Stochastic Population Forecast for Germany and its Consequence for the German Pension System

Wolfgang K. Härdle

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

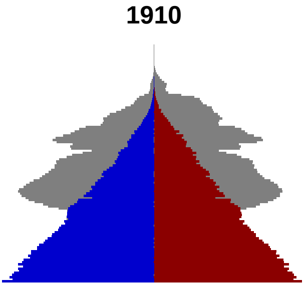
Population forecast needed for many purposes

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

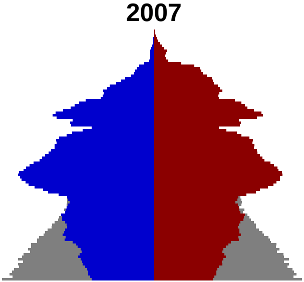
German Pension System

- established 1881 from O. v. Bismark
- pay-as-you-go financing brought 1957 from K. Adenauer:
  "*Kinder kriegen die Leute immer*"
- transfer of the system in 1990 on GDR citizens

German Family in 1910



German Family in 2007



Drastic consequences for the social system!

Conventional Population Projections

- using deterministic models
- 2 or 3 different scenarios for future vital rates
- forecast interval defined by the combination of the scenarios
- this technique faces substantial difficulties (no access probabilities, unrealistic correlations)

Stochastic Projection

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

Stochastic Projection

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

Stochastic Projection

- application of time series models
- modeling and forecasting of the vital rates separately
- demographic variables involving the size and the structure of population
  - Mortality
  - Fertility
  - Migration

Outline

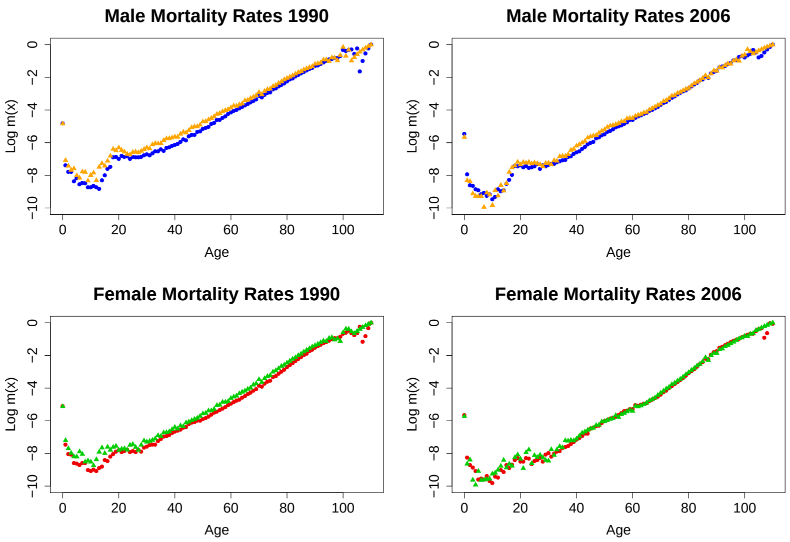
Mortality Data

- Age-specific Mortality Rate (ASMR)
- Source: Human Mortality Database (http://www.mortality.org/)
- Time period: 1956–2006
- Age groups: 0, 1, 2, . . ., 110+
- Data just from former West Germany considered
- Missing data (in the oldest age groups) replaced by a linear interpolation

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

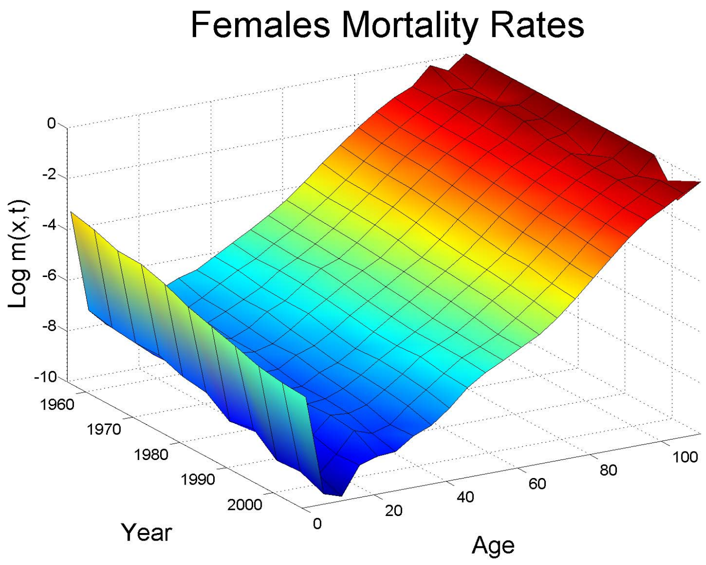
$m_{x,t}$ = total number of deaths per 1000 people of the age x in the time period t

- 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 in 1990 and 2006 for East Germany (triangles) and West Germany (circles).


