Stochastic Population Forecast for Germany

Alena Myšičková

Forschungsinstitut
für Neue Alterssicherungssysteme
und Rechtsbiometrik
in der Humboldt-Universität zu Berlin

nestor

HUMBOLDT-UNIVERSITÄT ZU BERLIN

C.A.S.E.
CENTER FOR APPLIED STATISTICS AND ECONOMICS

Motivation

Population forecast needed for many purposes

- financing of social systems
- labour market
- consumer demand
- financing of public infrastructure...

Stochastic Projection

- application of time series models
- modeling and forecasting of the vital rates separately
- demographic variables
  - Mortality

Stochastic Projection

- application of time series models
- modeling and forecasting of the vital rates separately
- demographic variables
  - Mortality
  - Fertility

Stochastic Projection

+ application of time series models
+ modeling and forecasting of the vital rates separately
+ demographic variables
  - Mortality
  - Fertility
  - Migration

Outline

1. Motivation ✓
2. Mortality
3. Fertility
4. Migration
5. Population forecast

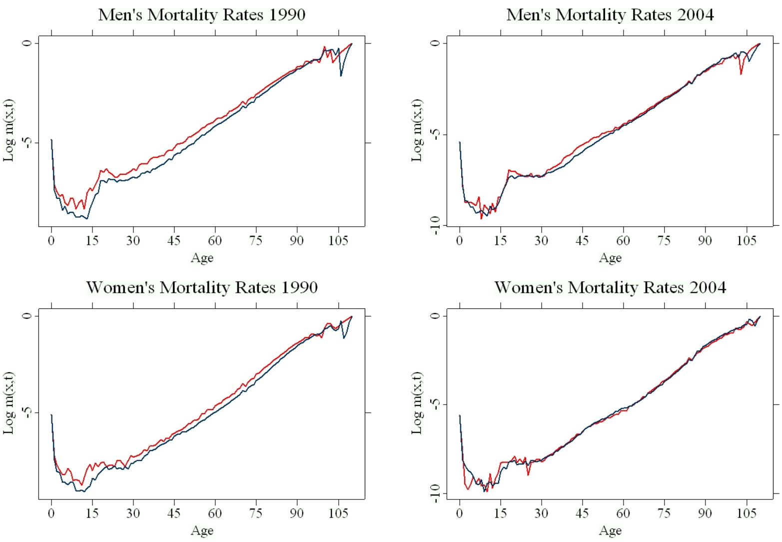
Mortality Data

- Age-specific Mortality Rate (ASMR)
- Source: Human Mortality Database (http://www.mortality.org/)
- Years 1956–2004
- Age groups: 0, 1, 2, ..., 110+
- Data just from old West German States considered
- Missing data (in the oldest age groups) replaced by a linear interpolation

Age-specific Mortality rate ($m_{x,t}$)

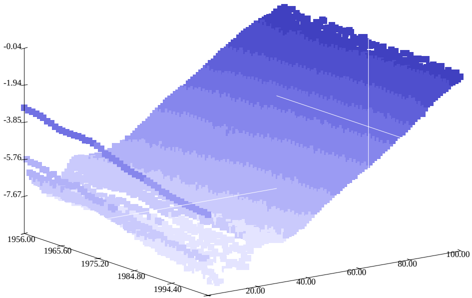
$m_{x,t}$ = total number of deaths per 1000 people of the age x in the time period t

- mortality relations in East Germany adapt on the relations in West Germany
- mortality decline in the 2nd half of the 20th century
- mortality decline in all age groups



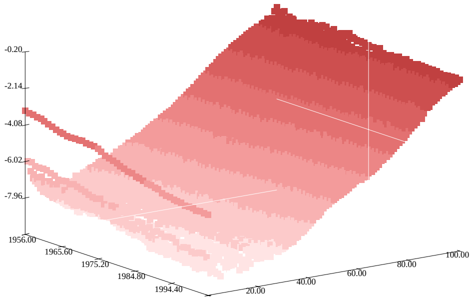
Age-specific mortality rates for 1990 and 2004:
East Germany red, West Germany blue.

Men's Mortality Rates



Log mortality rates for males.

