Hagtíðindi
Statistical Series

Working papers

**18 November 2015**

# Methodology of population projections

*Violeta Calian*
*Ómar Harðarson*

Statistics Iceland

## Icelandic summary
Í þessari greinargerð er lýst aðferðum og líkönum sem Hagstofan styðst við að til að setja saman mannfjöldaspá 2015–2065.

Mannfjöldaspáin byggir á því sem kallað er þáttaaðferð (e. *component method*). Þar er einungis tekið tillit til lýðfræðilegra breyta (þ.e. fæðinga, dauðsfalla og búferlaflutninga). Undirliggjandi áhrifaþættir mannfjöldaþróunarinnar eru sundurliðaðir en á lokastigi eru þeir lagðir saman. Þannig er mannfjöldi í lok árs = Mannfjöldi í byrjun árs + fæðingar – dauðsföll + (aðfluttir – brottfluttir).

Gerðar eru sérstakar spár um hvern þessara þátta, sem byggja á tímaraðalíkönum. Spárnar um dánarlíkur og frjósemi byggja á aðferð sem felur í sér sléttað, þverstætt niðurbrot (e. *smoothing, orthogonal function decomposition*) og tímaraðalíkönum. Frjósemisspáin gerir ráð fyrir þremur mismunandi forsendum um frjósemi til lengri tíma. Aðeins er gerð ein spá um dánarlíkur.

Gerð eru þrjú afbrigði um búferlaflutninga, sem byggja á mismunandi forsendum um aðflutning og brottflutning erlendra ríkisborgara. Aðeins eitt afbrigði er hins vegar gert um aðflutninga og brottflutninga íslenskra ríkisborgara. Spár um búferlaflutninga eru annars vegar til skamms tíma (næstu 5 ára) og til langs tíma (til 2065). Í skammtímaspánum er byggt á tímaraðalíkönum sem taka tillit til efnahagslegra þátta (vergrar landsframleiðslu - VLF, atvinnuleysi), en brottflutningur íslenskra ríkisborgara er að mestu skýrður með fjölda útskriftarnema úr menntaskólum tveimur árum áður.

Öllum spám sem byggja á líkönum fylgir tölfræðileg óvissa, sem táknuð er með öryggisbili. Öryggisbil fyrir frjósemisspá miðjuspár fellur hins vegar að háspánni og lágspánni fyrir frjósemi.

# 1. Introduction — methods and terminology

In this paper we describe the statistical methods used by Statistics Iceland for annual, short term and long term population projections. The central idea of our work is that by rigorous analysis of the time series data on migration, births and deaths, one can build valid statistical models, which can then be used for calculating point estimates and confidence intervals for predicted population components.

The terminology we use mirrors the underlying statistical methods. We call "population forecast" the predictions based on statistical modelling and the cohort component model. We do provide long term forecast of fertility and mortality rates, as well as of short term migration numbers, by age, sex and citizenship. We call "population projections" the predictions obtained from use of the standard cohort component model and certain expert assumptions on some components (such as long term migration) although including typical forecast rates (long term mortality and fertility rates) as well.

In recent years, more and more statistical offices are publishing official population predictions based on structural modelling, extrapolative methods as well as purely probabilistic models (see Booth (2006) for a review). Statistics Iceland integrates several techniques, according to the type of past data available for model fitting and testing. The short term net migration numbers are forecast by using time series dynamical models (see Calian (2013)), based on previous data analysis (see Hardarson 2010), while the fertility and mortality rates are forecast by using functional models with time series coefficients according to a functional data method proposed in Hyndman (2007).

For each of these components (and methods), we calculate prediction intervals accordingly. This calculation accounts for the statistical errors but not for the uncertainty in the main qualitative assumptions made for future developments. Predicting future shocks in any of the components depends on the existence of such strong fluctuations in the past data or on the assumptions about and distributions of various external factors employed by the models.

We also produce three projection *variants*, which are used in order to analyse the impact of different economic and demographic assumptions on future population development. They are point estimates, each with their own confidence intervals, for the total population (by age and sex), resulted from corresponding variants of the net migration, fertility and mortality rates.

The following variants are built for the short term migration:

a) the medium short term scenario is based on the current Statistics Iceland predictions for GDP growth and unemployment rates

b) the optimistic (or „high") scenario assumes higher GDP growth (double than the current Statistics Iceland predictions) and low unemployment

c) the pessimistic (or „low“) scenario assumes no GDP growth and high unemployment.

For long term predictions:

a) the long term values of net migration rates and fertility rates are constrained to converge to fixed values given by expert assumptions. In the case of fertility rates, the (independent) assumptions are very close to the estimated confidence interval bounds given by the model. We do not have any estimates of the uncertainty in the expert assumptions
b) no variants are assumed for the mortality rates, we only use the model based inference

The low total population projection variant is thus obtained from: low net migration predictions and low fertility rates predictions; the high population projection variant is obtained from: high net migration rates and high fertility rates predictions.

## 2. The total population

Given that we have modelled and predicted the values of net migration, mortality and fertility rates in the next years, the total population is calculated from the standard cohort component model which we briefly review here.

The following notations are used:

- the number of people of age *a* at the beginning of year $t$, $P_{\{a,t\}}$
- the number of people of age *a* which die during the year $t$, $D_{\{a,t\}}$
- the number of children who are born (age=0) during the year $t$, $B_{\{t\}}$
- the difference between the number of people who immigrate and emigrate (net migration numbers) during the year *t* having age *a*, $N_{\{a,t\}}$.

The basic demographic growth balance equation is then written as:

$P_{\{a,t\}} = P_{\{a,t-1\}} + N_{\{a,t\}} - D_{\{a,t\}}$, when $a > 0$, and $P_{\{0,t\}} = N_{\{0,t\}} + B_{\{t\}} - D_{\{0,t\}}$ when $a = 0$.

Starting with known population counts $P_{\{a,t_0\}}$, at the beginning of the current year $t_0$ one calculates the population for year $t_1 = t_0 + 1$ and repeats this step for the whole prediction horizon.

The net migration numbers $N_{\{a,t\}}$ are modelled as explained in the next section. The number of deaths and births $D_{\{a,t\}}$, $B_{\{t\}}$ are obtained by using the models and forecasts of the corresponding mortality and fertility rates and the relations defined below:

a) the number of children born in year $t$ is $B_{\{t\}} = \sum_a f_{\{a,t\}} P^{\{w\}}_{\{a,t\}}$, where $P^{\{w\}}_{\{a,t_0\}}$ is the total number of women of age $a$ in the total population, at time $t$ and the fertility rate $f_{\{a,t\}}$ is defined as the number of children born by women of age $a$ in year $t$ divided by the total number of women of age $a$ in year $t$.

b) the number of people who die in year $t$ is $D_{\{t\}} = \sum_a m_{\{a,t\}}\ P_{\{a,t\}}$ where $P_{\{a,t\}}$ is the total number of people (men, women or total) of age $a$, at time $t$ and the mortality rate $m_{\{a,t\}}$ is defined as the number of people of age $a$ (men or women) who die during year $t$ divided by the total number of people of age $a$.