Males Mortality Rates
Log m(x,t)
0
-2
-4
-6
-8
-10
1960
1970
1980
1990
2000
Year
0
20
40
60
80
100
Age


Females Mortality Rates
0
-2
-4
-6
-8
-10
Log m(x,t)
1960
1970
1980
1990
2000
Year
0
20
40
60
80
100
Age

Lee-Carter Model for Mortality

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

a_x - age specific parameter
$\exp(a_x)$ is the general shape of the mortality 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

Fitting the Model

Assumptions:

$$\sum_{t=t_1}^{t_n} k_t = 0$$

$$\sum_{x=x_0}^{x_p} b_x = 1$$

$$\widehat{a}_x = \frac{1}{T} \sum_{t=t_1}^{t_n} \log(m_{x,t})$$

$$t = \{1956 = t_1, 1957 \ldots, 2006 = t_n\} \ , \ T = t_n - t_1 + 1 \ ,$$
$$x = \{x_0 = 0, \ldots, 110+ = x_p\} \ .$$

Fitting the Model

Assumptions:

$$\sum_{t=t_1}^{t_n} k_t = 0$$

$$\sum_{x=x_0}^{x_p} b_x = 1$$

$$\widehat{a}_x = \frac{1}{T} \sum_{t=t_1}^{t_n} \log(m_{x,t})$$

$$t = \{1956 = t_1, 1957 \ldots, 2006 = t_n\}\ ,\ T = t_n - t_1 + 1\ ,$$
$$x = \{x_0 = 0, \ldots, 110+ = x_p\}\ .$$

Singular Value Decomposition (SVD) to derive k_t and b_x

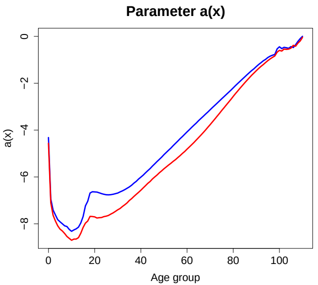
Fitting the Model

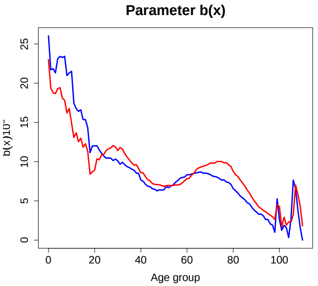
Singular Value Decomposition of Matrix $M(p \times T)$:

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

$\Gamma(p \times r)$ and $\Delta(T \times r)$ - orthonormal: $\Gamma^\top\Gamma = \Delta^\top\Delta = I_r$,
$r = \mathrm{rank}(M)$
$\Lambda = \mathrm{diag}(\lambda_1^{1/2}, \ldots, \lambda_r^{1/2})$, λ_i - eigenvalues of $M^\top M$
$b_x(p \times 1), k_t(T \times 1)$ - first column vectors of matrices Γ and Δ, resp.multiplied by $\lambda_1^{1/2}$

