

Population projections when using time series with extreme values

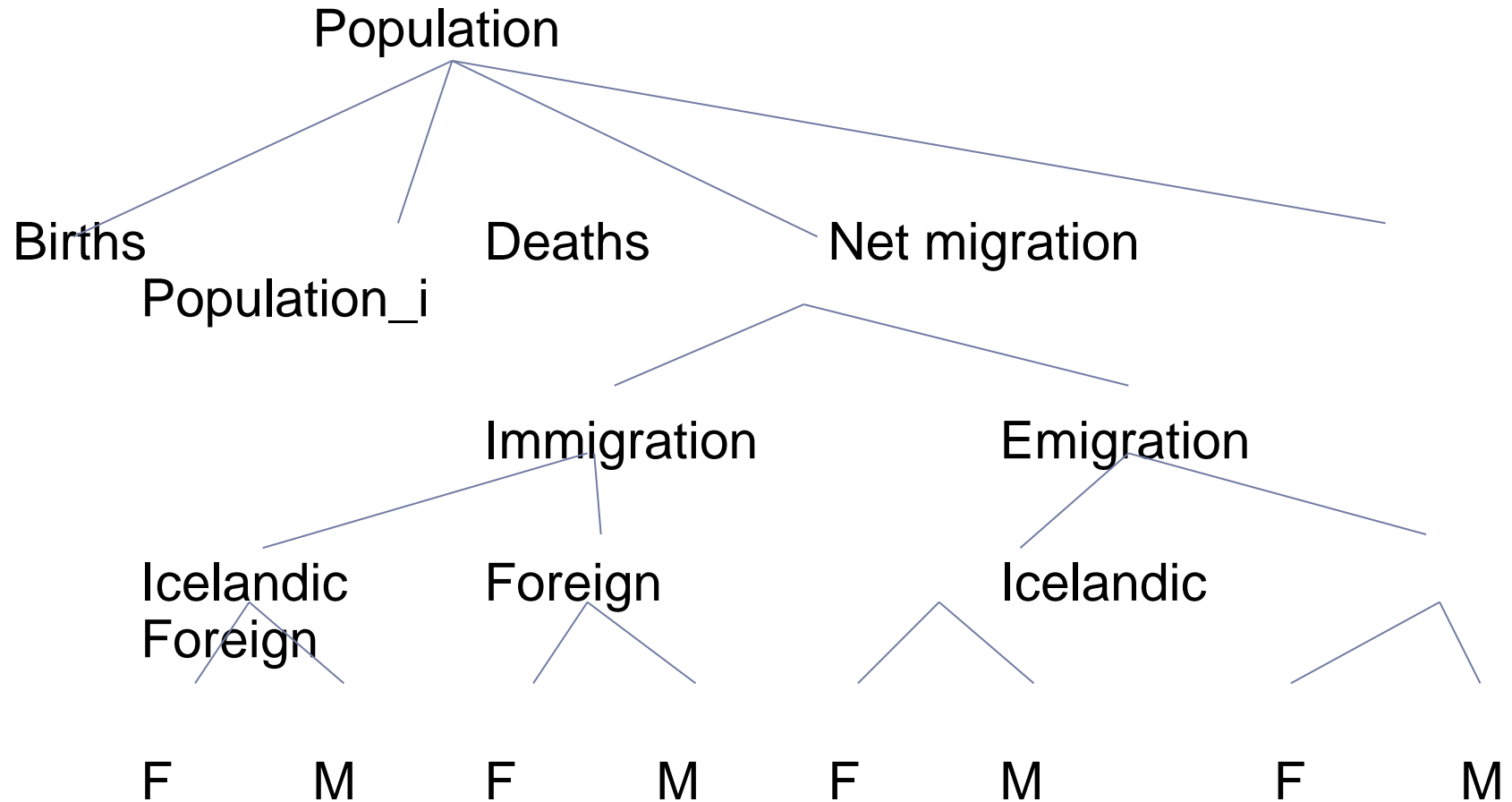
Violeta Calian, Statistics Iceland

Joint Eurostat/UNECE Work Session on Demographic Projections
Geneva, Switzerland, 18-20 April 2016

Outline

- ▶ Data and formulation of the mathematical problem
- ▶ Hypotheses testing and probability of rare events for time series: outliers and extreme values.
- ▶ Methods and models:
 - ▶ Vector dynamical (ARDL) models, vector arima models
 - ▶ functional data modelling with time series coefficient functions of orthonormal function expansions

Hierarchical time series data structure



The mathematical problem

Notation $y(t, x)$:

- the log of the observed mortality or fertility rate for age x and year t
- migration components by age, at time t
- a vector of several/all migration components (y_1, y_2, \dots)

Observations $\{t_i, x_a, ; y_{ia}\}$

$i = 1, \dots, N$ (years) and $a = M_0, \dots, M$ (ages),

at discrete points (t_i, x_a) of a two-dimensional domain

Predictions $y(t_i, x_a)$ with prediction intervals/densities

for same set of age values and for years t_i ($i = N + 1, \dots, N + h$),
where h = the length of the forecasting horizon

Principles

- *joint **modeling*** the (non-stationary) vector time series data (and its autocorrelations) so that the residuals are independently, identically and preferably normally distributed
- using (multiple) **hypotheses testing** (ergodicity, stationarity, outliers)
- scale independence
- data exchangeability when needed
- explaining and predicting
- data models depending on goal
- **$g(x, y) = 0$** $\rightarrow y = f(x)$ when it exists!

Outliers and extreme values

- Outlier

small probability of being generated by the same statistical distribution / stochastic process as the other observations in the data set