A cautionary note may be in order here, regarding the importance of the reference date: if, for some of the population counts, this date is not the beginning but the middle of the year, or if the counts are in fact average population numbers, then one has to calculate all the mortality and fertility rates accordingly, or use a transformation which connects the two type of calculation results. For instance, the mortality rates with respect to two different reference dates are related by:

$$m_{\{for\ mid\ year\ or\ average\}} = m_{\{for\ beginning\ of\ year\}}/(1 + 0.5 * m_{\{for\ beginning\ of\ year\}})\ .$$

# 3. Modelling migration. Short term forecast. Long term projections.

## 3.1. Short term migration

We use dynamical, or auto-regressive distributed lag (ARDL) models for the auto-correlated and non-stationary time series involved in migration processes, in order to give valid point estimates and prediction intervals of migration numbers.

This approach does not treat auto-correlation and non-stationarity as nuisance phenomena but includes them into the model. The dependent variable at time t is modelled as a function of its own values at different time lags and of the values of several simultaneous or lagged predictor variables.

We obtain short time predictions for the number of immigrants/emigrants of Icelandic and foreign citizenships as functions of several time series predictors: unemployment, change in GDP values, number of graduating high school students and dummy variables mirroring the EEA resizing in time and the Icelandic economic boom which ended in 2008, and had to be accounted for due to its uniqueness.

## 3.2. Data

The source of migration demographic data is the Icelandic National Register, which contains information on migration events since 1961 and is updated on a continuous basis, as opposed to once a year, since 1986. As showed in Hardarson (2010), the estimated values of migration flows are reasonably accurate, although the short term migration has an influence on the accuracy of the emigration figures. This effect is mainly due to lagged de-registration processes, but is well measured and stable in time. The net migration numbers are not affected by this phenomenon in a significant way.

The data concerning unemployment rates, gross domestic product and their short term forecast is provided by the department of national accounts and public finances of Statistics Iceland. The number of graduating high-school students and its predicted values for the next few years is provided by the education statistics of Statistics Iceland.

### 3.3. Data analysis and dynamical models for short term forecast

We use the following notation: $y_1$, $y_2$ – and alternatively, for the ease of interpretation, (*ImIceM*, *EmIceM*) - the number of Icelandic immigrants/emigrants, men; $y_3$, $y_4$ – (*ImIceW*, *EmIceW*) - the number of Icelandic immigrants/emigrants, women; $y_5$, $y_6$ – (*ImForM*, *EmForM*) - the number of foreign immigrants/emigrants, men, $y_7$, $y_8$ – (*ImForW*, *EmForW*) - the number of immigrants/emigrants, women of foreign citizenship; $x_4$ - *UnEmpl* -the unemployment rate; $x_8$ - *GDP,* a measure of GDP, $x_5$, $x_6$ – (*GradM*, *GradF*), the number of graduating students, men and women respectively; *boom* – an indicator variable coupled to the Icelandic economic boom, reflecting also temporary changes in the registration process; *eea* – an indicator variable which reflects the entrance of Iceland into the *EEA,* and thus free movement of persons within that area.

All these (ten) time series of 44 years length (see Appendix 1, Figure 1 and 2 for their first order differences) were tested for: (i) stationarity, by using augmented Dickey - Fuller and Kwiatkowski – Philips – Schmidt - Shin (KPSS) and (ii) auto-correlation of first and higher order, by using Durbin-Watson and Breusch - Godfrey tests.

The auto-regressive distributed lag (*ARDL*) models are legitimate candidates for inference, see Pesaran (1995, 1999). They can be used to test for co-integration and to estimate long-run and short-run dynamics, even if the variables are stationary and non-stationary time series. We made sure none of these series is I(2). These are necessary but not sufficient conditions (see Johansen 2010), for un-biased and consistent point estimates and independent and identically distributed residuals. Choosing the structure and the order of the ARDL model by a consistent model selection criterion is a crucial step, too.

We have built, in a consistent and parsimonious way, the dynamical models given by a general formula:

$$y_i(t) \sim \sum y_k(t-j) + \sum x_p\,(t-l)\,,$$

where $k = i$ and $j = 1, 2, \ldots$ up to maximum lag order in $y_i$, or/and $k \neq i$ and $j = 0,1,2, \ldots$ up to maximum lag order in other dependent variables than $y_i$, $p = 4, 8$ and $l = 0,1, \ldots$ up to maximum lag order in exogenous variables.

They take particular simple forms, here written by using again the standard R notation (for dynamical models in our case) but the more meaningful variable names:

$$y_1 \sim ImIceM(t) \sim ImIceM(t-1) + UnEmpl(t) + UnEmpl(t-1) + EmIceM(t-1)$$

$$y_2 \sim EmIceM(t) \sim EmIceM(t-1) + EmIceM(t-2) + GradM(t-2) + EmIceW(t)$$

$y_3 \sim ImIceW(t) \sim ImIceW(t-1) + UnEmpl(t) + GDP(t) + UnEmpl(t-1) + GDP(t-1)$

$y_4 \sim EmIceW(t) \sim EmIceW(t-1) + EmIceW(t-2) + GradW(t-2)$

$y_7 \sim ImForW(t) \sim ImForW(t-1) + UnEmpl(t) + GDP(t) + boom(t) + eea(t)$

$+ UnEmpl(t-1) + GDP(t-1)$

$y_8 \sim EmForW(t) \sim EmForW(t-1) + ImForW(t) + ImForW(t-1) + UnEmpl(t)$
$+ GDP(t) + UnEmpl(t-1) + GDP(t-1)$

The coefficients of the models are given in Appendix 2 and the models are illustrated by Figure 3 (all models' fitting) and 4 (serial independence of residuals) of Appendix 1. The variables $y_5$ and $y_6$ were not modelled separately, as using the male time-series turned out the be less robust than optimal. Instead, they were obtained by using the empirically verified correlation between men and women migration numbers and the results of models $y_7$, $y_8$.

A ***note of caution*** is in order here: interpretation and model diagnostics when using ARDL are very different from the classical so-called *static* models. In particular, the effects of various exogenous or endogenous variables are measured by complex *combinations* of the dynamical model coefficients. Collinearity, short and long term effects are key ingredients which have to be treated appropriately. One way to apply the classical notions is to transform the dynamical model into the equivalent error correction model (ECM). We give an example of such equivalent model in Appendix 3.

Our calculation of prediction intervals is correct and optimal in some sense, see Pesaran (1997), for each dynamical model of $y_1$ , … to $y_8$, and it requires significant additional work for the net migration when not analysed as a time series on its own but as a sum of correlated time series.