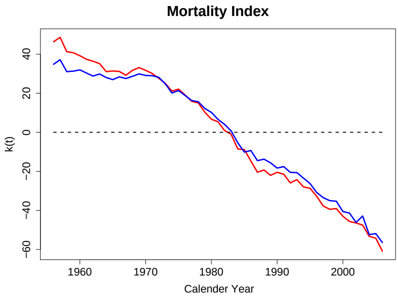
	Males		Females	
Age	a_x	$b_x * 10^3$	a_x	$b_x * 10^3$
0	-4.27	27.24	-4.56	22.99
5	-7.87	24.33	-8.22	19.43
10	-8.27	21.91	-8.70	14.88
15	-7.65	15.58	-8.15	11.85
20	-6.61	11.75	-7.70	8.93
25	-6.73	10.30	-7.67	11.42
30	-6.66	9.97	-7.40	11.40
35	-6.40	8.86	-7.00	10.25
40	-5.98	7.30	-6.57	8.60
45	-5.52	6.23	-6.09	7.20
50	-5.03	6.28	-5.65	6.85
55	-4.55	7.38	-5.26	7.03
60	-4.06	8.26	-4.81	7.84
65	-3.58	8.53	-4.32	9.20
70	-3.11	8.21	-3.77	9.82
75	-2.65	7.53	-3.17	10.00
80	-2.18	6.31	-2.56	8.74
85	-1.72	5.03	-1.98	6.90
90	-1.28	3.44	-1.46	4.66
95	-0.90	2.63	-1.03	3.39
100	-0.43	2.19	-0.60	4.27
105	-0.42	1.96	-0.50	2.29





Mortality Index k_t

Aim: find adequate ARIMA Time Series Model for the forecast



$\longrightarrow$ Random Walk with Drift appropriate for both genders
Unit root test (ADF) for $I(1)$ rejected (p-value < 0.01)

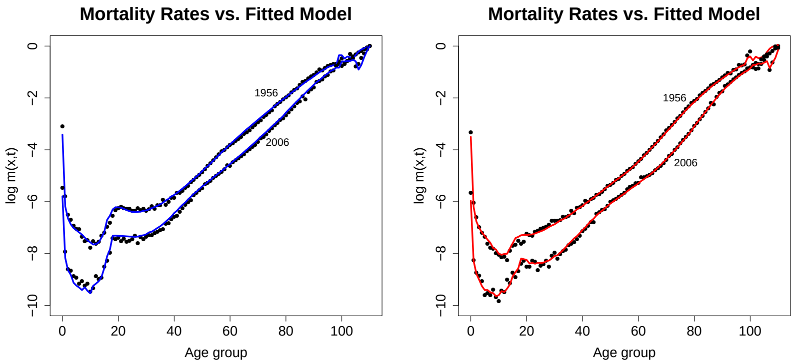
Random Walk with Drift

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

δ - slope of the deterministic trend; $u_t \sim (0, \sigma_u^2)$
The fitted model:

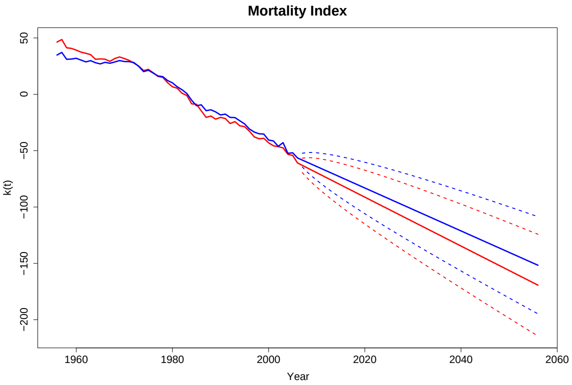
- males: $k_t = -1.83 + k_{t-1} + u_t$; $\widehat{\sigma}_{u^m} = 3.11$
- females: $k_t = -2.14 + k_{t-1} + u_t$; $\widehat{\sigma}_{u^f} = 3.26$
- $\delta =$ the average annual changes in k
- $\widehat{\sigma}_u =$ uncertainty associated with a one-year forecast k_t

The Fitted Model



Actual and fitted mortality rates for all age groups in 1956 and 2006.

Forecast of Mortality Index



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

Life Expectancy

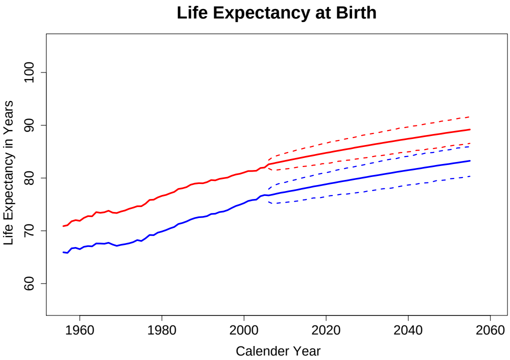
Computing formula:

$$e_{x,t} = \frac{1}{2} + \sum_{y=x}^{x_p-1} \prod_{i=x}^{y} (1 - q_{i,t}) \ ,$$

with the probability of death: $q_{x,t} = \frac{2m_{x,t}}{2+m_{x,t}}$.