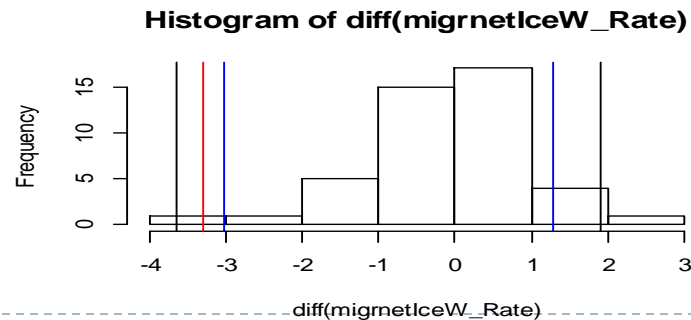
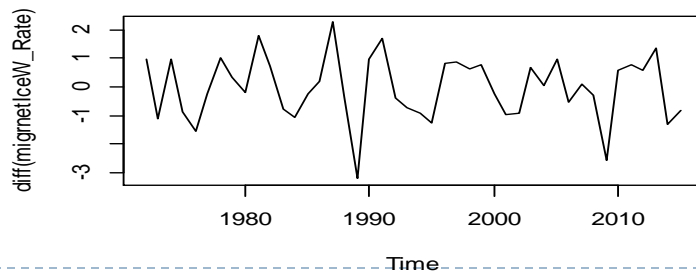
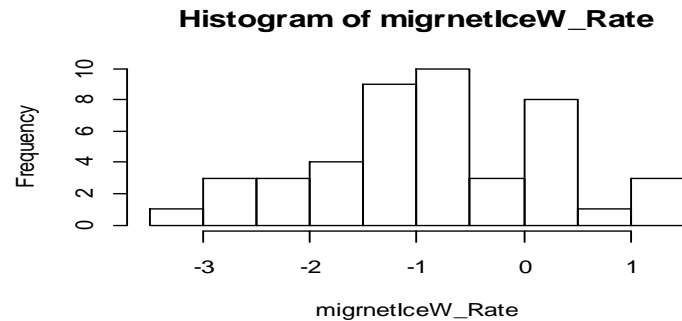
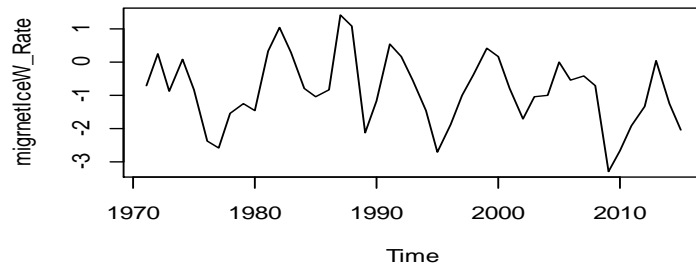
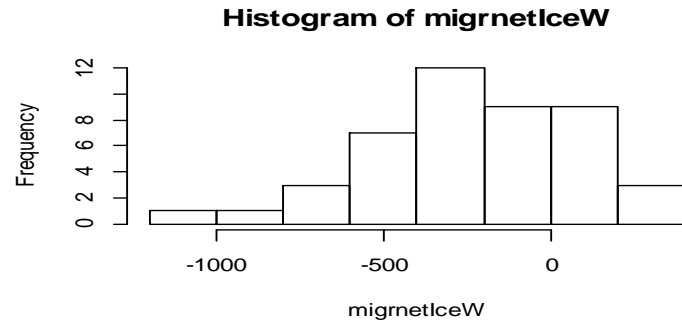
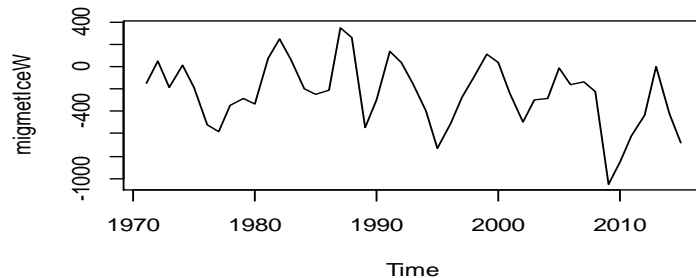
- Time scales of outlier **effects** in time series data

- Extreme value

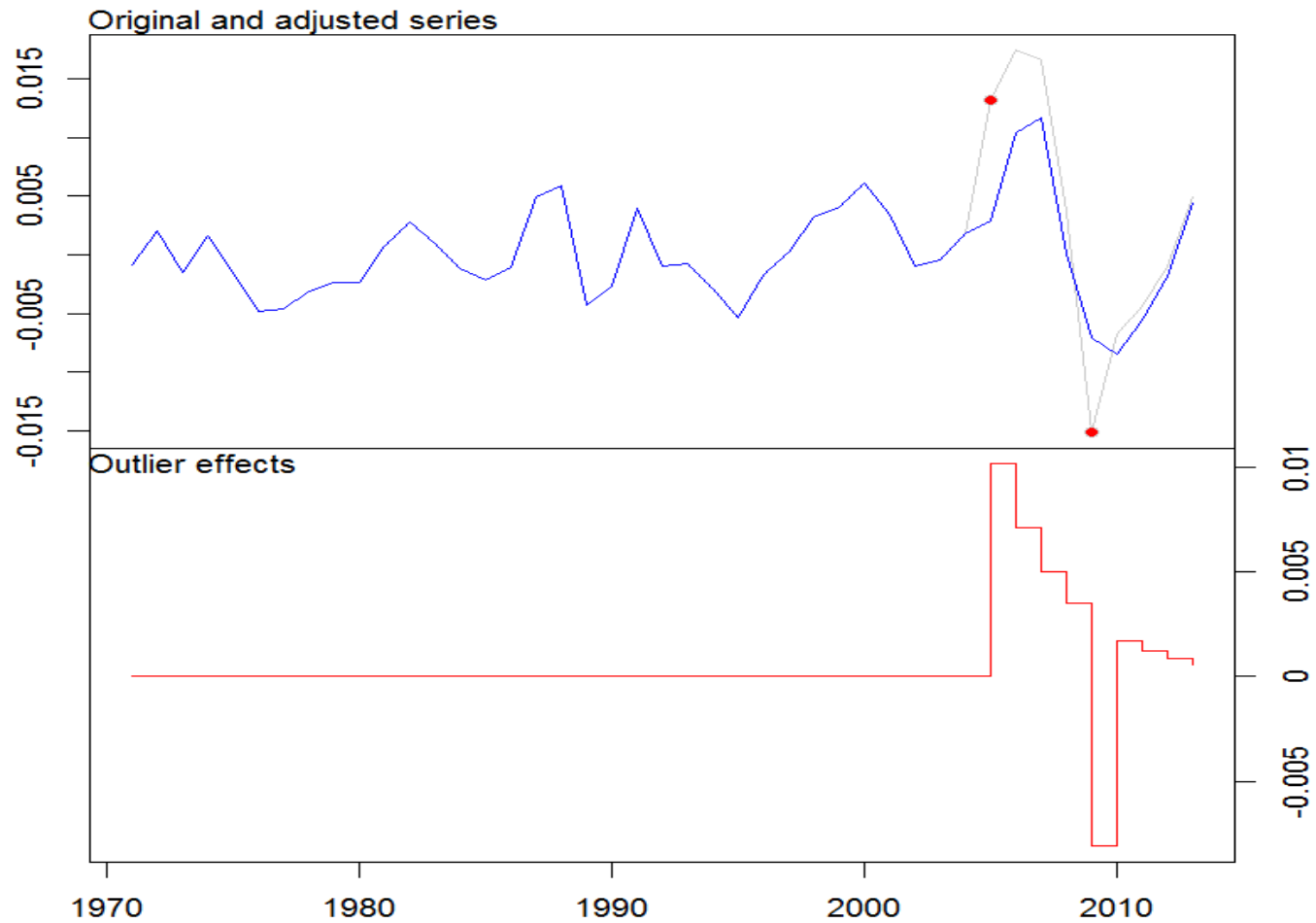
observation with low probability of occurrence but generated from the same statistical distribution / stochastic process as the other observations in the data set