We applied several tests in order to establish the goodness of fit and the behaviour of residuals for all our models:

A) Stationarity of residuals: the KPSS tests do *not* reject the hypothesis of stationarity of residuals' distributions for any of our models (all p-values greater than 0.1). Supporting this conclusion, augmented Dickey-Fuller tests reject *non*-stationarity (all p-values smaller than 0.01)
B) Normality of residuals: Jacques-Bera tests do *not* reject normality for any of the model residuals' distributions. The p-values of the tests are greater than 0.08 and reflect extremely well the general aspect of the empirical distributions; *q-q* plots.
C) Autocorrelation of residuals: Box-Ljung tests do *not* reject the hypothesis of random residuals. Same conclusion is supported by the direct calculation of autocorrelation for residuals shown in Figure 4 of Appendix 1.
D) Goodness of fit: rainbow tests imply that we can*not* reject any of the models.

The test results thus confirm the assumption that the models can be used for valid inference. The tests have to be interpreted with great care and flexibility, since most are themselves based on some models and null hypotheses rejections can be caused by more than one reason.

The results of the forecasting show that the economic factors have a strong effect (although not de-stabilizing) on the migration rates, as illustrated by the predicted values of migration under different scenarios which are created by modifying the unemployment and GDP growth values.

### 3.4. Long term migration

Structural models perform very well on short term but their main limitation is that they do require data on the future values of exogenous variables. These are not always easy to obtain. One could use instead purely probabilistic models, but one can argue that it is more efficient to take advantage, at least on the short term, of the information regarding various factors which influence the migration process. A new method based on alternative modelling is not yet sufficiently tested. Such a method could be based on a generalization of space-time series methods to the case of age-time series such as the migration components, or could use Bayesian priors for the factors we identified as significant for modelling migration. We are currently investigating the viability of these options.

In lieu of modelling the long term migration we rely on the expert opinion of Statistics Iceland's advisory committee on population projections, which was given in 2011. In this opinion, the three scenarios for the long term net migration are set by 400, 800 and 1,200 for the low, medium and high variants, respectively. In modelling the different migration patterns of Icelandic and foreign citizens, the long term net migration of Icelandic citizens is set as -800 for all variants.

## 4. Modelling fertility and mortality rates. Long term forecast and assumptions

### 4.1. Formulation of the problem

#### a) The data problem

Data on number of births and deaths by age and sex poses some challenges when calculating fertility, mortality rates and life tables. The main reason consists of zero counts for very small or very high ages. This happens due to the small size of the population and to the fact that most of these counts refer to rare events.

Standard solutions to such problem are to aggregate data over several years or to borrow data from similar and bigger populations, such as from other Nordic countries. Borrowing data is limited by the type of analysis though, and needs to be preceded by a proof of validity, i.e. testing hypotheses about the distributions of the needed variables in the populations. This analysis is ongoing.

### b) The mathematical problem

Let $y(t, x)$ denote the log of the observed mortality or fertility rate for age $x$ and year $t$. In general, we could assume there exists an underlying smooth function $f(t, x)$ which is observed with error and at discrete points $(t_i, x_a)$ of a (time-age) two-dimensional domain, giving the values $\{t_i, x_a, ; y_{ia}\}$, with $i = 1, \ldots, N$ and $a = M_0, \ldots, M$. We need to predict $y(t, x)$ for the same set of age values $x_a$ ($a = M_0, \ldots, M$) and for years $t_i$ ($i = N + 1, \ldots, N + h$), where $h$ is the length of the forecasting horizon.

To this effect, we should ideally fit one of the following types of model :

(i) a model $f(t, x)$ of time and age dependent rates;
(ii) a multivariate model $(f_t(x_{M_0}), \ldots, f_t(x_M))$ of time dependent rates, for each age;
(iii) a multivariate model $(f_{t_1}(x), \ldots, f_{t_N}(x))$ of age dependent rates, for each year.

Due to the asymmetry of the spatial-temporal domain in a forecasting context, to consistency issues and correlations of the rates across time and age values, it is difficult to find a parametric model for the function a) or the vector functions b) or c) without using simplifying assumptions.

## 4.2. The modelling solution

We use a functional data approach which was proposed in Hyndman et al (2007), Hyndman et al (2008). This method is robust to outliers, and it has been tested in a sufficiently extensive way.

Ideally, a general decomposition of the (smooth) function in a) should be $f(t, x)$ $=\sum_{j,b} \omega_{j,b}\varphi_{j,b}\ (t, x)$, with $\varphi_{j,b}$ being functions of an orthonormal basis over the bi-dimensional domain. However, a simpler and more efficient form is given by the factorization: $f_0(t, x)$=$\sum_{k=1,\ldots K} \beta_k(t)\varphi_k\ (x)$, where the number of orthogonal functions $K$ is reasonably small, and this is the solution adopted in Hyndman (2007, 2008).

Then the observations are modeled as: $y_{ia} = \mu(x_a) + f_0(t_i, x_a) + \mathrm{e}_{t_i}(\mathrm{x}_a) + \alpha_{t_i}(\mathrm{x}_a)\varepsilon_{t_i, x_a}$, where $\mu(x)$ is the mean of $f_0(t, x)$ across time (years), $\mathrm{e}_t(\mathrm{x})$ is the residual modelling error (assumed serially uncorrelated), the coefficient functions $\beta_k(t)$ are independent (by construction), $\varepsilon_{t,x}$ represents the random variation in birth or death rates and $\alpha_t(\mathrm{x})$ allow the variance to change with age and time.

The method has several steps:

1) Smoothing the raw data, i.e. log of crude mortality or fertility rates, by using spline functions with constraints on concavity and monotonicity, as functions of time and age. This reduces the observational noise.

2) Expressing the smooth functions as series expansions over a basis of orthonormal functions (see $f_0(t,x)$) above). Fitting time series models for the coefficient functions $\beta_k(t)$ of these series expansions and using these models for forecasting.

3) Using the forecast values of the coefficients to predict the values of the smooth functions and thus to predict mortality or fertility rates. Calculate prediction intervals based on the estimated variances of the error terms of step 1 and step 2.

## 4.3. Fertility rates models

Figure 5 shows the past and forecast values of fertility rates by age. The variation due to orthonormal basis functions used in modelling is: 76.4%, 15.3%, 3.7%, 1.4%, 0.8% and 0.7%.

We see that the increase in mothers' modal age with time will continue in the next 50 years. It is also clear that the fertility is predicted to decline for almost 30 years and then slightly increase again. This increase is due to the local peak in birth rate which occurred in 2008-2010 and to the average age (around 30 years) of mothers.

## 4.4. Mortality rates models

Figures 6 and 8 show the mean age pattern, the orthogonal basis functions and model coefficients for female and male mortality rates, respectively. The variation due to orthonormal basis functions is: 79.1%, 9.4%, 3.9%, 2.7%, 1.9%, 1.0% for the female model and 89.0%, 3.3%, 2.6%, 1.5%, 1.2%, 0.9% for the male model. The residuals (visualised in Figures 7 and 9) prove that we can use the fitted models to make predictions for future values of the mortality rates.

The first coefficient function and basis function show in both cases that the mortality has consistently decreased over time but the speed of this improvement depends on age. Thus very small ages and people 40 to 80 years old are the main beneficiaries of the trend, a similar conclusion to another population analysed with the same method (Hyndman (2008)).