Year	Males	Females
1956	65.9	70.9
2006	77.2	82.3
2056	83.2	89.1

Life Expectancy

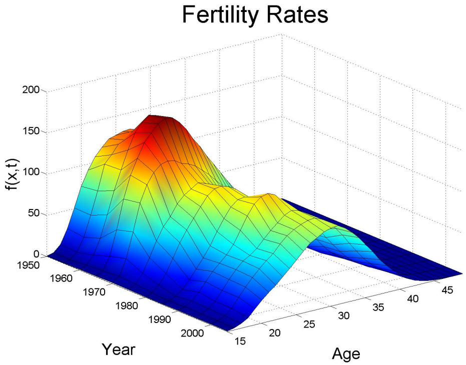


Forecast of the life expectancy at birth for boys and girls with 95% forecast intervals.

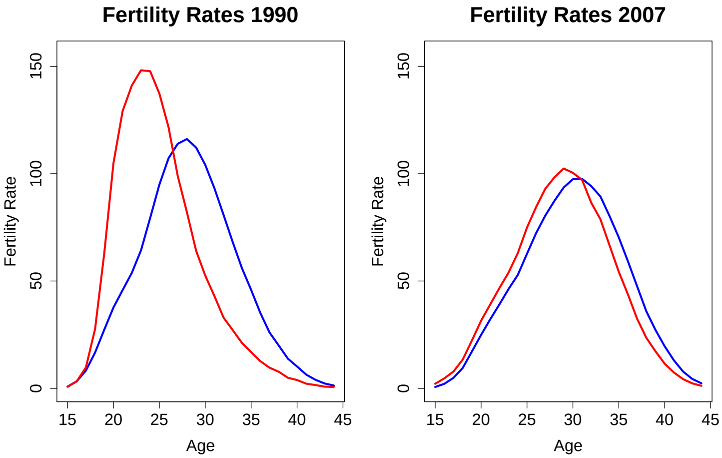
Fertility Data

- Age-specific Fertility Rate (ASFR)
- Source: Federal Statistical Office (www.destatis.de)
- Time period: 1950–2007
- Mothers at the age of: 15, 16, . . ., 49
- Data just from former West Germany 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

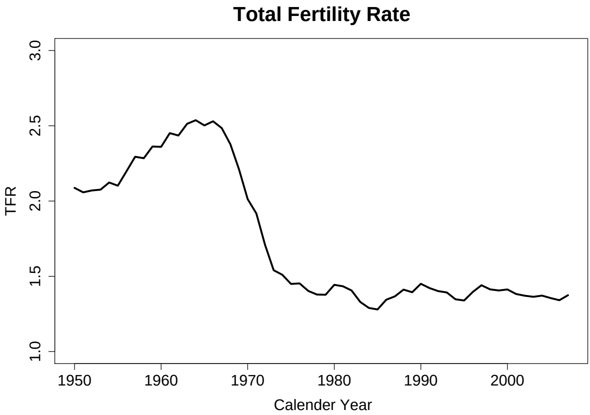


Age-specific Fertility Rates in 1990 and 2007 for East Germany (red) and West Germany (blue).

Total Fertility Rate

$$\mathrm{TFR}_t = 10^{-3} \cdot \sum_{x=15}^{49} f_{x,t}$$

- the average number of children that would be born to one woman over her lifetime if she were to experience the exact current age-specific fertility rates (ASFRs) through her lifetime
- 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



$$\begin{aligned} \max(\mathrm{TFR}) &= 2.54 \quad \text{in } 1964 \\ \min(\mathrm{TFR}) &= 1.28 \quad \text{in } 1985 \end{aligned}$$

Lee-Carter Model for Fertility

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

f_t - time-varying fertility index

a_x, b_x - age-specific parameters

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

Lee-Carter Model for Fertility

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

f_t - time-varying fertility index

a_x, b_x - age-specific parameters

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

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

Lee-Carter Model for Fertility

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

f_t - time-varying fertility index

a_x, b_x - age-specific parameters

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

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

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

Fitting the Model

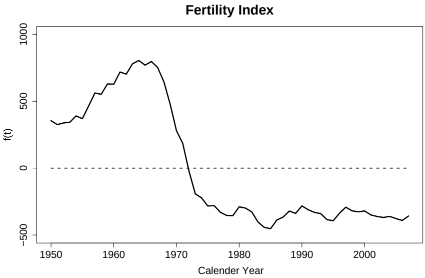
Analogue assumptions to mortality model:

$$\sum_t f_t = 0$$

$$\sum_x b_x = 1$$

$$\widehat{a}_x = \frac{1}{T} \sum_t f_{x,t}, \text{ with } t = \{t_0 = 1950, \dots, t_n = 2007\}$$