- **Predicting** extreme values occurrence for time series data

Example: Outliers and extreme values in migration data



Net migration rate



Model $f(t, x)$ types

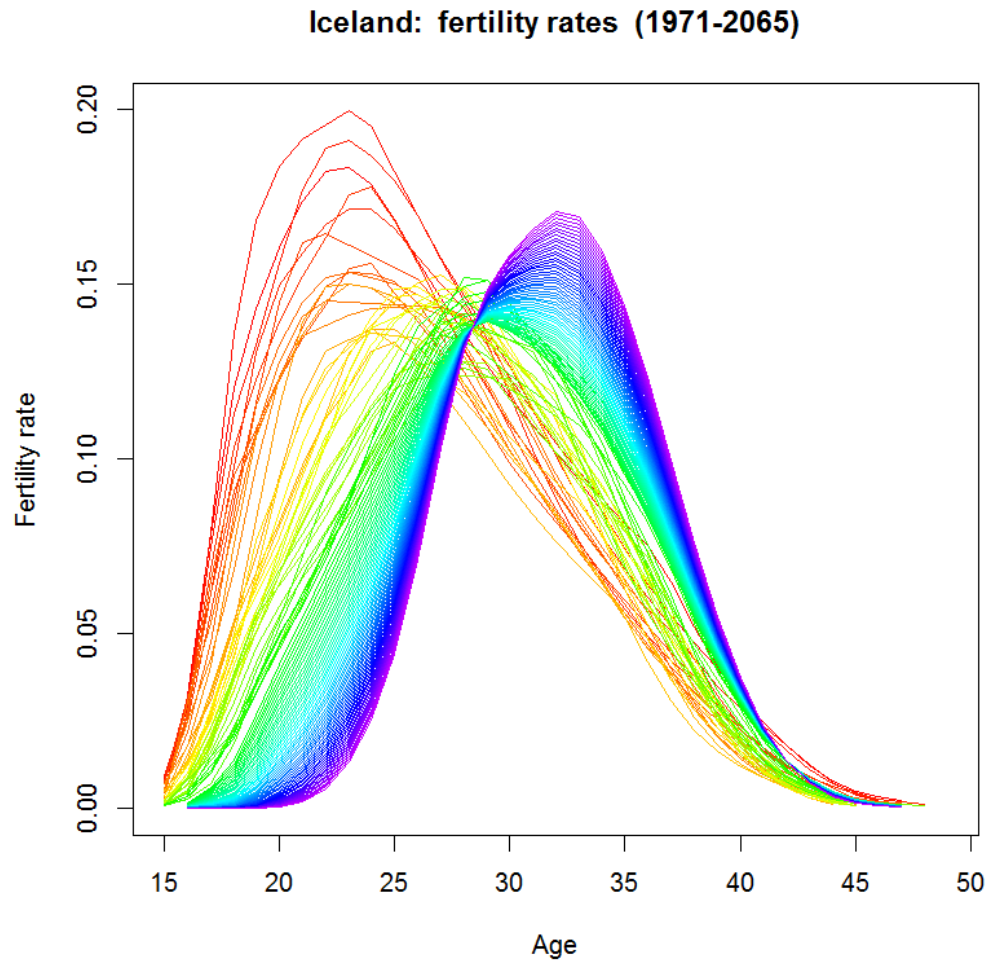
- ▶ a model $f(t, x)$ of time and age dependent rates (trick: which can be written as an expansion with time-series coefficient functions which are independent by construction)*
- ▶ a multivariate model $(f_{x_{M_0}}(t), \dots, f_{x_M}(t))$ of time dependent rates
- ▶ a multivariate model $(f_{t_1}(x), \dots, f_{t_N}(x))$ of age dependent rates
- ▶ a (multivariate-) model $f(t)$ for $\sum_x f(t, x)$ of time dependent (vector of-) rates for the sums over all ages (trick: when age distribution stable in time, disaggregation is possible) as functions of exogenous variables Z_1, \dots and their lags.*
- ▶ a multivariate model of „everything“ $(f_1, f_2, \dots, Z_1, \dots)$ as functions of (t, x) ☺

Models for fertility and mortality rates

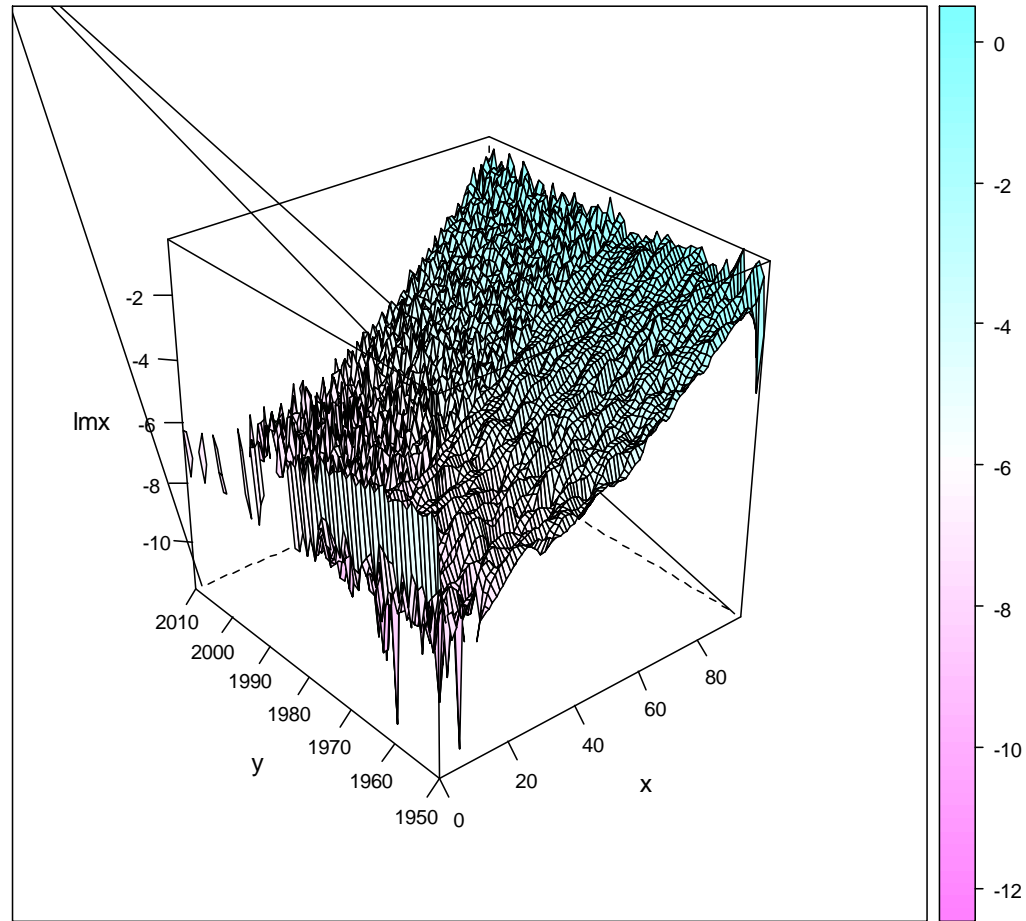
Functional data approach – Hyndman et al (2007, 2008)

- ▶ $y_{ia} = \mu(x_a) + f_0(t_i, x_a) + e_{t_i}(x_a) + \alpha_{t_i}(x_a)\varepsilon_{t_i, x_a}$
- ▶ $\mu(x)$ - the mean of $f_0(t, x)$ across time (years)
- ▶ $e_t(x)$ is the residual modelling error (assumed serially uncorrelated)
- ▶ the coefficient functions $\beta_k(t)$ of the expansion $f_0(t, x) = \sum_{k=1, \dots, K} \beta_k(t) \varphi_k(x)$ are independent (by construction)
- ▶ $\varepsilon_{t, x}$ - the random variation in birth or death rates
- ▶ $\alpha_t(x)$ allow the variance to change with age and time.