In Figures 10 and 11, the past and forecast of female and male logarithm death rates are represented. We notice again the way the decrease in mortality over time depends on age but also that mortality changes are smaller around young adult ages than older.

## 4.5. Long term prediction intervals and assumptions

For both fertility and mortality rates we obtain short and long term forecasts, i.e. point estimates and prediction intervals, based on functional data models with time series coefficients, which do not depend on exogenous variables or any subjective inputs.

In the case of fertility rates we have three variants for the long term expert assumptions as well. The difference between the forecast point estimate and the medium value assumption is of the order of $10^{-2}$ and differences between the lower/upper bounds of the prediction intervals and the low/high values given be experts are of the order of $10^{-1}$. Therefore, for the long term predictions we smoothly connected our prediction intervals to the expert assumptions.

In the long term, total fertility rates are expected to converge to 1.8, 1.95 and 2.1 for the low, medium and high variants, respectively.

## 5. Conclusions

We have described in this paper the status of the population projection methodology used by Statistics Iceland. A detailed study of the performance of the employed models is the object of a future paper and it is based on using shorter time series in order to predict the (known) values of population components for recent years.

The dynamical models of short term migration can still be improved, especially if one aims to include more informative factors connected to internal and external social and economic processes. We are investigating the possibility of building models for long term migration as well and improving the ones for mortality and fertility rates. This can be done by using generalised age-time autoregressive models in all these cases, Bayesian prior distributions for exogenous factors and by changing the function basis in the functional data approach.

# References


Alho, J.M. (1997) Scenarios, uncertainty and conditional forecasts of the world population, *Journal of Royal Statistical Society* A 160, Part 1, 71–85.

Brunborg, H. and Cappelen, A. (2010) Forecasting migration flows to and from Norway using an economic model, *Joint Eurostat/UNECE work session on demographic projections*, Lisbon, Portugal, 28–20 April 2010.

Booth, H. (2006) Demographic forecasting: 1980 to 2005 in review, *International Journal of Forecasting*, 22(3), 547–581.

Calian, V. (2013) Dynamical models for migration projections, *proceedings of Joint Eurostat/UNECE Work Session on Demographic Projections*, Rome, 29.-31. október 2013, http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.11/2013/WP_4.4.pdf

Hardarson, O. (2010) Long-term and short term migration in Iceland - Analysis of estimation methods of Statistics Iceland, *Working paper 13 at the Joint UNECE/Eurostat Expert Group Meeting on Register-Based Censuses*, The Hague, The Netherlands, 10–11 May 2010.

Hyndman, R.J., & 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.

Johansen, S. (1996) Likelihood-based inference in cointegrated vector autoregressive models, *Oxford University Press*, Oxford.

Johansen, S. (2010) The analysis of nonstationary time series using regression, correlation and cointegration with an application to annual mean temperature and sea level, *Discussion Papers*, no.10–27, Department of economics, University of Copenhagen.

Keilman, N., Pham, D. Q., Hetland, A. (2002) Why population forecasts should be probabilistic — illustrated by the case of Norway, *Demographic Research*, Vol. 6, Article 15.

Pesaran, M. H. (1999) An autoregressive distributed lag modelling approach to cointegration analysis, *Symposyum of the Norwegian Academy of Science and Letters*, Oslo, 3–5 March 1995, and DAE Working Paper Series No. 9514. Department of Econometrics, University of Cambridge, 1999.

Pesaran, M.H., Shin, Y. and Smith, R.J. (1996) Testing the existence of long-run relationships, *DAE Working Paper Series No. 9622*, Department of Econometrics, University of Cambridge.

Pesaran, M.H., Shin, Y. and Smith, R.J. (2001) Bounds testing approaches to the analysis of level relationships, *J. Applied Econometrics*, 16: 289–326.