SVD to determine 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

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
- $F_t \longrightarrow U$ as $g_t \longrightarrow \infty$
- $F_t \longrightarrow L$ as $g_t \longrightarrow -\infty$
  - modeling of time series g_t
  - consider the data after 1976 (after the pill took effect)

ARMA(1,1) Model

$$g_t = \delta + \phi g_{t-1} + \theta u_{t-1} + u_t \ ,$$

with parameters δ, ϕ, θ and $u_t \sim (0, \sigma_u^2)$.

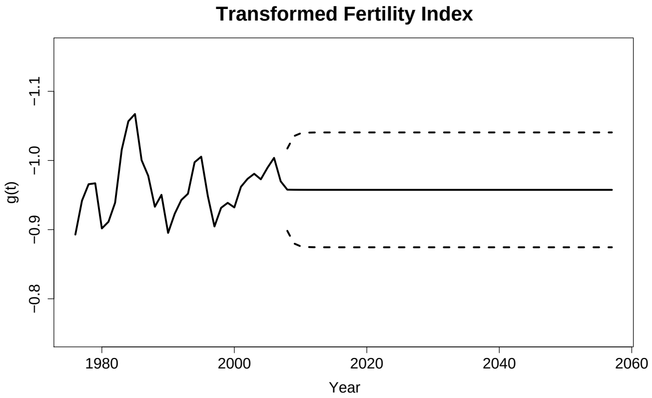
The Fitted Model

with $L = 0$, $U = 5$:

$$g_t = -0.96^* + 0.51^* g_{t-1} + 0.33^* u_{t-1} + u_t \ ,$$

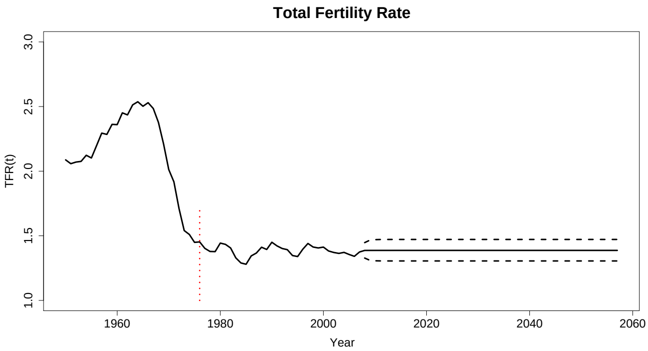
$\widehat{\sigma}_u^2 = 9.22 \cdot 10^{-4}$
Unit root test (ADF) rejected (p-value $= 0.07$)

$*$ = significant at 5%



Transformed fertility index with 95%-confidence interval.

Forecast of TFR



TFR with 95%-confidence interval.

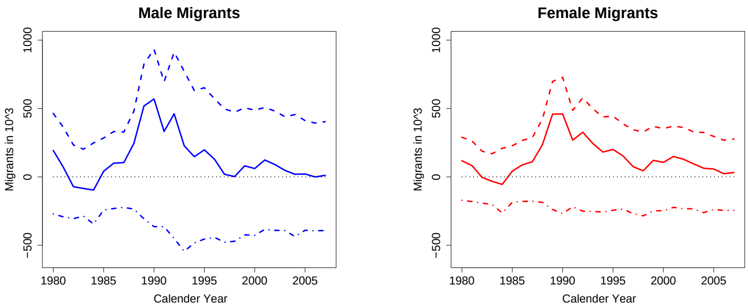
Migration

Influenced by many factors

- migration policy and labour market development in Germany
- political, economic, demographic development in migration countries
- strongly regulated migration in the former German Democratic Republic
- before 1990 only data from former West Germany

Migration Data

- Source: Federal Statistical Office
- Time period: 1980–2007
- Age groups: 0, 1, . . . , 110+
- Data from former West Germany (1980–1989)



Net migration (solid), immigration (dashed) and emigration level (dotdashed).