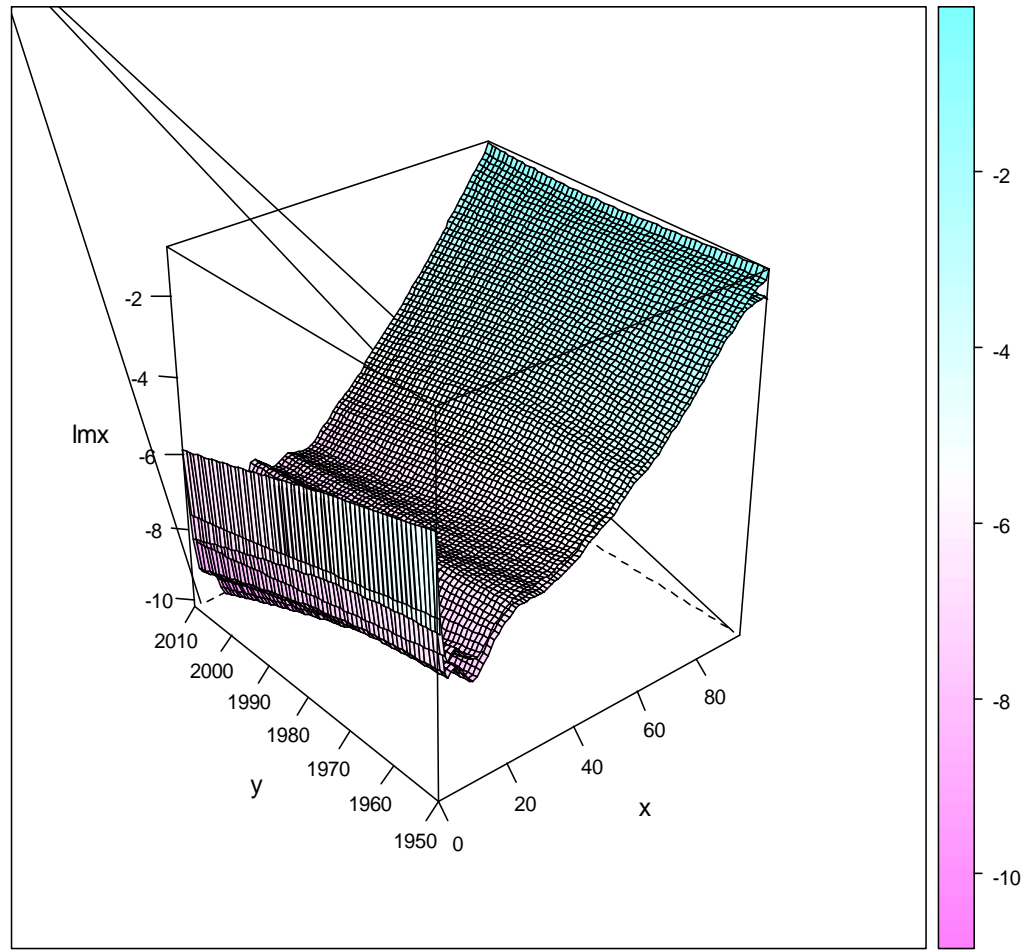
Fertility



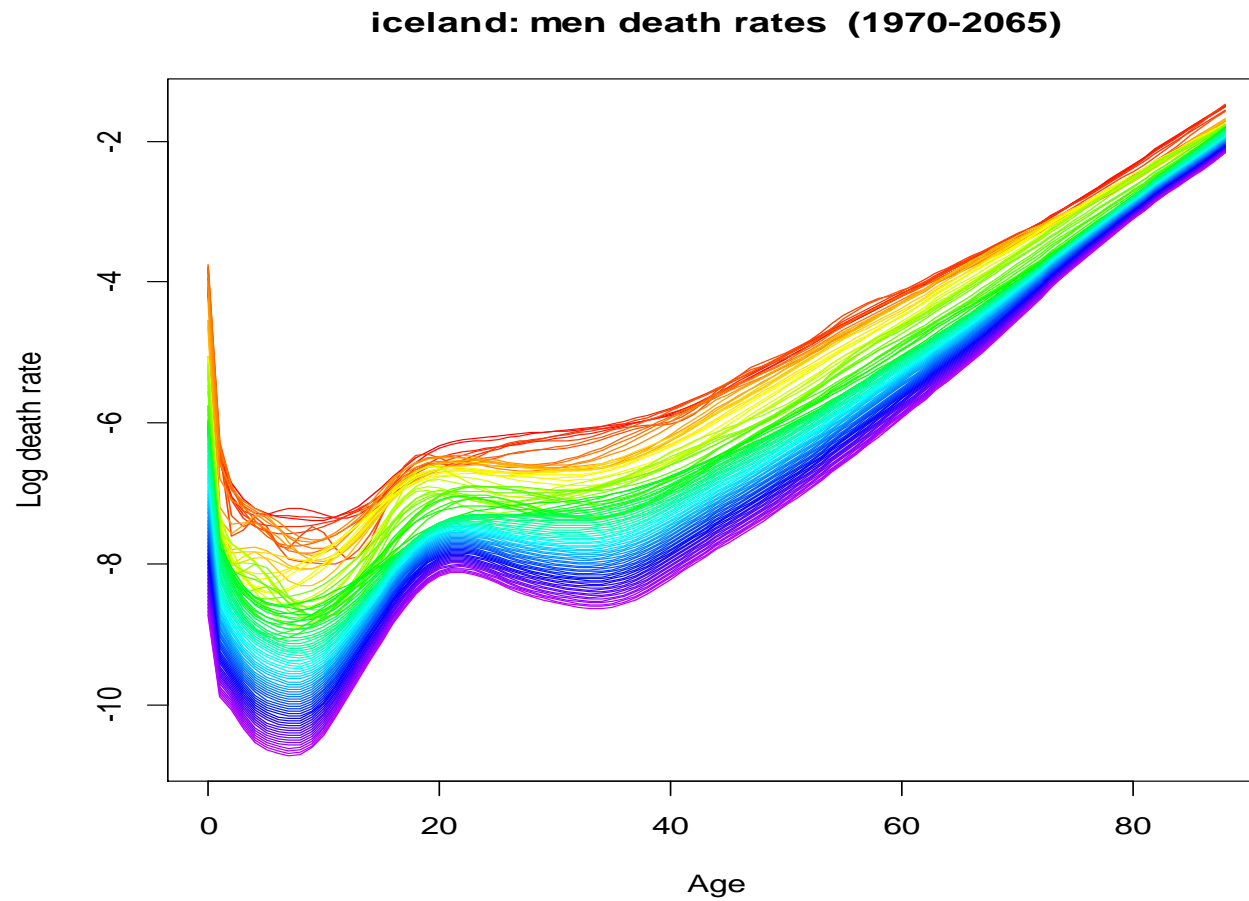
Raw data example and why extreme values matter