Women's Mortality Rates



Log mortality rates for females.

Lee-Carter Model for Mortality

$$\log(m_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t}$$

a_x - age specific parameter
e^{a_x} - the general shape of the mortality function across the age

b_x - age specific parameter
- how fast declines the rate with respect to changes in k_t

k_t - time-varying mortality index

$\varepsilon_{x,t} \sim (0, \sigma_\varepsilon^2)$ - error term
- particular age-specific historical influences not captured by the model

Estimation of the Model

Assumptions:

$$\sum_t k_t = 0$$

$$\sum_x b_x = 1$$

$\longrightarrow$ $a_x = \frac{1}{T} \sum_t \log(m_{x,t})$, with $t = 1, \ldots, T$

Estimation of the Model

Assumptions:

$$\sum_t k_t = 0$$

$$\sum_x b_x = 1$$

$\longrightarrow$ $a_x = \frac{1}{T} \sum_t \log(m_{x,t})$, with $t = 1, \ldots, T$

$\longrightarrow$ Singular Value Decomposition (SVD) to estimate k_t and b_x

Estimation of the Model

Singular Value Decomposition of Matrix $M(p \times T)$:

$$M = \log(m_{x,t}) - a_x = \Gamma \, \Lambda \, \Delta^\top$$

$\Gamma(p \times r)$ and $\Delta(T \times r)$ - orthonormal: $\Gamma^\top \Gamma = \Delta^\top \Delta = I_r$, $r = \mathrm{rank}(M)$

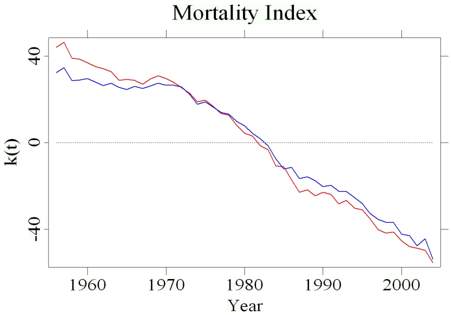
$\Lambda = \mathrm{diag}(\lambda_1^{1/2}, \dots, \lambda_r^{1/2})$, λ_i - eigenvalues of $M^\top M$

$\longrightarrow$ b_x, k_t - first column vectors of matrices Γ and Δ, resp. multiplied by $\lambda_1^{1/2}$

	Males		Females	
Age	a_x	$b_x * 10^3$	a_x	$b_x * 10^3$
0	-4.27	27.24	-4.52	23.58
5	-7.87	24.33	-8.18	19.42
10	-8.27	21.91	-8.67	14.91
15	-7.65	15.58	-8.13	11.96
20	-6.61	11.75	-7.67	8.38
25	-6.73	10.30	-7.64	11.23
30	-6.66	9.97	-7.38	11.65
35	-6.40	8.86	-6.98	10.05
40	-5.98	7.30	-6.54	8.42
45	-5.52	6.23	-6.07	7.20
50	-5.03	6.28	-5.64	6.91
55	-4.55	7.38	-5.25	7.17
60	-4.06	8.26	-4.80	7.94
65	-3.58	8.53	-4.30	9.06
70	-3.11	8.21	-3.74	9.75
75	-2.65	7.53	-3.14	9.97
80	-2.18	6.31	-2.54	8.70
85	-1.72	5.03	-1.96	6.96
90	-1.28	3.44	-1.46	4.76
95	-0.90	2.63	-1.03	3.48
100	-0.43	2.19	-0.59	4.37
105	-0.42	1.96	-0.51	2.84

Mortality index k_t

Aim: find adequate ARIMA Time Series Model for the forecast (Box-Jenkins Method)