Model for Migration

- estimate the age density of in-moving and out-moving individuals
- kernel density estimator:

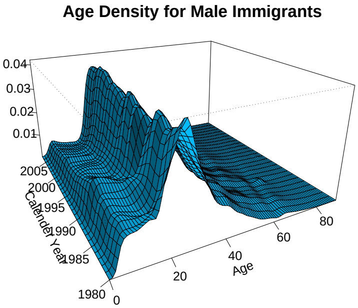
$$\widehat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

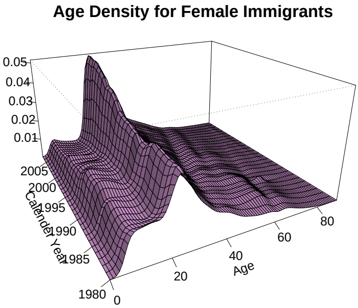
K - kernel function (Gaussian)

h - bandwidth (Silverman's rule of thumb)

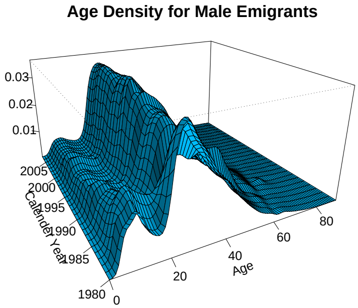
n - number of observations

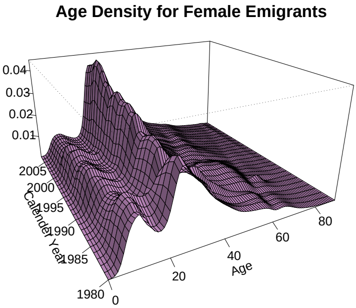
- no significant trend in the age structure of migrants





Estimated density of immigrants.





Estimated density of emigrants.

Forecast of Net Migration

- appropriate time series model for the level of immigrants (data 1994-2007) and emigrants (data 1980-2007)
- forecast of the number of in-moving and out-moving people
- number of immigrants and emigrants in the single age group estimated by the kernel density estimator (from 2007)
- net migrants = immigrated ind. - emigrated ind.

AR(1) Model

$$y_t = \delta + \phi y_{t-1} + \varepsilon_t \ ,$$

with parameters δ, ϕ and $\varepsilon_t \sim (0, \sigma_\varepsilon^2)$.

The Fitted Model

$$\begin{aligned}
I_t^m &= 562.85^* + 0.94^* I_{t-1}^m + \varepsilon_t^m \ , & \widehat{\sigma}_{\varepsilon^m} &= 52.52 \\
I_t^f &= 377.93^* + 0.94^* I_{t-1}^f + \varepsilon_t^f \ , & \widehat{\sigma}_{\varepsilon^f} &= 31.18 \\
E_t^m &= 359.76^* + 0.86^* E_{t-1}^m + e_t^m \ , & \widehat{\sigma}_{e^m} &= 42.70 \\
E_t^f &= 227.55^* + 0.60^* E_{t-1}^f + e_t^f \ , & \widehat{\sigma}_{e^f} &= 26.07 \ .
\end{aligned}$$

Unit root tests (ADF) rejected (p-values $\leq$ 10%).

Cohort-Component Method

For females:

$$\begin{pmatrix} N_{1,t+1} \\ N_{2,t+1} \\ N_{3,t+1} \\ \vdots \\ N_{k,t+1} \end{pmatrix} = \begin{pmatrix} 0 & \dots & s \cdot F_{15,t} & \dots & s \cdot F_{49,t} & 0 \\ P_{1,t} & 0 & \dots & \dots & \dots & 0 \\ 0 & P_{2,t} & \ddots & 0 & \dots & 0 \\ 0 & 0 & \dots & \dots & P_{k-1,t} & 0 \end{pmatrix} \cdot \begin{pmatrix} N_{1,t} \\ N_{2,t} \\ N_{3,t} \\ \vdots \\ N_{k,t} \end{pmatrix} + \begin{pmatrix} NI_{1,t} \\ NI_{2,t} \\ NI_{3,t} \\ \vdots \\ NI_{k,t} \end{pmatrix}$$

