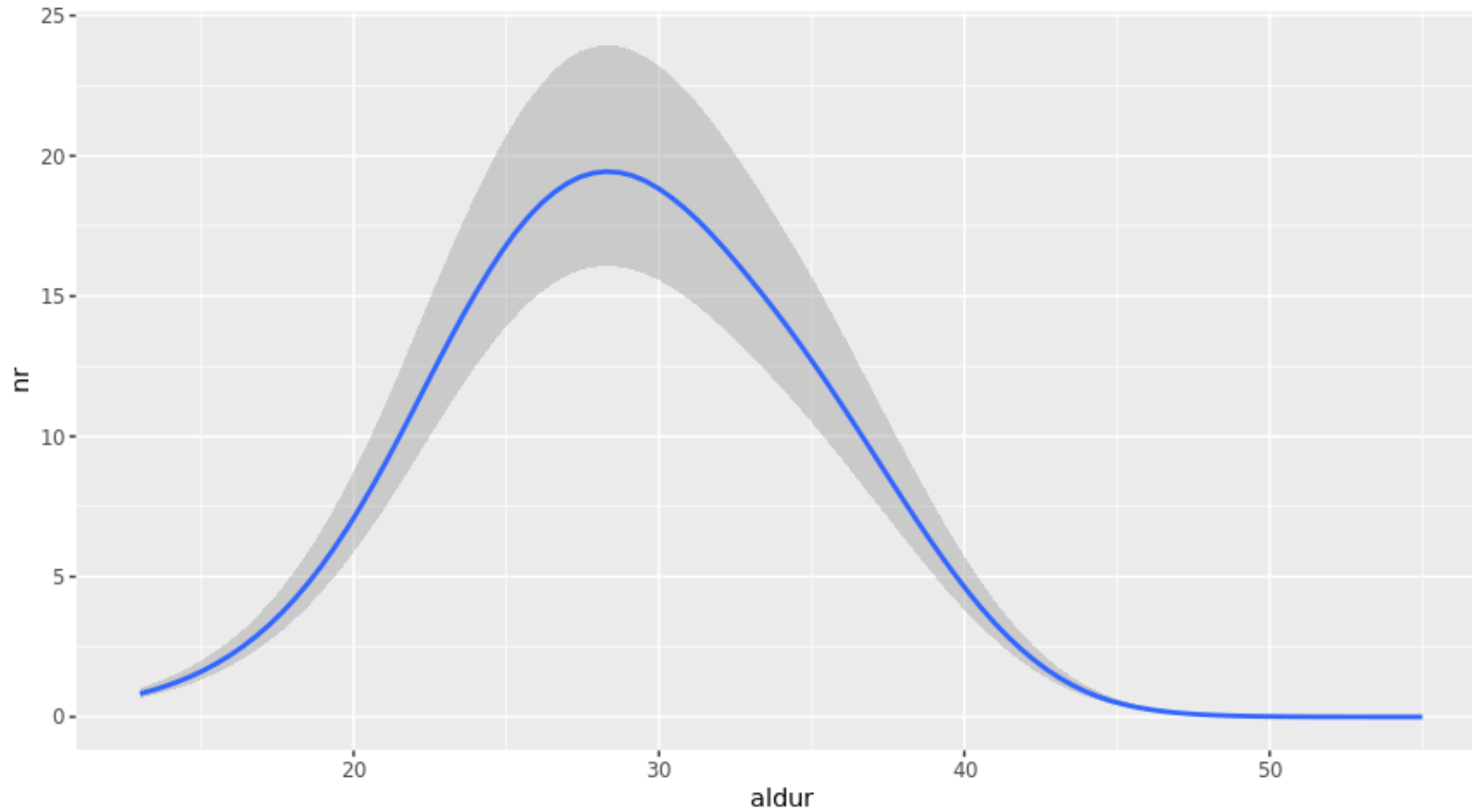


Counts/rates models: all migration categories (f_{em_ut})