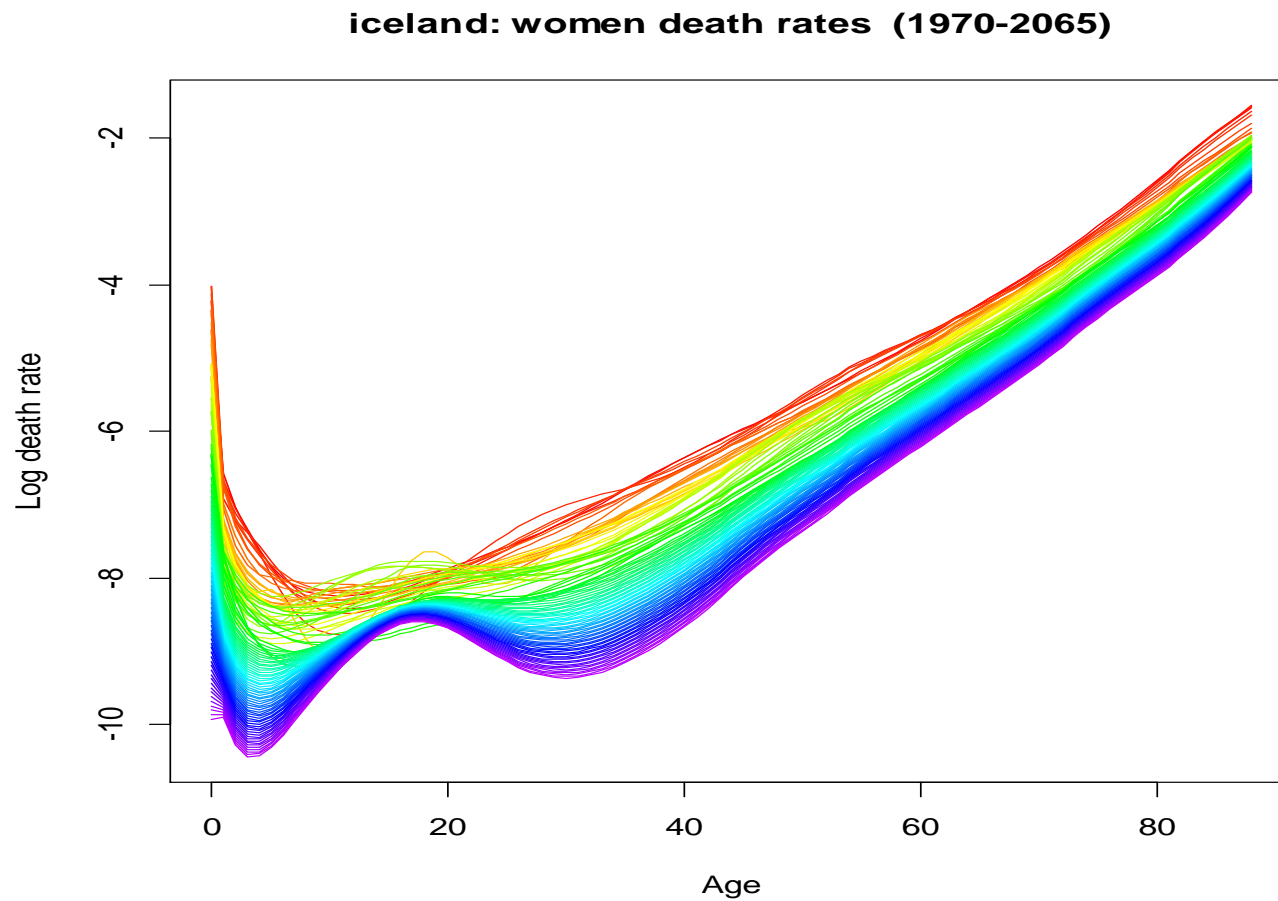
Smooth age-time mortality surface



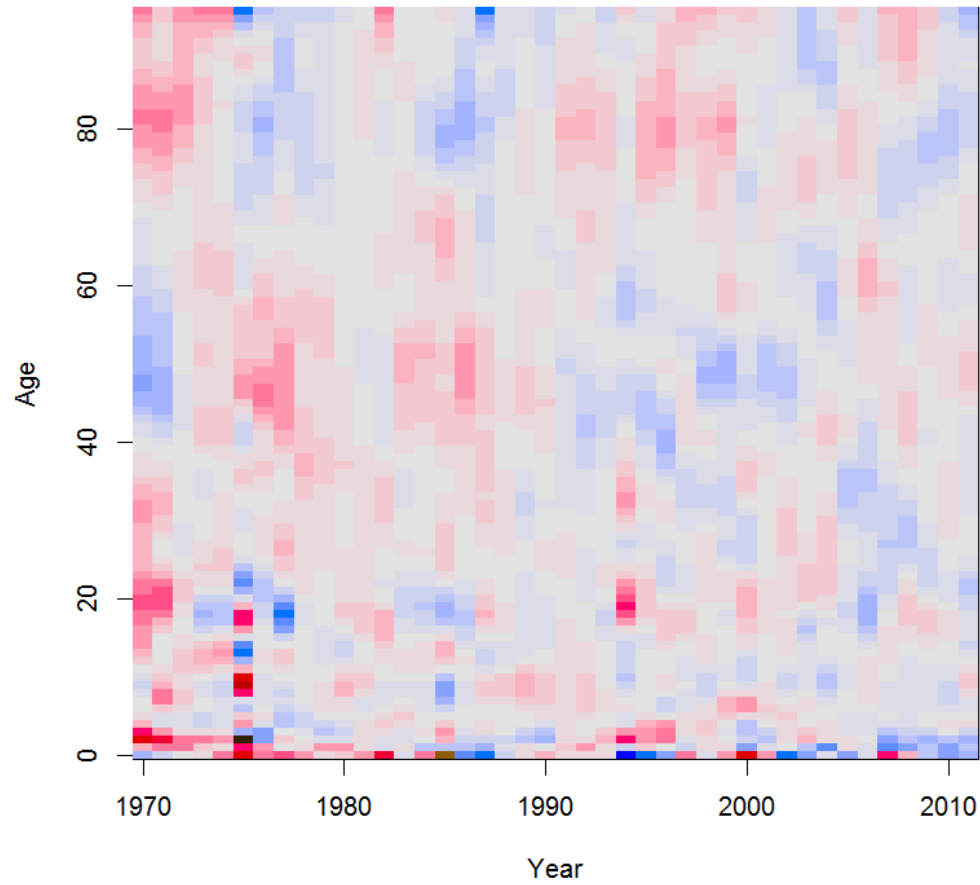
Mortality rates



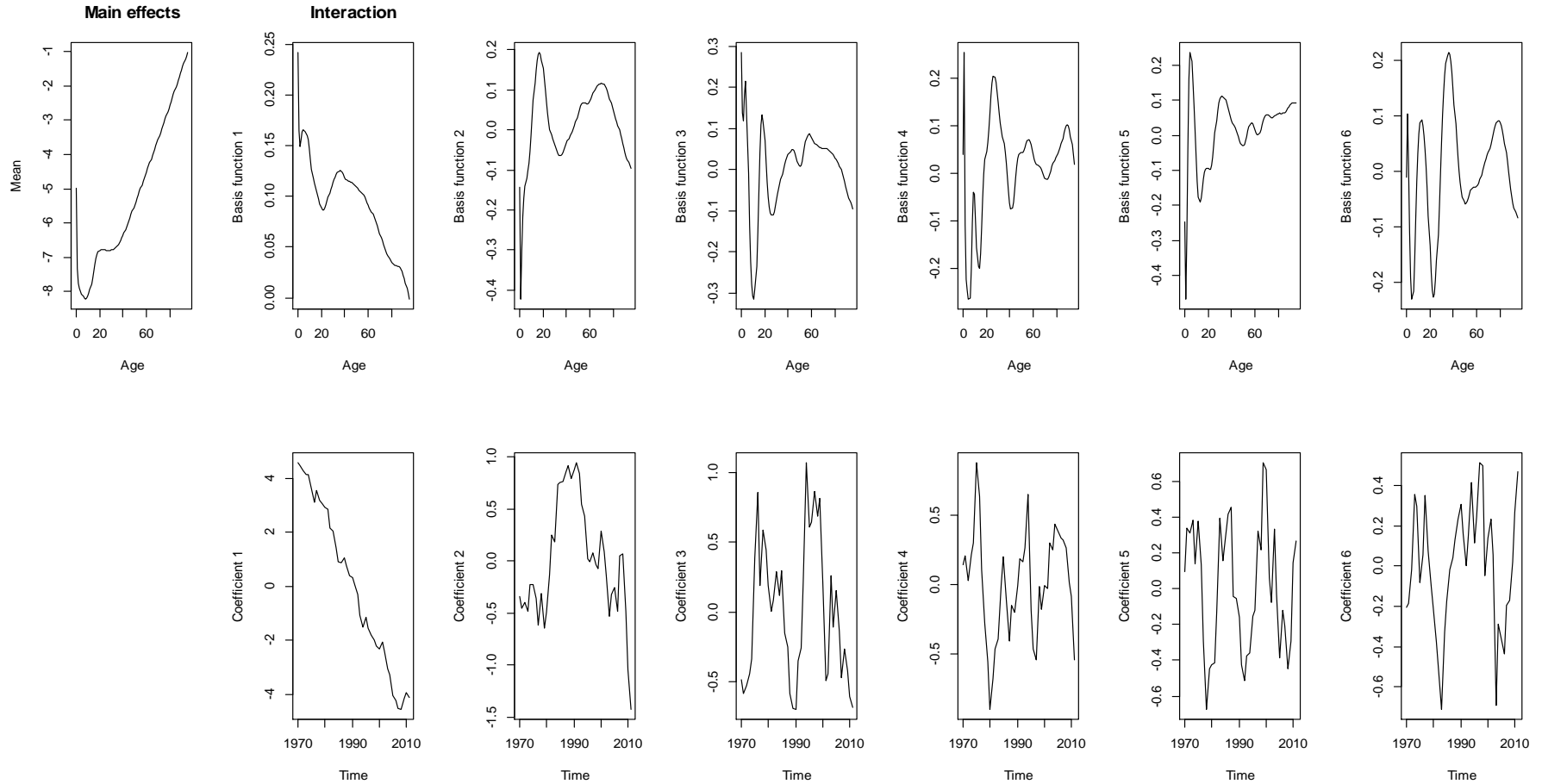
Mortality rates



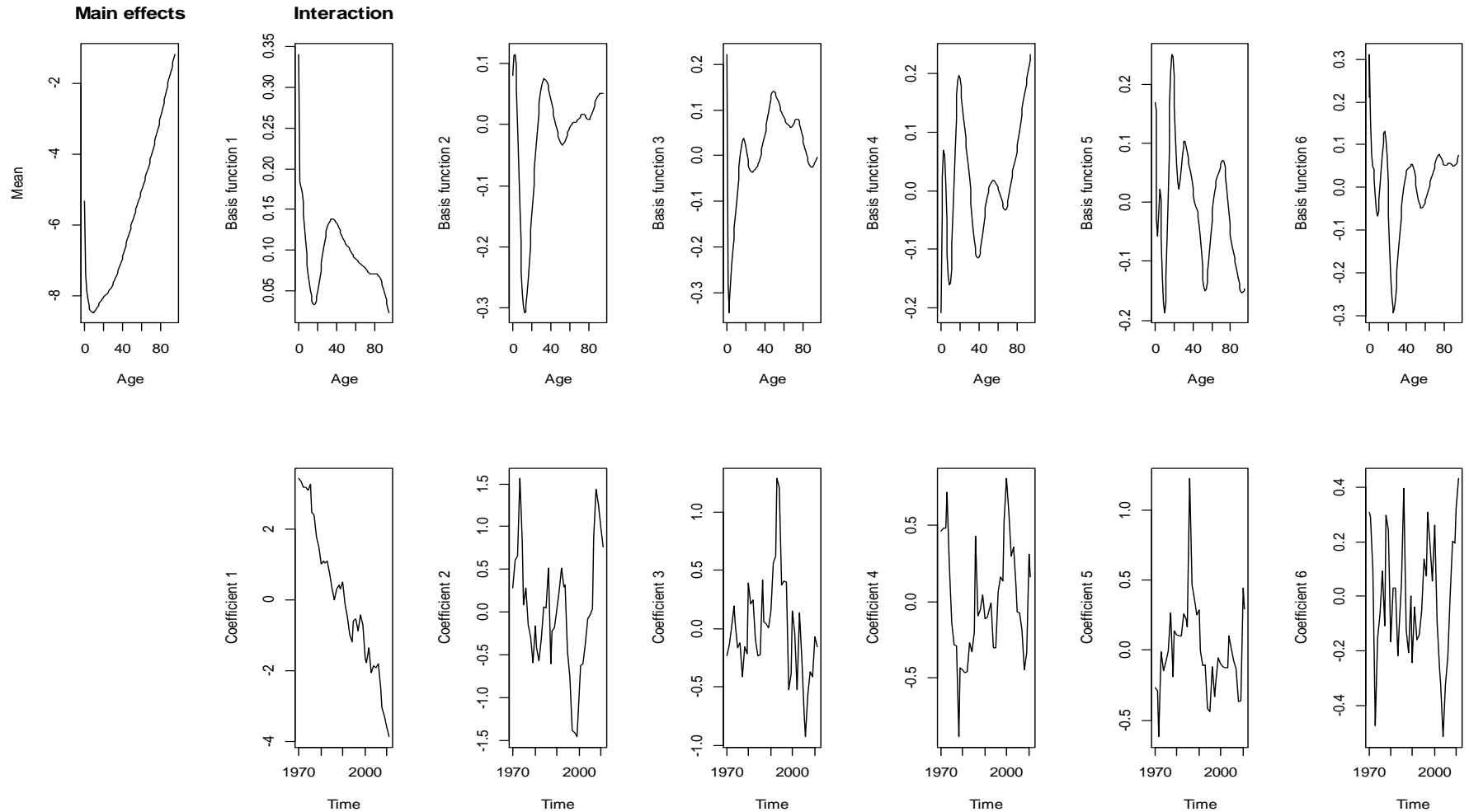
Residuals of mortality model



Functions of mortality model - men



Functions of mortality model - women



Migration models

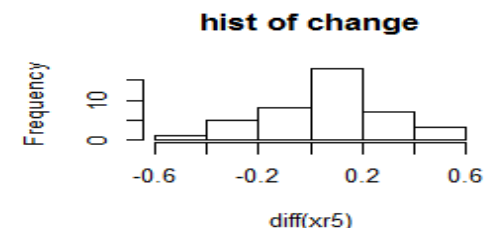
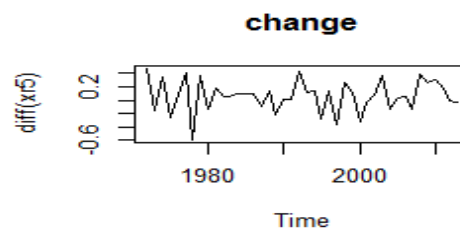
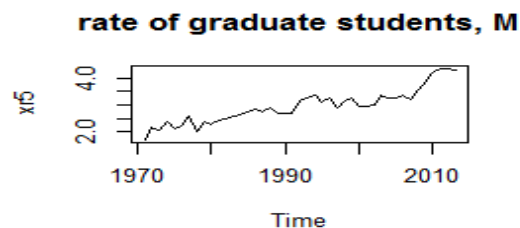
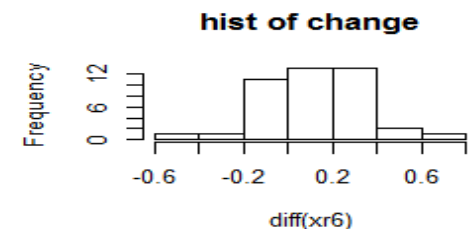
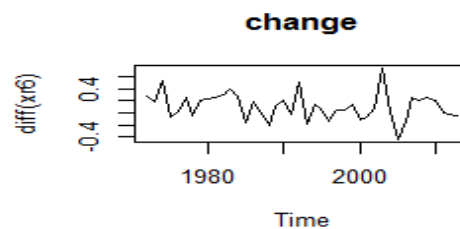
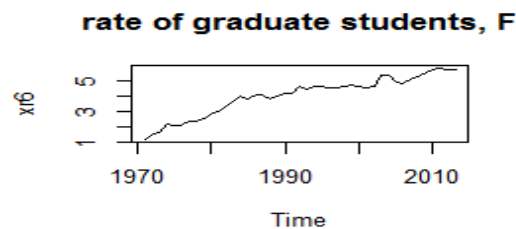
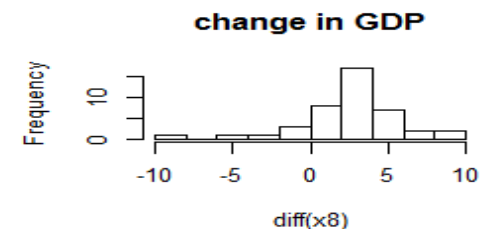