$\longrightarrow$ Random Walk with Drift appropriate for both genders

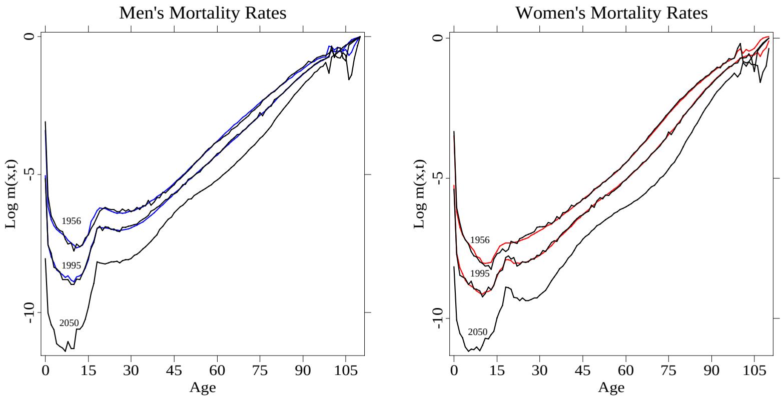
Random Walk with Drift

$$k_t = \delta + k_{t-1} + u_t$$

δ - slope of the deterministic trend; $u_t \sim \mathrm{WN}(0, \sigma_{\mathrm{u}}^2)$
The fitted model:

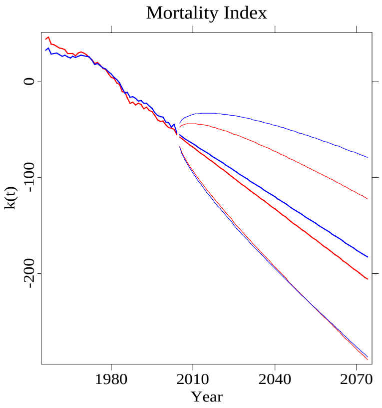
- males: $k_t = -1.84 + k_{t-1} + u_t$ with $\widehat{\sigma}_{u^m} = 6.33$
- females: $k_t = -2.15 + k_{t-1} + u_t$ with $\widehat{\sigma}_{u^f} = 5.11$
- $\delta =$ the average annual changes in k
- standard deviation $\widehat{\sigma}_u$: uncertainty associated with a one-year forecast k_t

The Fitted Model



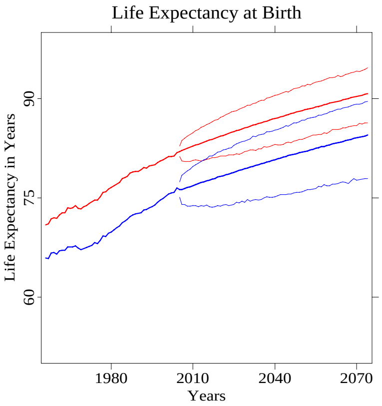
Actual and fitted mortality rates for all ages in 1956 and 1985 with forecast for 2050.

Forecast of Mortality Index

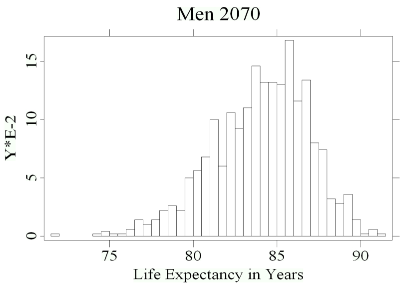


Mortality index for men and women with 95% forecast intervals.

Life Expectancy



Life expectancy at birth for boys and girls with 95% forecast intervals.



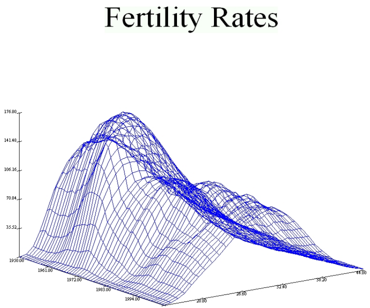
Histogram for the life expectancy of newborn boys (left) and girls (right) in year 2070.