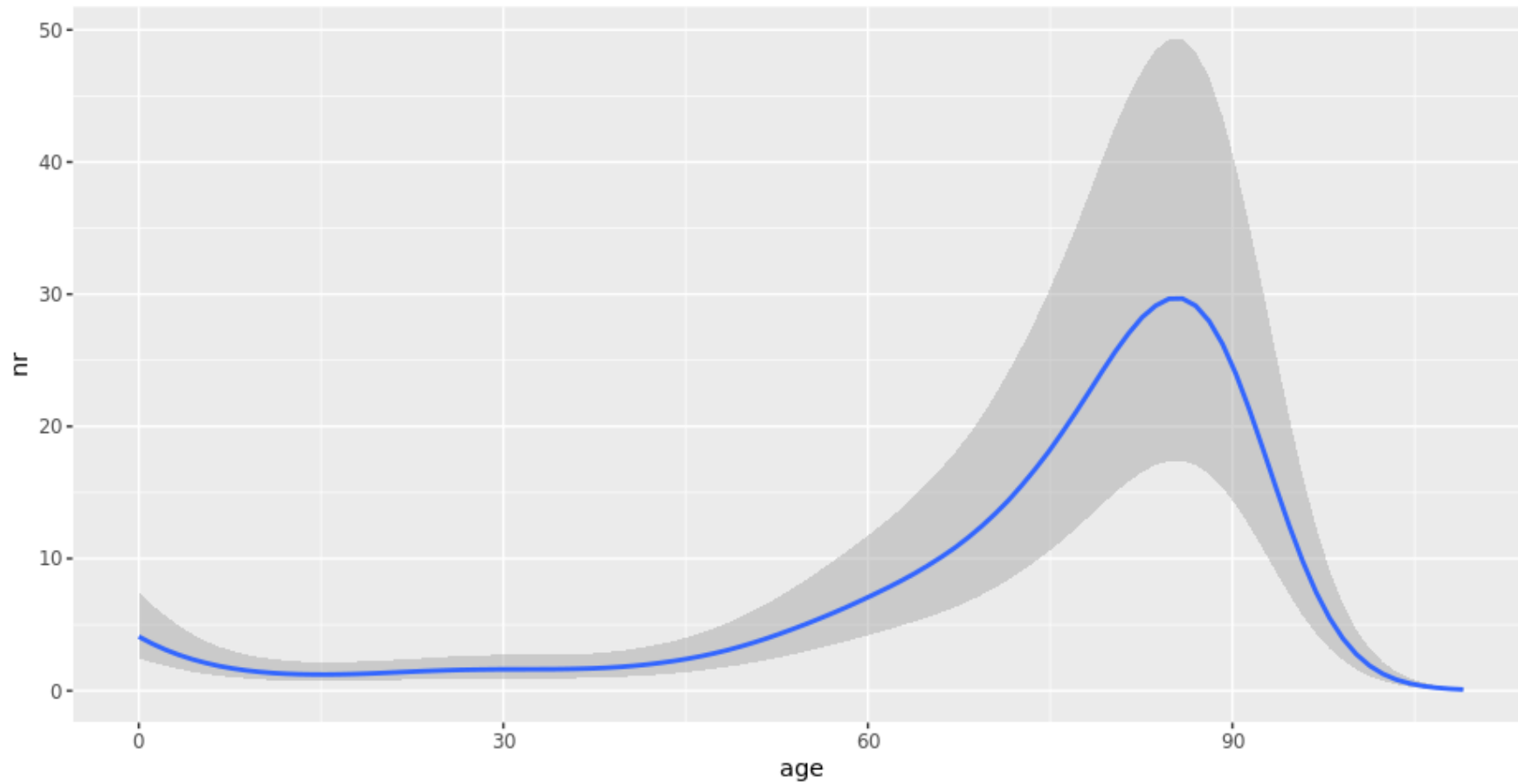
<https://cran.r-project.org/web/packages/brms/index.html>