**Appendix 1:** Figures of variables, model fitting, residuals and forecasts.

**Figure 1.** Changes in exogenous variables 1971–2014

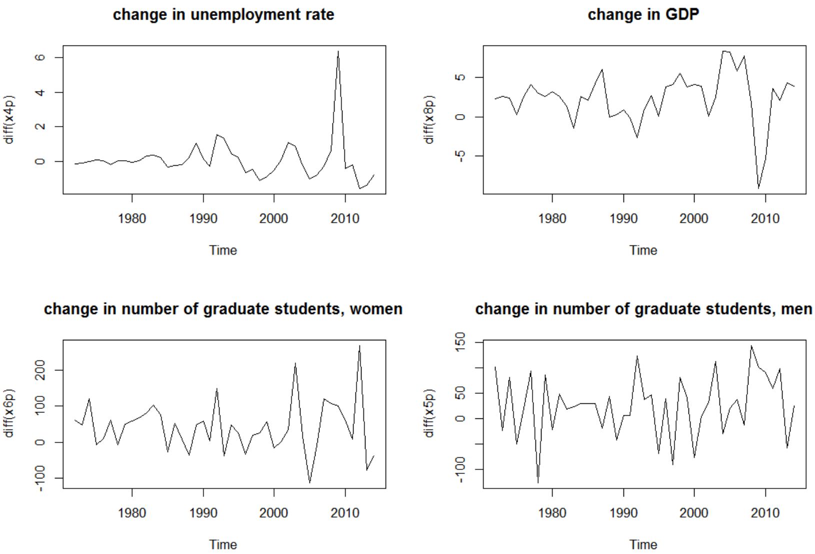



**Figure 2.** Changes in modelled variables 1971–2014

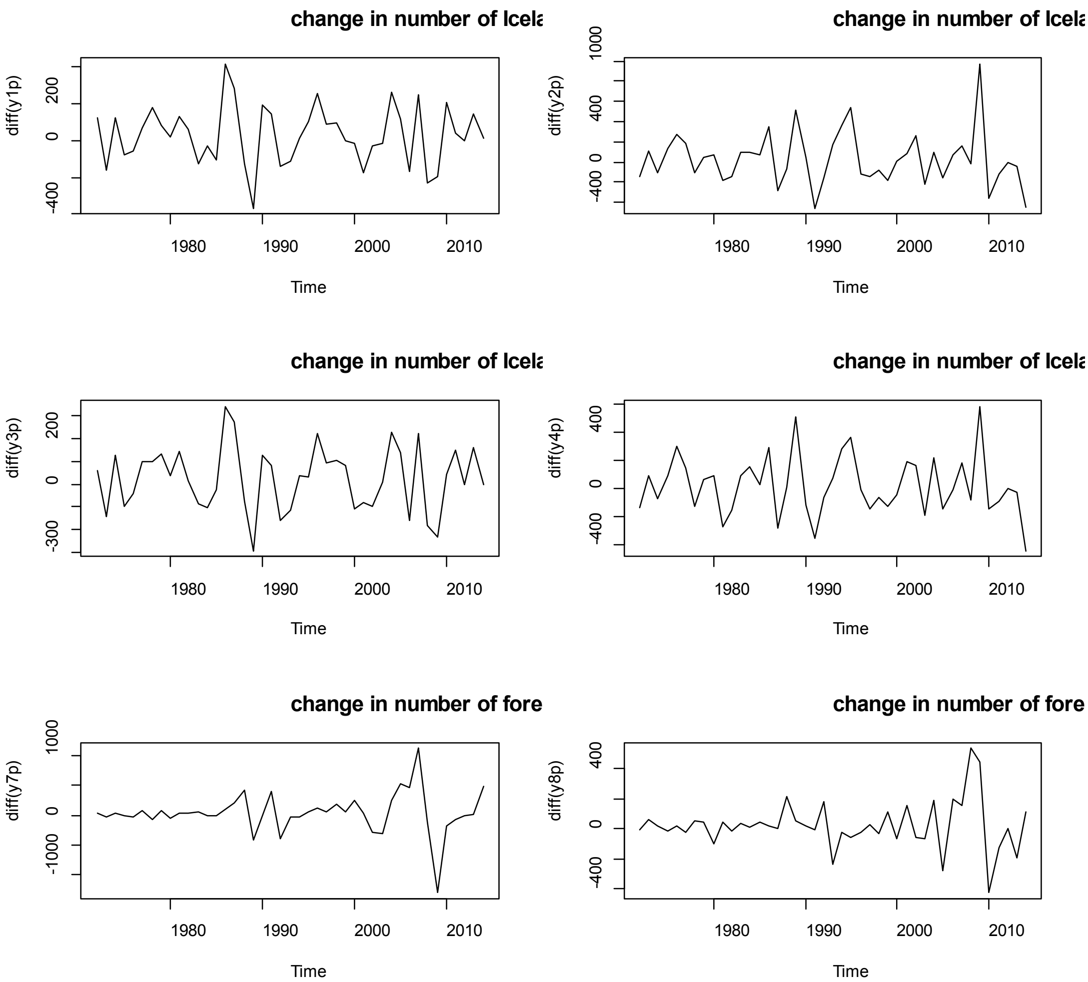


**Figure 3.** Model fitting for migration components 1971–2014

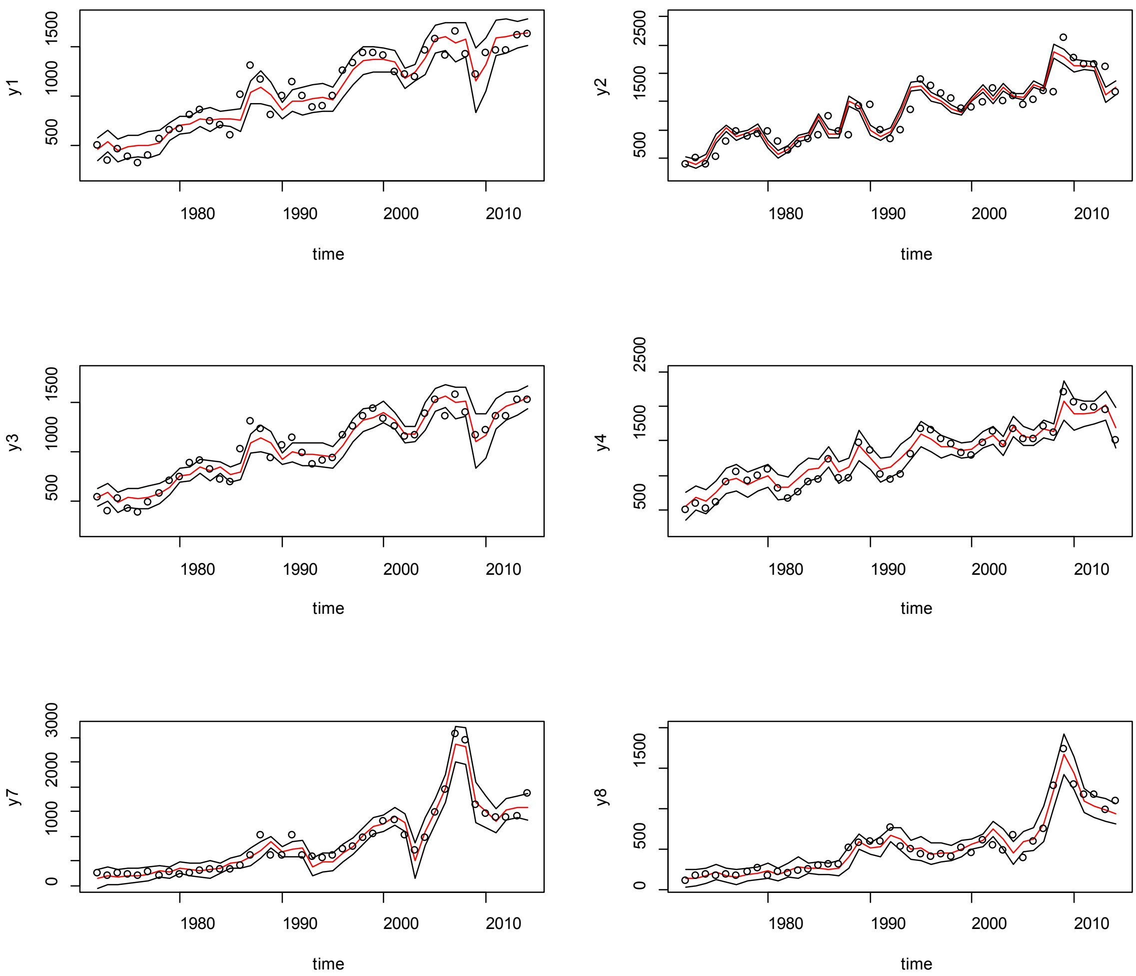