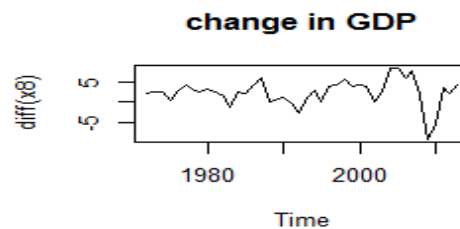
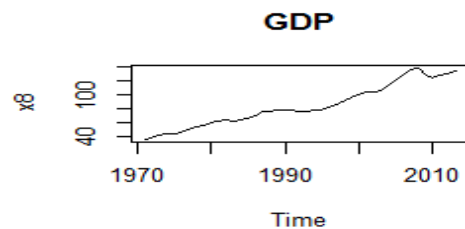
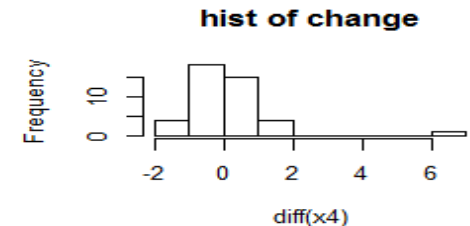
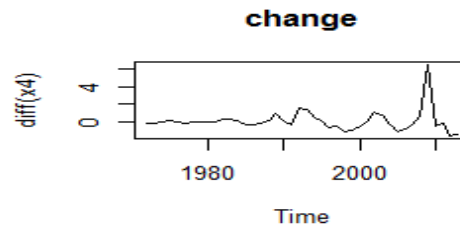
Vector ARDL for short term and *vector ARIMA* for long term, i.e. *(S)VAR*

Vector ARDL models $y(t) \sim y(t_k) + x(t_p)$

on components: $y_i(t) = \sum a_k y_k(t_k) + \sum b_p x_p(t_p) + \varepsilon(t)$

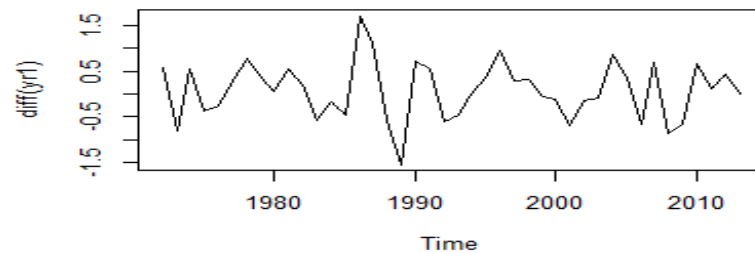
- ▶ $y_1 \sim ImIceM(t) \sim ImIceM(t-1) + UnEmpl(t) + UnEmpl(t-1) + EmIceM(t-1)$
- ▶ $y_2 \sim EmIceM(t) \sim EmIceM(t-1) + EmIceM(t-2) + GradM(t-2) + EmIceW(t)$
- ▶ $y_3 \sim ImIceW(t) \sim ImIceW(t-1) + UnEmpl(t) + GDP(t) + UnEmpl(t-1) + GDP(t-1)$