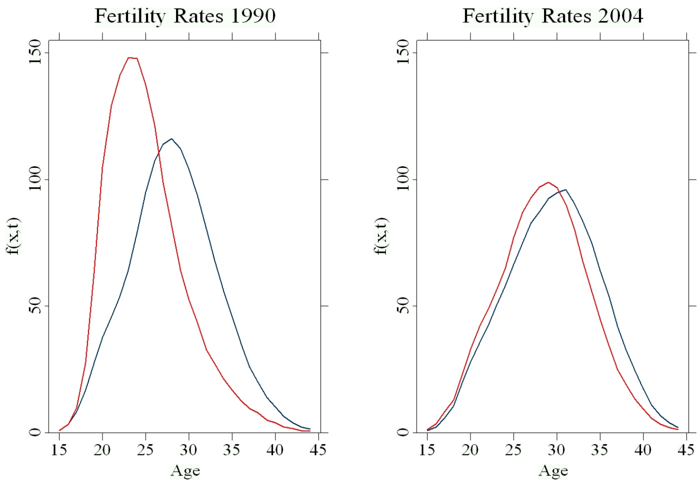
Fertility Data

- Age-specific Fertility Rate (ASFR)
- Source: Statistisches Bundesamt (http://www.destatis.de/)
- Years 1950–2005
- mothers at the age of: 15, 16, ..., 44
- old West German States data considered
- no missing data

Age-specific Fertility Rate ($f_{x,t}$)

$f_{x,t}$ = number of births from mothers at the age of x per 1000 women at the same age in the time period t





Age-specific Fertility Rates in 1990 and 2004.

East Germany red, West Germany blue.

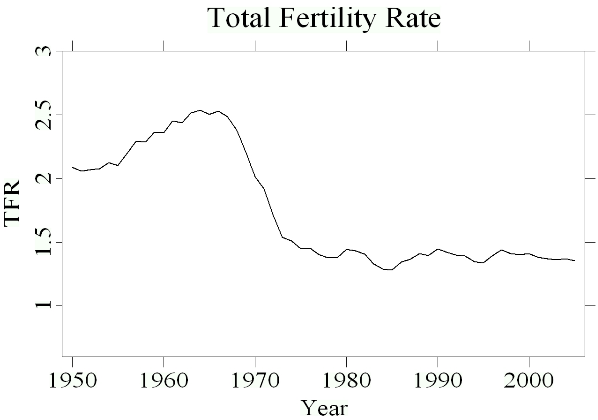
Totale Fertility Rate

$$\mathrm{TFR}_t = \sum_{x=15}^{44} f_{x,t}$$

- the sum of the age-specific rates for the given time period t
- the average number of children that would be born to a woman over her lifetime if she were to experience the exact current age-specific fertility rates (ASFRs) through her lifetime

$\longrightarrow$ Interpretation: mean number of children a woman is expected to bear during her child-bearing years

- independent from the age-sex structure of the population



Max(TFR) = 2.54 in 1964
Min(TFR) = 1.28 in 1985.

Lee-Carter Model for Fertility

→ Lee-Carter Model:

$$f_{x,t} = a_x + b_x f_t + \varepsilon_{x,t}$$

a_x - age-specific parameter
$A = \sum_x a_x$ - average value of the TFR over the sample period
b_x - age-specific parameter
f_t - time-varying fertility index
$\varepsilon_{x,t} \sim (0, \sigma^2_\varepsilon)$ - error term

Lee-Carter Model for Fertility

→ Lee-Carter Model:

$$f_{x,t} = a_x + b_x f_t + \varepsilon_{x,t}$$

a_x - age-specific parameter

$A = \sum_x a_x$ - average value of the TFR over the sample period

b_x - age-specific parameter

f_t - time-varying fertility index

$\varepsilon_{x,t} \sim (0, \sigma_\varepsilon^2)$ - error term

$\longrightarrow$ $\text{TFR}_t = A + f_t + E_t$, with $E_t = \sum_x \varepsilon_{x,t}$

Lee-Carter Model for Fertility

$\longrightarrow$ Lee-Carter Model:

$$f_{x,t} = a_x + b_x f_t + \varepsilon_{x,t}$$

a_x - age-specific parameter
$A = \sum_x a_x$ - average value of the TFR over the sample period
b_x - age-specific parameter
f_t - time-varying fertility index
$\varepsilon_{x,t} \sim (0, \sigma_\varepsilon^2)$ - error term
$\longrightarrow$ $\mathrm{TFR}_t = A + f_t + E_t$, with $E_t = \sum_x \varepsilon_{x,t}$
$\longrightarrow$ f_t - deviation in period t of the TFR from its long term average A