**Figure 4.** Autocorrelation of models' residuals

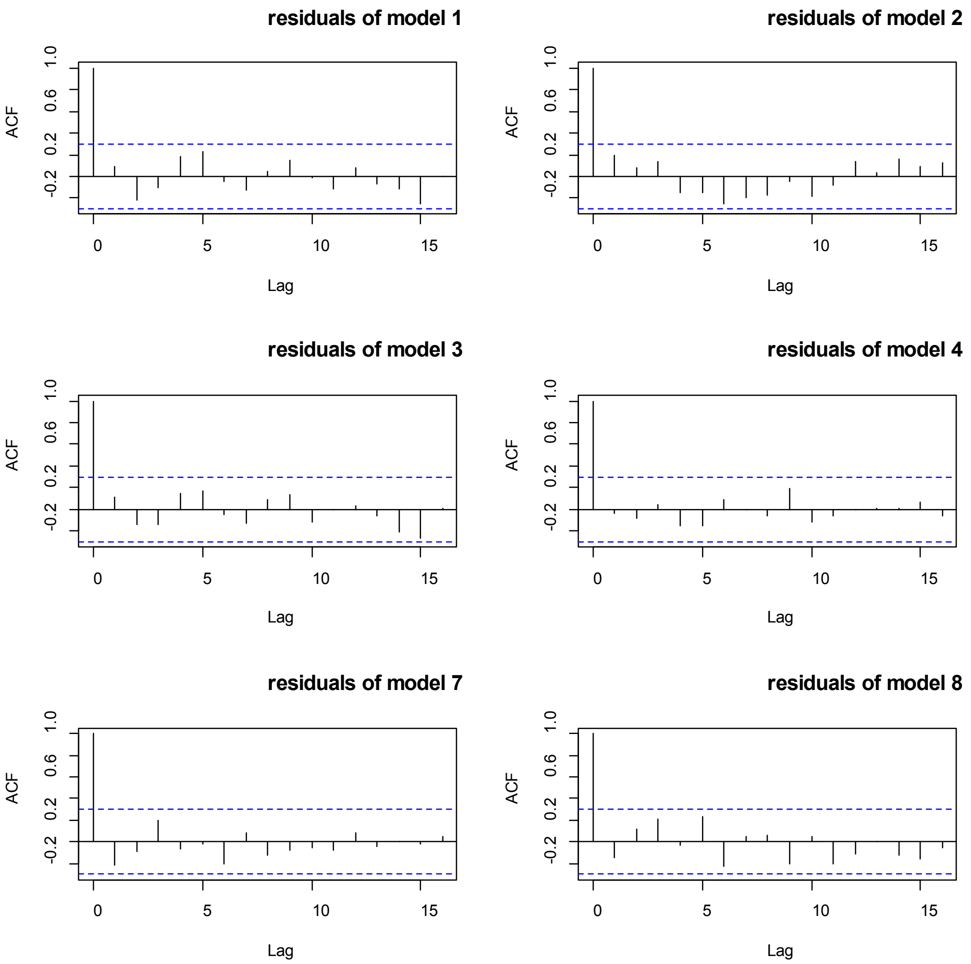


**Figure 5.** Fertility rates 1971–2065

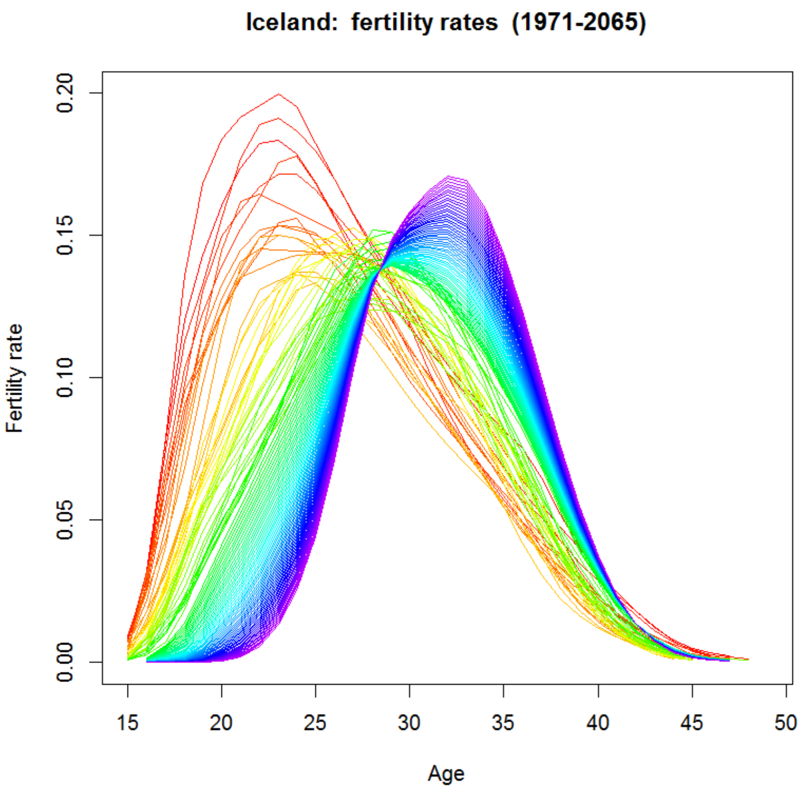



**Figure 6.** The basis functions and model coefficients for the female mortality rates

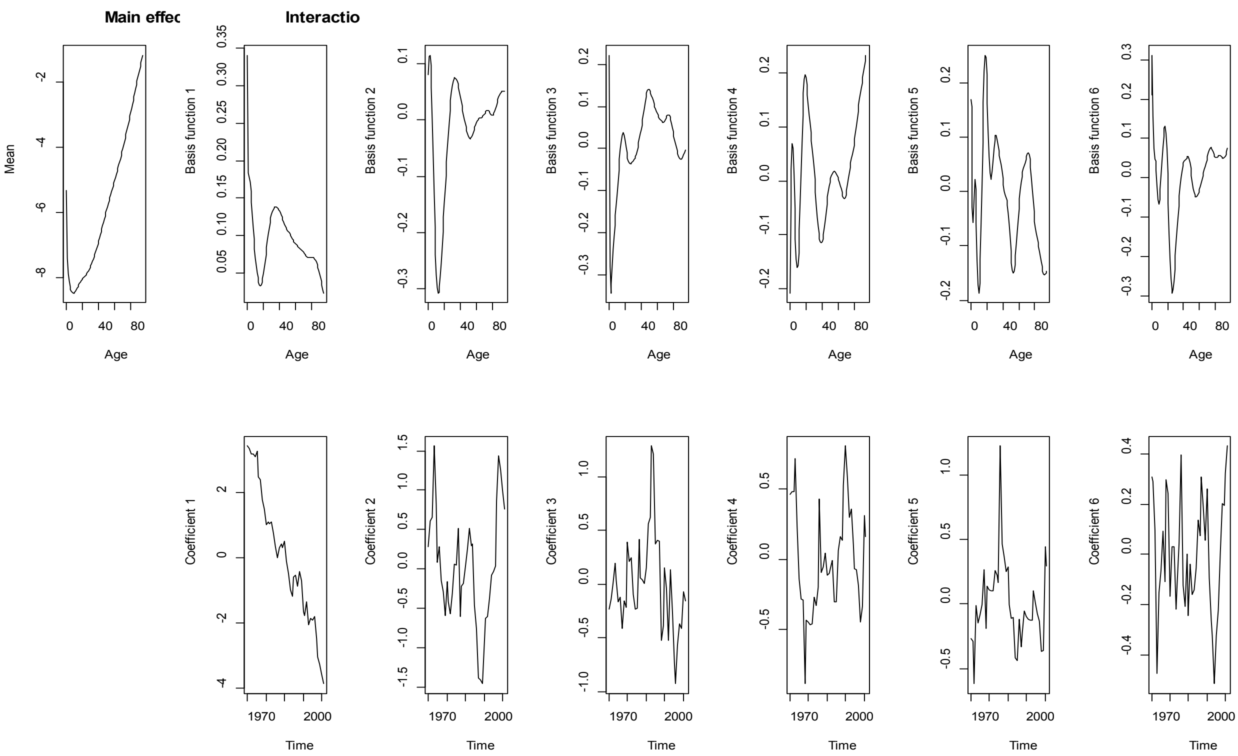


**Figure 7.** The residuals of the model for female mortality rates.

Darker colours mean higher values, in positive and negative directions (red or blue colours).