Counts/rates: fertility models



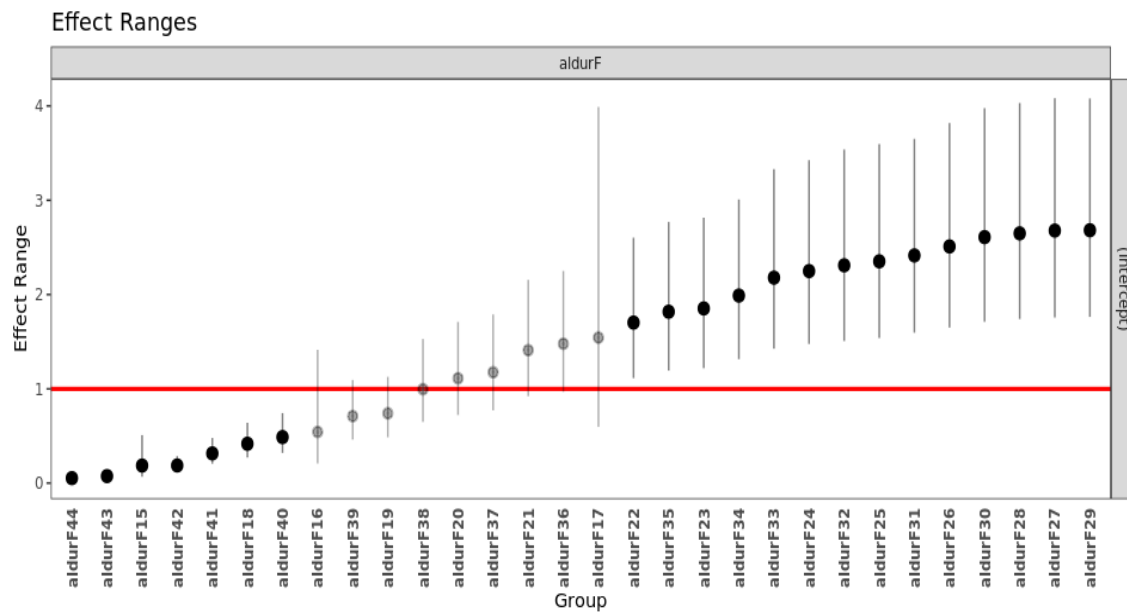
Counts/rates: mortality models (addit, s)



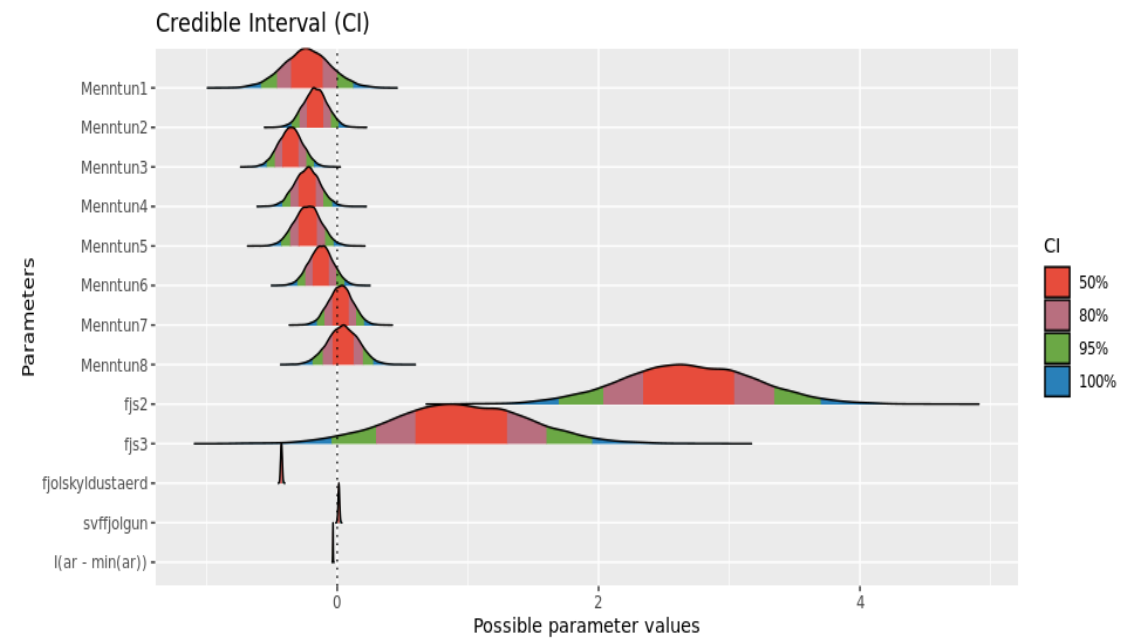
V2. Microdata input

Example 1. Significance of effects for:

Age group effect on fertility (o.r. units)



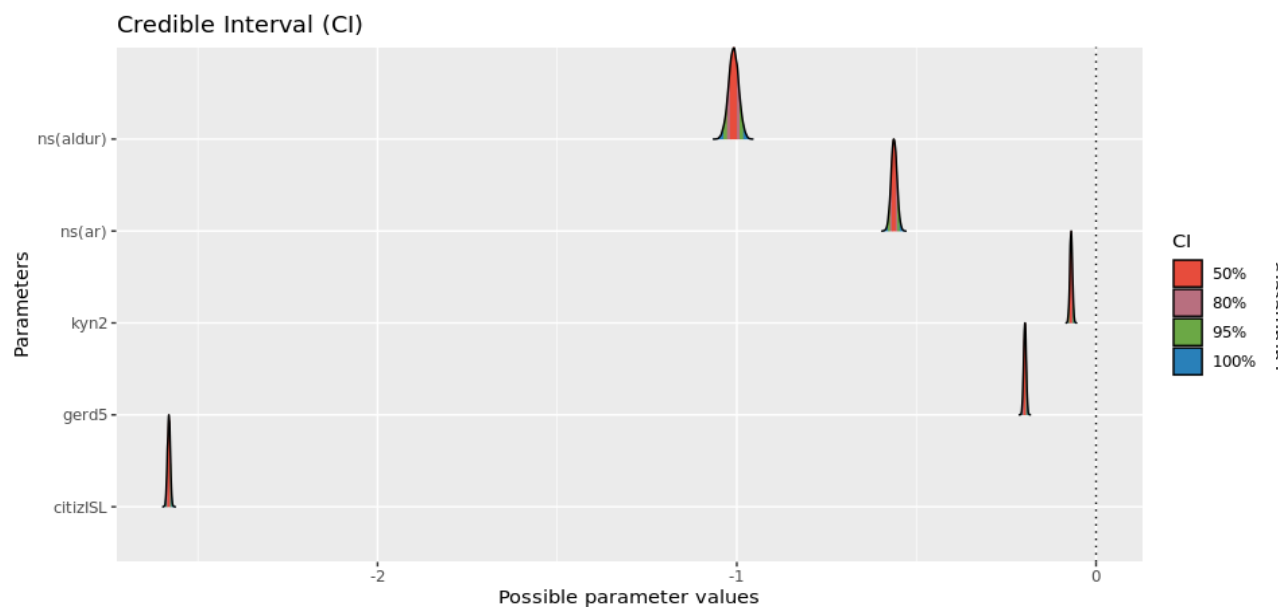
Characteristics' effects on fertility



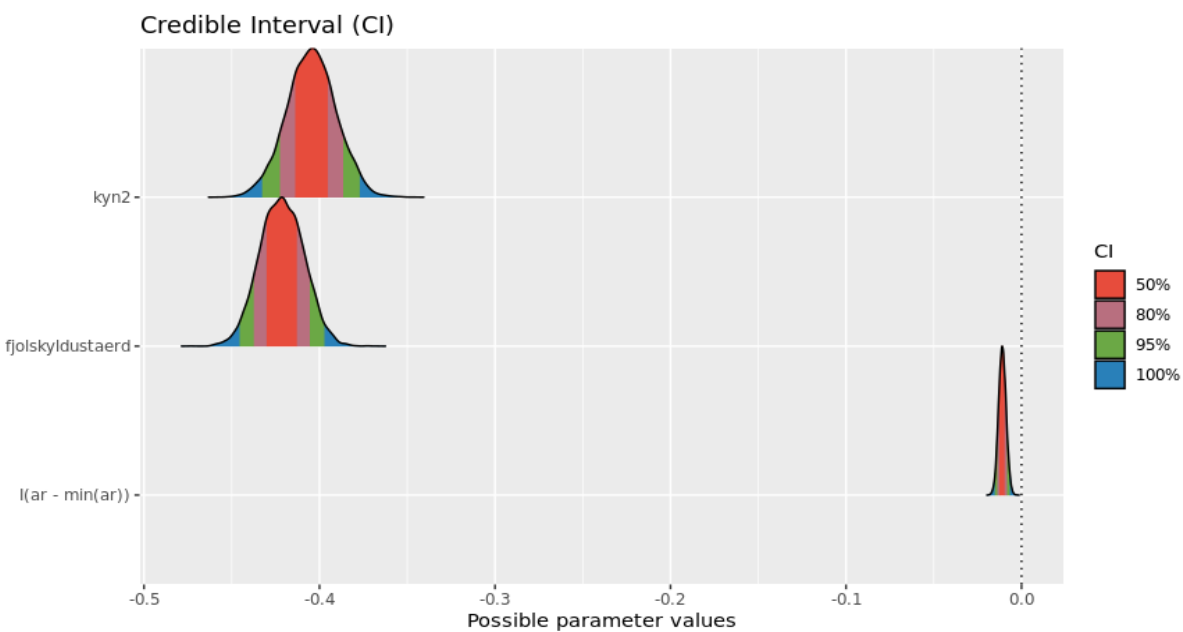
Microdata input

Example 2. Characteristics' effects on:

Migration

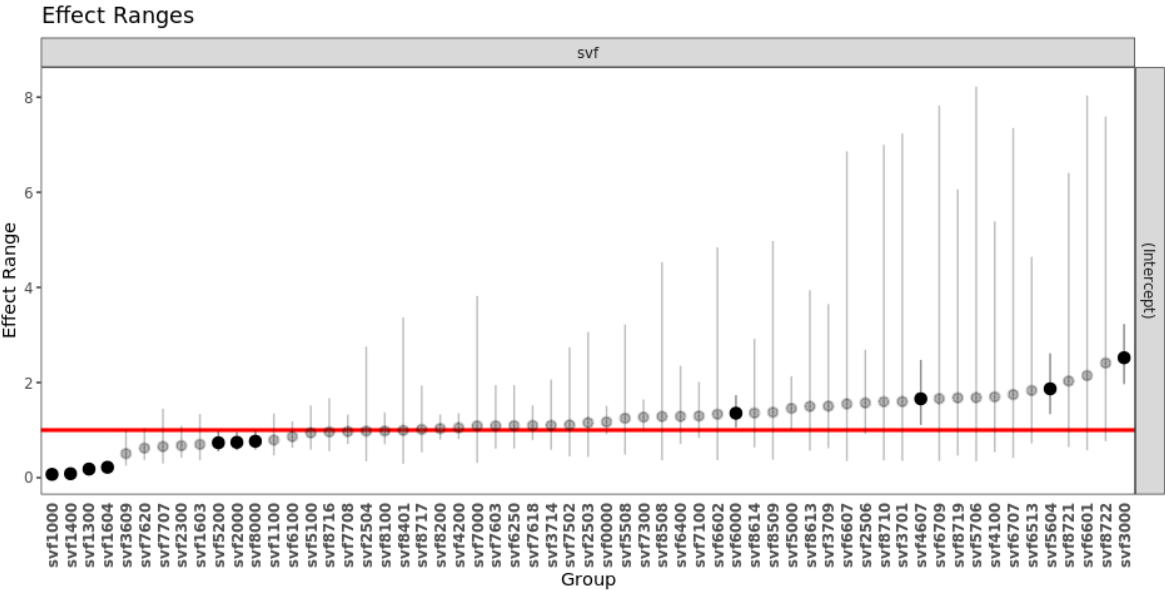


Mortality

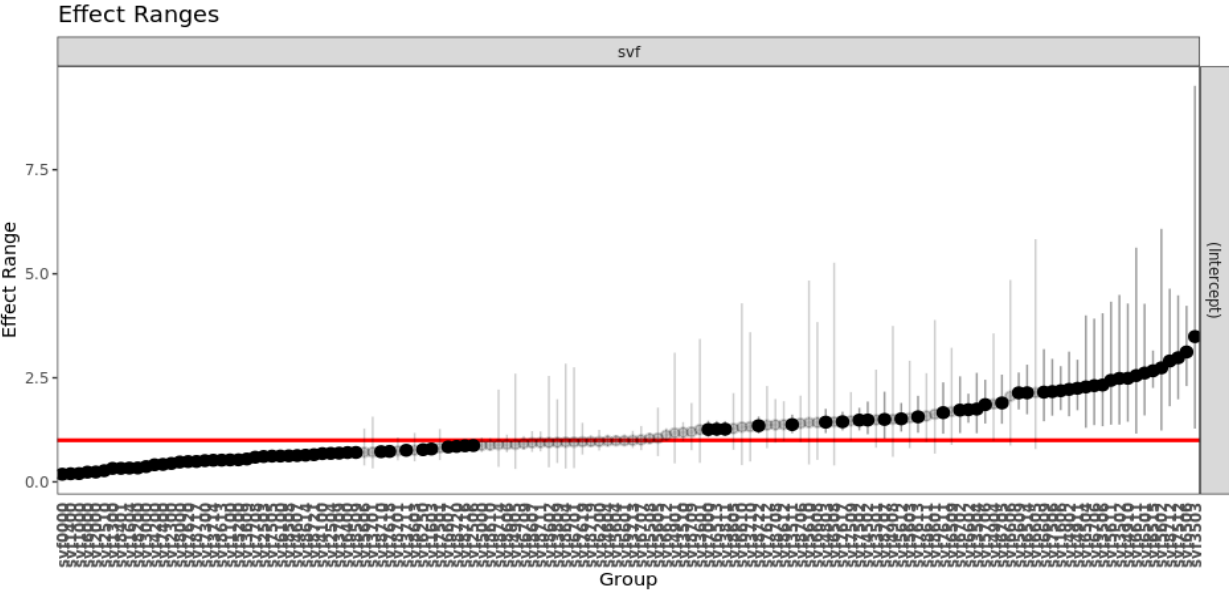


One more example: testing for location effects
(although not with *mrf* yet!)

municipality - effect on fertility



municipality – effect on migration



Future/work in progress/*

- finalise stochastic *population* predictions by combining high number of values sampled from the posterior distributions of the demographic components (migration, fertility, mortality) as in *ccpm*
- finalise the model *testing*
- forecasting with multiple *model averaging**

References

- Rob J. Hyndman, Heather Booth (2008). Stochastic population forecasts using functional data models for mortality, fertility and migration, *International Journal of Forecasting*, Volume 24, Issue 3, <https://doi.org/10.1016/j.ijforecast.2008.02.009>
- <https://cran.r-project.org/web/packages/bsts/index.html>
- Wood S (2017). *Generalized Additive Models: An Introduction with R*, 2 edition. Chapman and Hall/CRC.
- <https://cran.r-project.org/web/packages/mgcv/index.html>