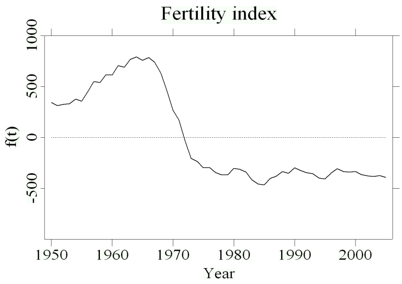
Estimation of the model

Analogue to mortality model:

- Assumptions: $\sum_t f_t = 0 \, , \quad \sum_x b_x = 1$

$\longrightarrow$ $a_x = \frac{1}{T} \sum_t f_{x,t}$, with $t = 1, \ldots, T$

$\longrightarrow$ SVD to estimate b_x and f_t



- Fertility index f_t - large variance, problems with direct forecasting $\longrightarrow$ not demographically plausible results
- definition of bounds for TFR:
  - L - lower bound
  - U - upper bound
  - F^* - ultimate level

Transformed Fertility Index

$$g_t = \log\left(\frac{F_t - L}{U - F_t}\right) \Leftrightarrow F_t = \frac{U \cdot \exp(g_t) + L}{1 + \exp(g_t)}$$

- $F_t = f_t + A$ - fitted value of the TFR
- $g_t \longrightarrow \infty \Rightarrow F_t \longrightarrow U$
- $g_t \longrightarrow -\infty \Rightarrow F_t \longrightarrow L$

$\longrightarrow$ modeling of time series $(g_t - G^*)$

- with G^* calculated as: $G^* = \log\left(\frac{F^* - L}{U - F^*}\right)$

ARMA(1,1) Model

$$g_t - G^* = \delta + \phi(g_{t-1} - G^*) + \theta u_{t-1} + u_t \ , \quad u_t \sim \mathrm{WN}(0, \sigma_\mathrm{u}^2)$$

δ, ϕ, θ - Parameters

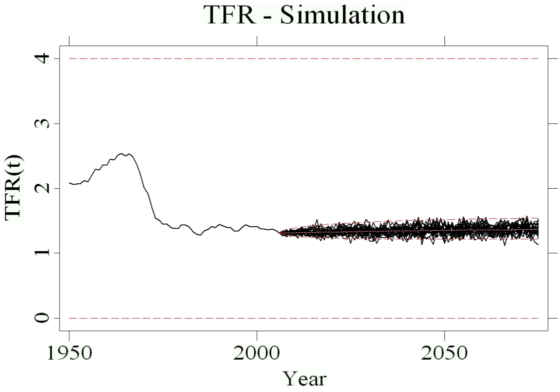
The Fitted Model

with $L = 0$, $U = 4$, and $F^* = 1.395$ (TFR average in 1973-2004)

$$g_t + 0.625 = 0.360 + 0.987(g_{t-1} + 0.625) + 0.234 u_{t-1} + u_t \ ,$$

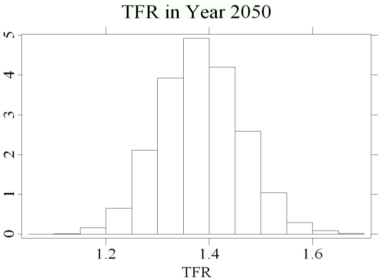
$\widehat{\sigma}_u = 0.081$.

Forecast of TFR

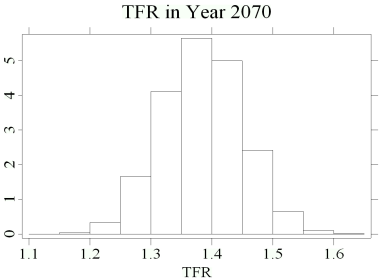


95%-Forecast Interval with Monte Carlo Simulation.

TFR Distribution in 2050 and 2070



Median = 1.375



Median = 1.381

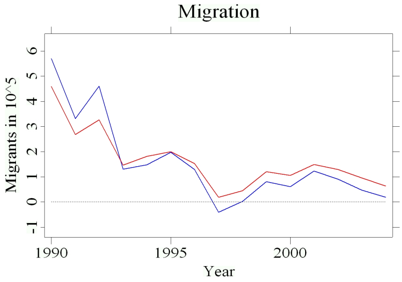
Migration

- modeled **constant**
- influenced by many factors
  - political development in Germany
  - political development in migration countries
  - development of the labour market
  - ...