- ▶ $y_4 \sim EmIceW(t) \sim EmIceW(t-1) + EmIceW(t-2) + GradW(t-2)$
- ▶ $y_7 \sim ImForW(t) \sim ImForW(t-1) + UnEmpl(t) + GDP(t) + boom(t) + eea(t) + UnEmpl(t-1) + GDP(t-1)$
- ▶ $y_8 \sim EmForW(t) \sim EmForW(t-1) + ImForW(t) + ImForW(t-1) + UnEmpl(t) + GDP(t) + UnEmpl(t-1) + GDP(t-1)$
- ▶ $y_{net}(t) \sim y_{net}(t-1) + y_{net}(t-2) + UnEmpl(t) + GDP(t) + GDP(t-1) + UnEmpl(t-1) + GDP(t-2) + UnEmpl(t-2) + GradM(t-2) + GradF(t-2) + boom + bam + boom(t-1) + bam(t-1)$

Factors

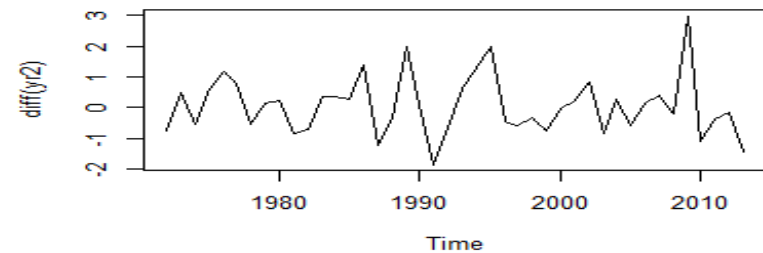


Migration rate components

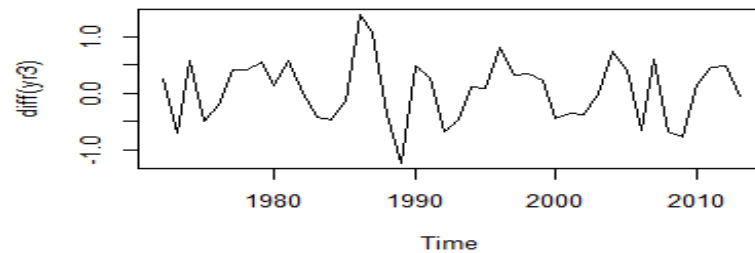
change in immigration rate of Icelandic men