- Bürkner P (2017). “brms: An R Package for Bayesian Multilevel Models Using Stan.” *Journal of Statistical Software*, **80**(1), 1–28. doi: [10.18637/jss.v080.i01](https://doi.org/10.18637/jss.v080.i01)
- Bürkner P (2018). “Advanced Bayesian Multilevel Modeling with the R Package brms.” *The R Journal*, **10**(1), 395–411. doi: [10.32614/RJ-2018-017](https://doi.org/10.32614/RJ-2018-017)
- Bates D, Mächler M, Bolker B, Walker S (2015). “Fitting Linear Mixed-Effects Models Using lme4.” *Journal of Statistical Software*, **67**(1), 1–48. doi: [10.18637/jss.v067.i01](https://doi.org/10.18637/jss.v067.i01)
- Zhang, Junni & Bryant, John (2020). Bayesian Disaggregated Forecasts: Internal Migration in Iceland. 10.1007/978-3-030-42472-5_10.
- Bryant, J., & Zhang, J. L. (2016). BAYESIAN FORECASTING OF DEMOGRAPHIC RATES FOR SMALL AREAS: EMIGRATION RATES BY AGE, SEX, AND REGION IN NEW ZEALAND, 2014-2038. *Statistica Sinica*, 26(4), 1337–1363. <http://www.jstor.org/stable/44114337>
- Lam, K.K.; Wang, B. (2021). Robust Non-Parametric Mortality and Fertility Modelling and Forecasting: Gaussian Process Regression Approaches. *Forecasting* 2021, 3, 207-227. <https://doi.org/10.3390/forecast3010013>

Thank you!