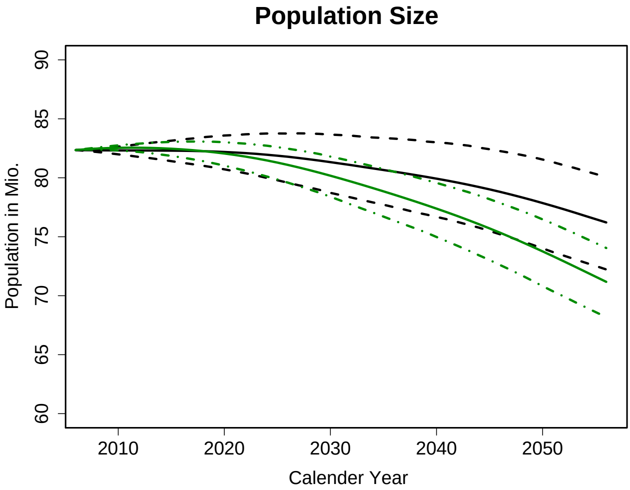
Cohort-Component Method

- $N_{i,t}$ - 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
- $P_{i,t}$ - 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

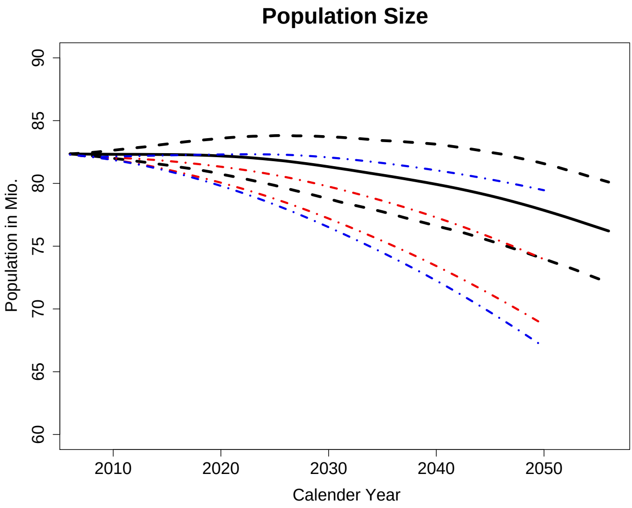
Population Size

Estimation after 1000 simulations compared with the results of the 11th Coordinated Projection of Statistisches Bundesamt

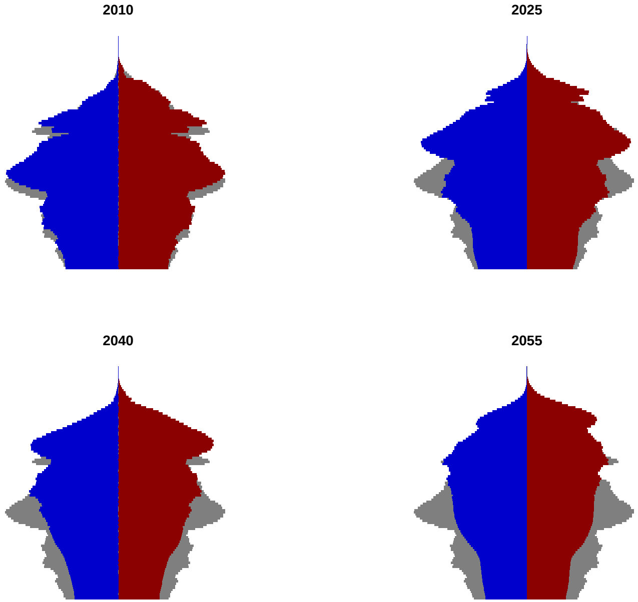
Year	5%-Quantile	Mean	95%-Quantile	L-Bound	U-Bound
2010	81.78	82.52	83.27	81.89	82.04
2025	79.67	82.55	85.30	78.77	80.67
2040	75.83	80.14	84.25	73.42	77.28
2050	72.46	77.48	82.27	68.74	73.96