$\longrightarrow$ average of last 10 years

Assumption for Immigration:

- 51 000 men and 95 000 women per year
- altogether: 146 000 persons
- Assumptions from Statistisches Bundesamt: 100 000 and 200 000



Cohort-Component Method

For females:

$$\begin{pmatrix} N_{1,t+1} \\ N_{2,t+1} \\ N_{3,t+1} \\ \vdots \\ N_{k,t+1} \end{pmatrix} = \begin{pmatrix} 0 & \dots & s \cdot F_{15,t} & \dots & s \cdot F_{44,t} & 0 \\ P_{1,t} & 0 & \dots & \dots & \dots & 0 \\ 0 & P_{2,t} & \ddots & 0 & \dots & 0 \\ 0 & 0 & \dots & \dots & P_{k-1,t} & 0 \end{pmatrix} \cdot \begin{pmatrix} N_{1,t} \\ N_{2,t} \\ N_{3,t} \\ \vdots \\ N_{k,t} \end{pmatrix} + \begin{pmatrix} NI_{1,t} \\ NI_{2,t} \\ NI_{3,t} \\ \vdots \\ NI_{k,t} \end{pmatrix}$$

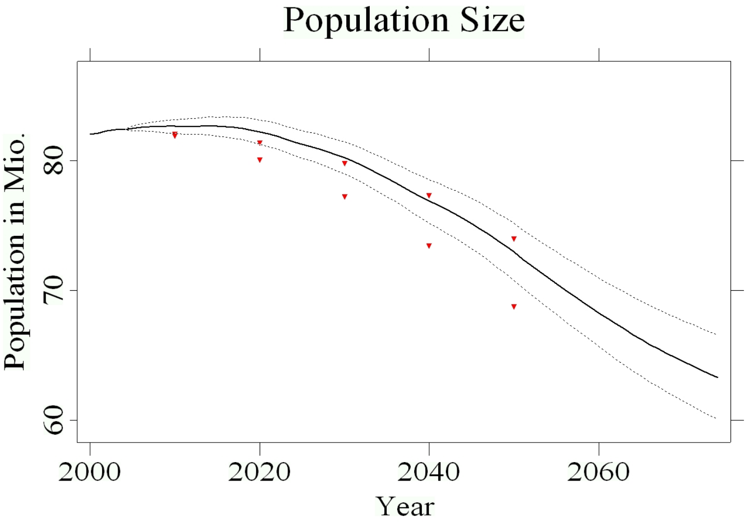
Cohort-Component Method

- $N_{i,t}$ - population in the i-th age group in year t
- $F_{i,t}$ - age-specific fertility rate for mother in the i-th age group in year t multiplied by the number of woman
- $P_{i,t}$ - probability for one person in the i-th age group to achieve the next year
- $NI_{i,t}$ - number of immigrants in the i-th age group in year t
- s - sex ratio at birth (taken as 100:106)
- for males a similar matrix without fertility rates, male newborns calculated from the number of female birth

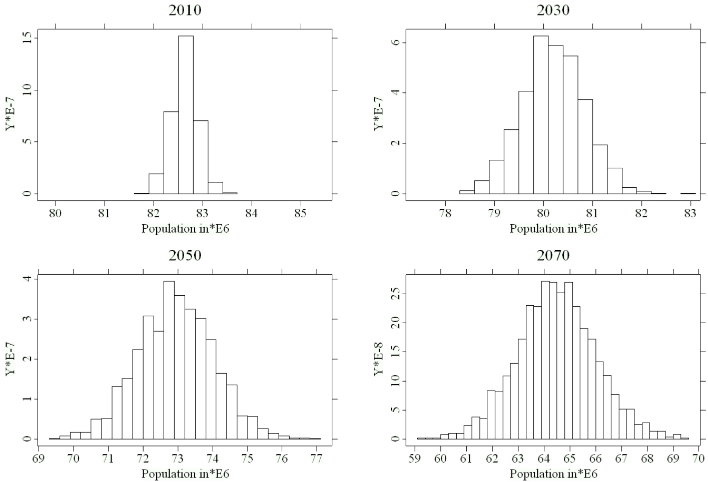
Population Size

Estimation after 5000 simulations compared with the results of the 11th Coordinated Projection of Statistisches Bundesamt