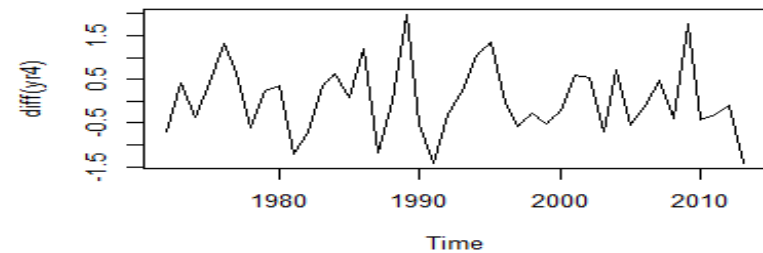
change in emigration rate of Icelandic men



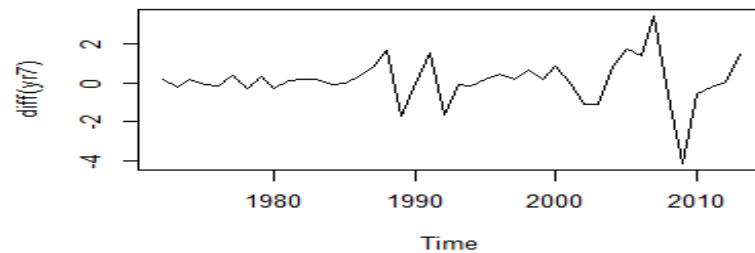
change in immigration rate of Icelandic women



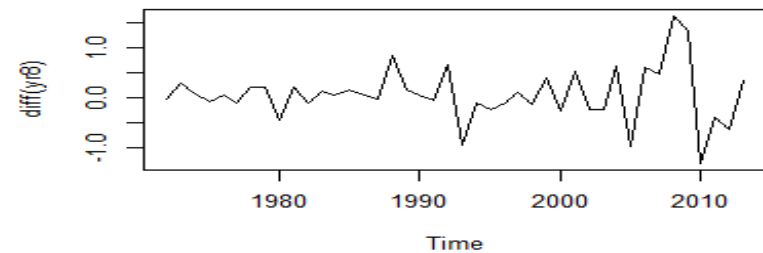
change in emigration rate of Icelandic women



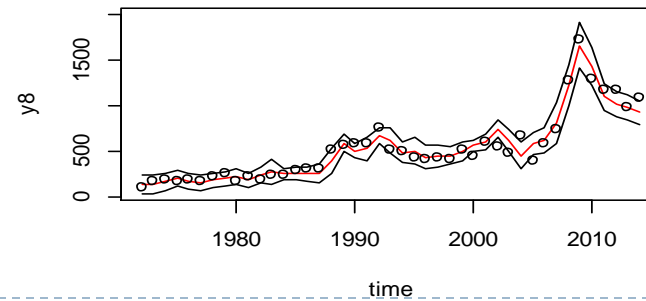
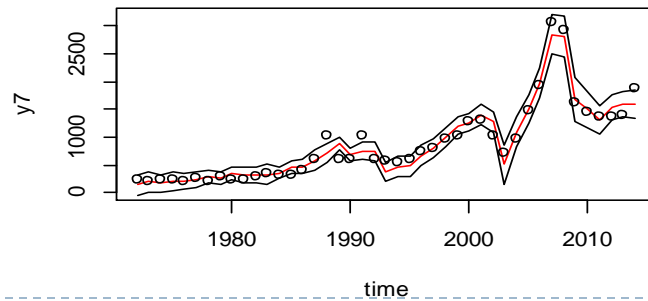
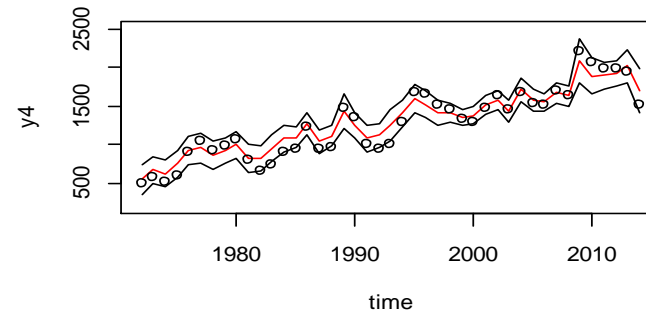
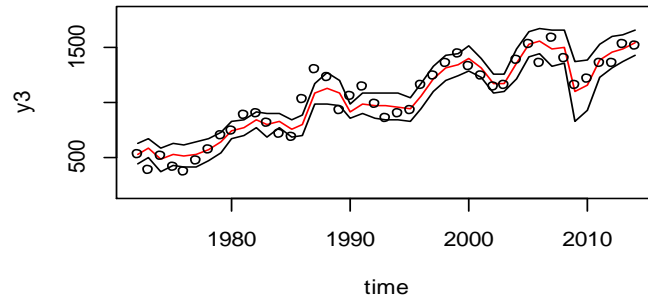
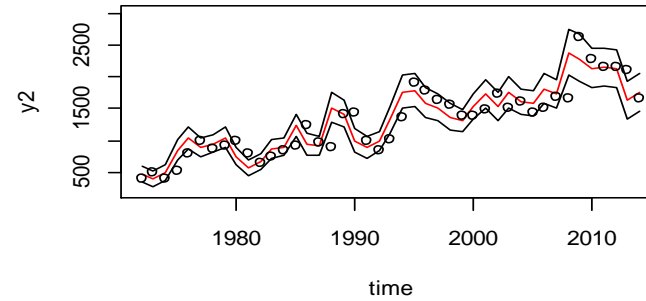
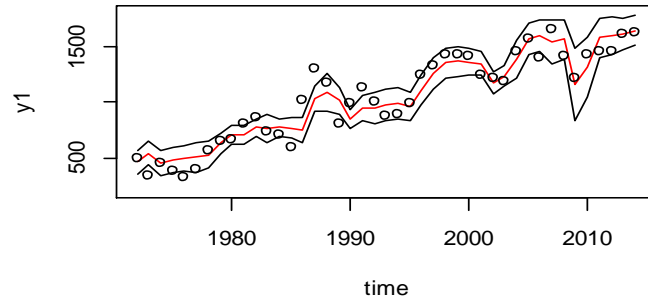
change in immigration rate of foreign women



change in emigration rate of foreign women



Migration components modeled



Long term migration



References

- ▶ Methodology of population projections,
http://www.statice.is/media/49266/hag_151118.pdf
- ▶ Population projections 2015-2065,
http://www.statice.is/media/49265/hag_151118_en.pdf
- ▶ Hyndman, R.J. and Ullah, M. S. (2007) Robust forecasting of mortality and fertility rates: a functional data approach, *Computational statistics & Data Analysis*, 61, 4942–4956.
- ▶ Hyndman, R. J., Booth, H., (2008) Stochastic population forecasts using functional data models for mortality, fertility and migration, *International Journal of Forecasting* 24 (2008) 323–342.