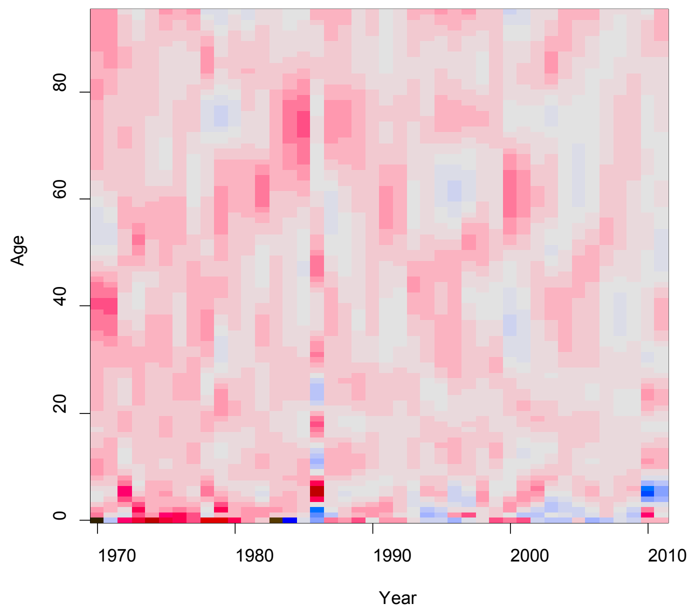



**Figure 8.** The basis functions and coefficients of the model for the mortality rates of men

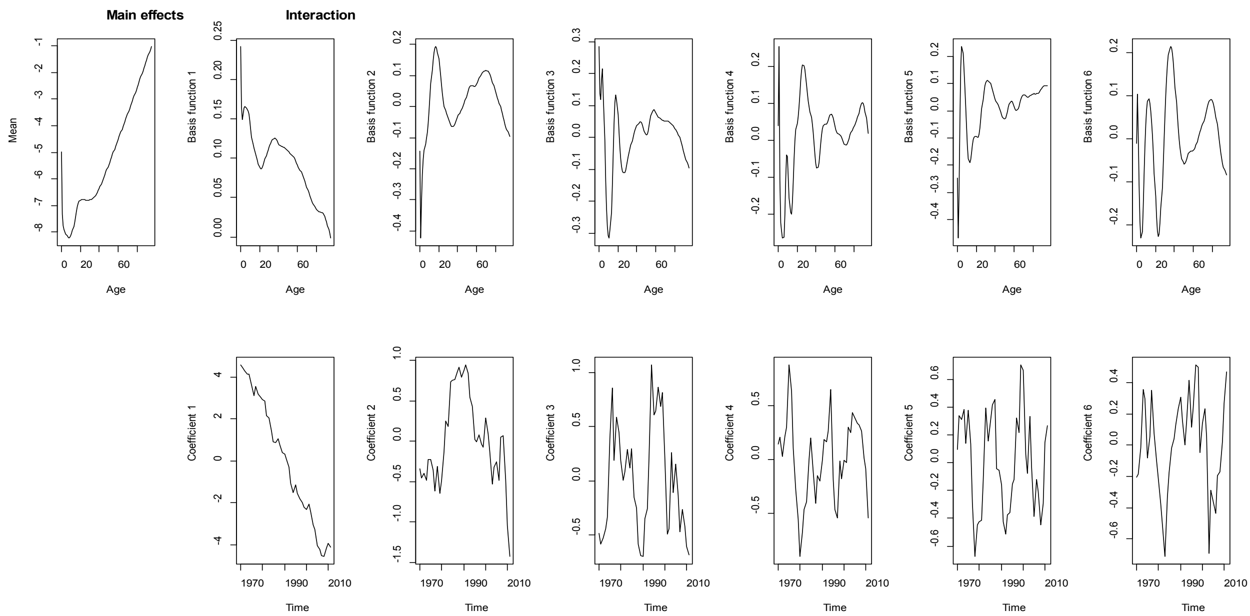


**Figure 9.** The residuals of the model for mortality rates of men

Darker colours mean higher values, in positive and negative directions (red or blue colours).