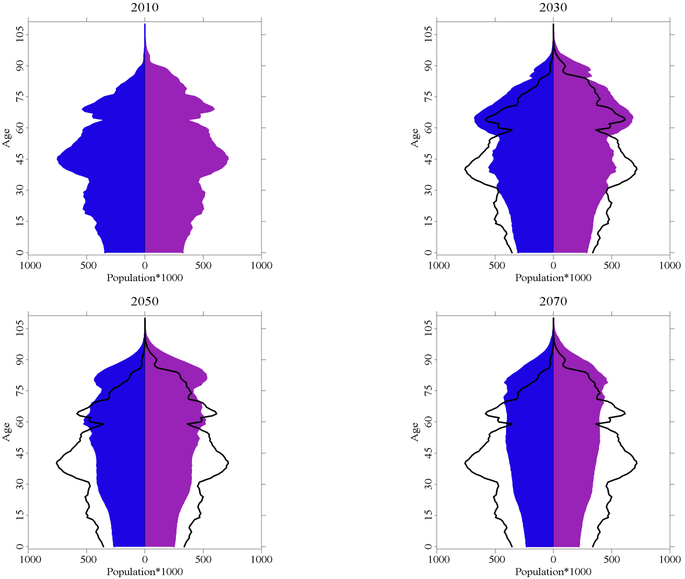
Year	5%-Quantile	Mean	95%-Quantile	L-Bound	U-Bound
2010	82.01	82.53	83.04	81.89	82.04
2025	79.94	81.05	82.17	78.77	80.67
2040	74.81	76.55	78.26	73.42	77.28
2055	67.10	69.51	71.95	–	–
2070	60.81	63.83	66.90	–	–



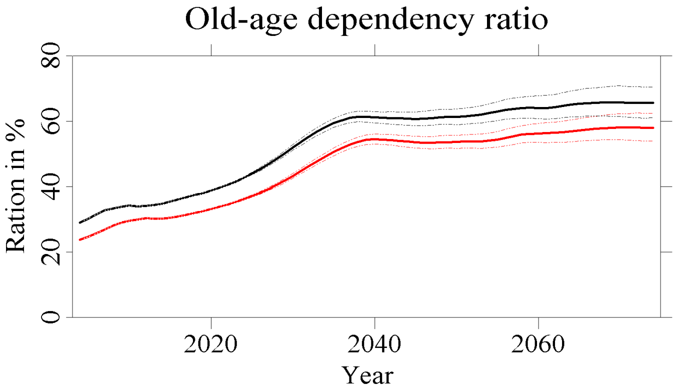
Population Size
Population in Mio.
60
70
80
2000
2020
2040
2060
Year



Distribution of population in mio. in years 2010, 2030, 2050 and 2070



Population pyramids in years 2010, 2030, 2050 and 2070.



Age quotient with age limit 65 years (black) and with age limit 67 years (red).

References

B. Babel.
Bevölkerungsvorausberechnungen unter besonderer Berücksichtigung von Unsicherheiten
EUL Verlag, 2007

R.D. Lee, L.R. Carter.
Modelling and Forecasting U.S. Mortality.
Journal of the American Statistical Association, 87:659–675, 1992.

R.D. Lee.
Modelling and forecasting the time series of US fertility: Age distribution, range, and ultimate level.
International Journal of Forecasting, 9:187–201, 1993.

References

- O. Lipps und F. Betz
  Stochastische Bevölkerungsprojektion für West- und Ostdeutschland
  Zeitschrift für Bevlkerungswissenschaft, 30(1):3–42, 2005.

- P. Pflaumer
  Confidence Intervals for Population Projections based on Monte Carlo Methods
  International Journal of Forecasting, 4:135–142, 1988.

- Statistisches Bundesamt
  11. koordinierte Bevölkerungsvorausberechnung
  Annahmen und Ergebnisse 2006.