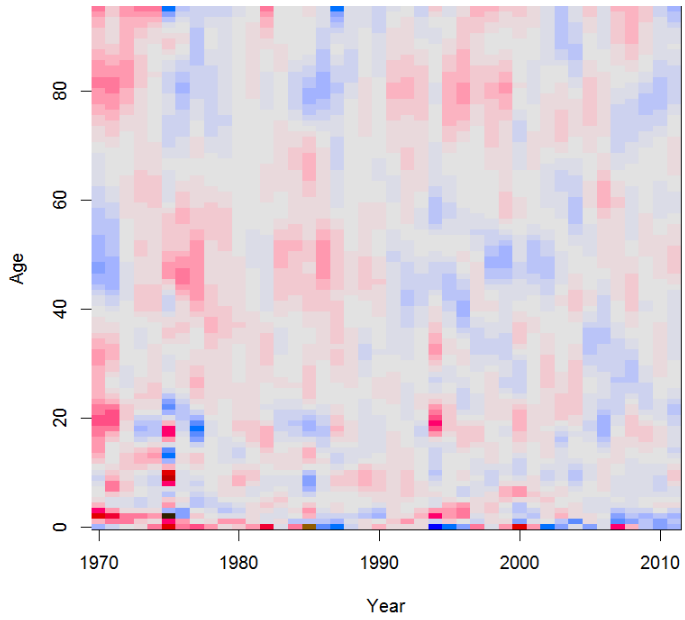



**Figure 10.** Logarithm of female death rates 1971–2065

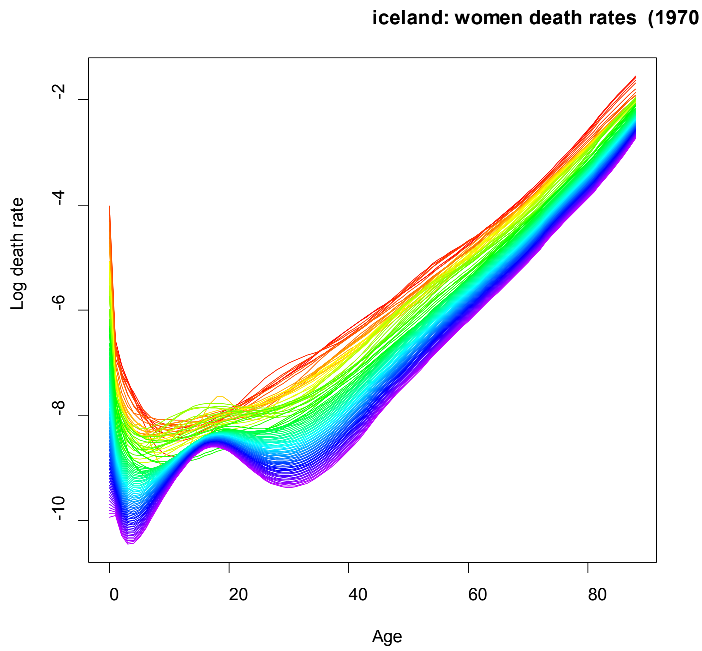
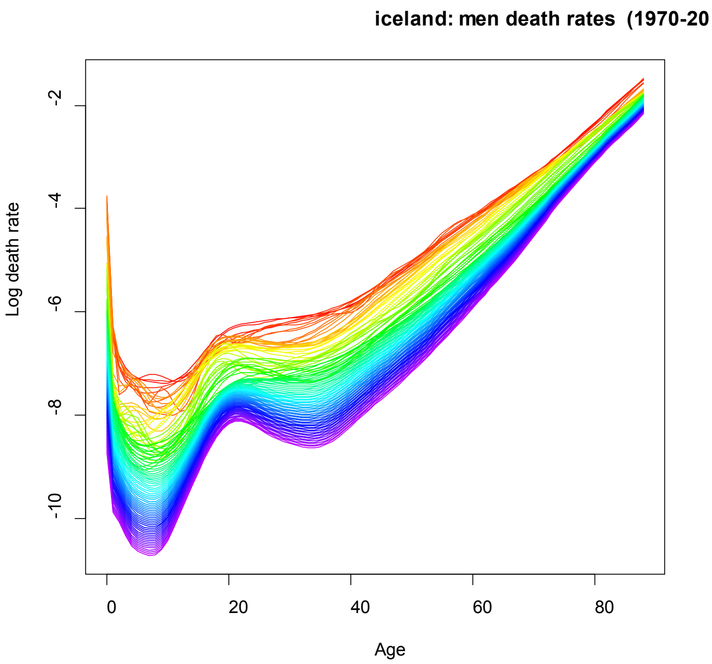


**Figure 11.** Logarithm of male death rates 1971–2065



**Appendix 2:** The coefficients of the ARDL models

In what follows, the notation $L(z, n)$ means the value of time series $z$ lagged by $n$ time units.

We must stress again that these coefficients are **not** measures of effects as it is the case for static models. By transforming an ARDL into an equivalent error correction model, one may find a combination of coefficients of the ARDL which gives information about the effects of the corresponding factors on a given response variable.

**Model 1**

| (Intercept) | L(y1, 1) | x4 | x8 | L(x4, 1) | L(x8, 1) |
|---|---|---|---|---|---|
| 37.113 | 0.567 | -14.068 | 20.425 | 35.437 | -16.393 |

**Model 2**

| (Intercept) | L(y2, 1) | L(y2, 2) | L(x5, 2) | y4 |
|---|---|---|---|---|
| -226.54211 | 0.06309 | -0.03785 | 0.05065 | 1.13987 |

**Model 3**

| (Intercept) | L(y3, 1) | x4 | x8 | L(x4, 1) | L(x8, 1) |
|---|---|---|---|---|---|
| 59.7953 | 0.6936 | -5.6251 | 24.0032 | 22.3355 | -21.8384 |

**Model 4**

| (Intercept) | L(y4, 1) | L(y4, 2) | L(x6, 2) |
|---|---|---|---|
| 274.6727 | 0.6570 | -0.2367 | 0.4387 |

**Model 7**

| (Intercept) | L(y7, 1) | x4 | x8 | boom | bam | L(x4, 1) | L(x8, 1) | L(boom, 1) |
|---|---|---|---|---|---|---|---|---|
| -160.882 | 0.782 | -176.596 | 1.408 | 528.284 | 155.928 | 129.114 | 6.146 | -807.380 |

**Model 8**

| (Intercept) | L(y8, 1) | y7 | L(y7, 1) | x4 | x8 | L(x4, 1) | L(x8, 1) |
|---|---|---|---|---|---|---|---|
| 25.465199 | 0.429888 | 0.008287 | 0.233925 | 18.517708 | -11.745838 | -8.048180 | 12.803516 |

**Appendix 3:** Examples of error correction models corresponding to the dynamic migration models

The ARDL models used in modelling short term migration can be written in one of the equivalent forms of an error correction model, such that their interpretation is closer to the classic regressions.

1. For example, the model for emigrating Icelandic women ($y_4$ or *EmIceW*),

$y_{4(t)} = \alpha_0 + \gamma_1 y_4(t-1) + \gamma_2 y_4(t-2) + \beta_2 x_6(t-2) + \varepsilon(t),$

where $x_6(t)$ is the number of graduating female students at any year *t*, is equivalent to:

$$y_{4(ECM)}(t) = A_0 + b_0\, x_6(t) - B_1\, \Delta x_{6(t)} - B_2\, \Delta x_6(t-1) - \Gamma_1\, \Delta y_4(t) - \Gamma_2\, \Delta y_4(t-1) + \varepsilon(t)$$

Here, the first two terms are the long run equilibrium terms ($A_0 + b_0\, x_6(t)$) while the rest are describing the influence of short term „shocks“ i.e variations, for two consequtive years in our case, in the number of graduating and emigrating women.

2. As a second example, the number of immigrating Icelandic women ($y_3$ or *ImIceW*),

$$y_3(t) = \alpha_0 + \gamma_1\, y_3(t-1) + \beta_{4,0}\, x_4(t) + \beta_{4,1}\, x_4(t-1) + \beta_{8,0}\, x_8(t) + \beta_{8,1}\, x_8(t) + \varepsilon(t),$$

where $x_4(t)$ and $x_8(t)$ are measures of unemployment and GDP in year *t*, is equivalent to

$$y_{3(ECM)}(t) = A_0 + b_{4,0}\, x_4(t) + b_{8,0}\, x_8(t) - B_{4,1}\, \Delta x_4(t) - B_{8,1}\, \Delta x_8(t) - \Gamma_1\, \Delta y_3(t) + \varepsilon(t)$$

The long run equilibrium ($A_0 + b_{4,0}\, x_4(t) + b_{8,0}\, x_8(t)$) is driven by the unemployment and GDP, while the local shocks are due to variations of same factors and recent past immigration numbers between successive years.

The exact values of the above constants are:

1) For model $y_{3(ECM)}$, we have: $A_0 \approx 425$ and $b_0 \approx 0.6$; $B_1 = B_2 \approx 0.4$; $\Gamma_1 \approx 0.4$; $\Gamma_2 \approx -0.2$.
2) For model $y_{4(ECM)}$, we find: $A_0 \approx 200$, $b_{4,0} \approx 56$ and $b_{8,0} \approx 10$; $B_{4,1} \approx 22.3$; $B_{8,1} \approx -21$; $\Gamma_1 \approx 0.7$.

Statistical Series **Working papers**
Vol. 100 • Issue 43 18 November 2015
ISSN 1670-4770
Supervision: Violeta Calian • violeta.calian@statice.is
Ómar Harðarson • omar.hardarson@statice.is
www.statice.is
Telephone (+354) 528 1000
Fax (+354) 528 1099