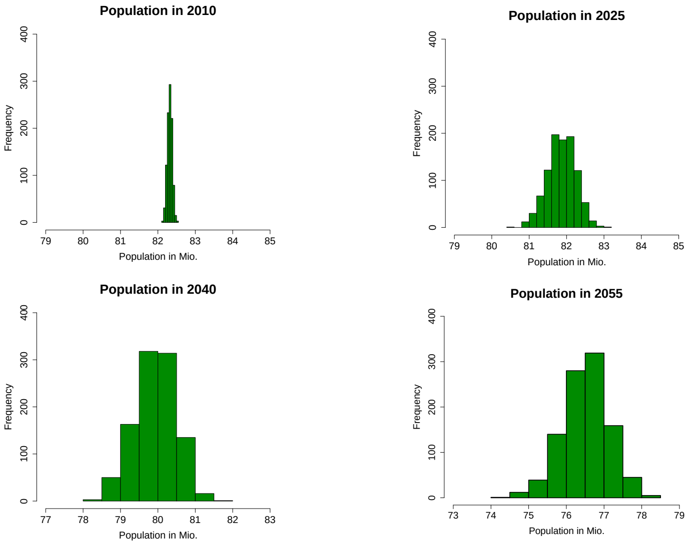
Forecast of the population size compared to the forecast with a deterministic constant migration assumption (green).



Forecast of the population size compared to the forecast of the Federal Statistical Office (red and blue).



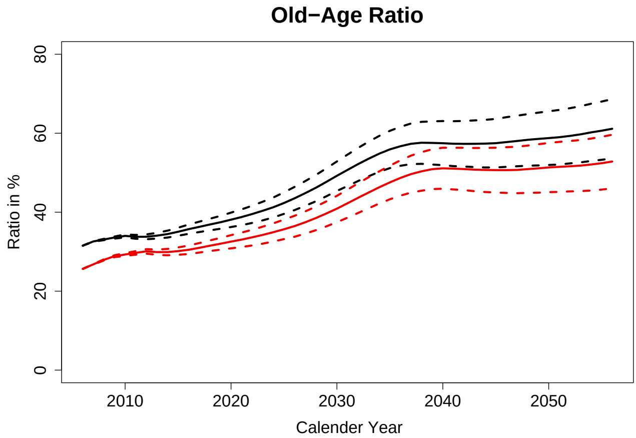
2010
2025
2040
2055


Population in 2010
Population in 2025
Population in 2040
Population in 2055
Frequency
Population in Mio.
0
100
200
300
400
79
80
81
82
83
84
85
77
78
73
74
75
76

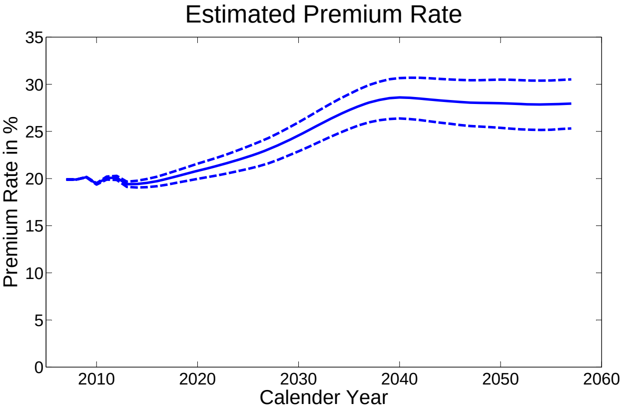
Consequence for the German Pension System

Some facts about the actual system

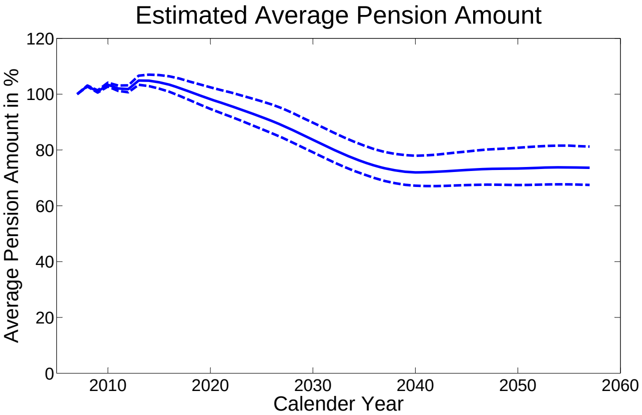
- pay-as-you-go financing
- pension costs in 2006: 190 bn.€
- government subventions in 2006: 55 bn.€
- averaged pension payment amount in 2006: 720€/month (men: 985€ , women: 510€)



Old-age dependency ratio with age limit 65 years (black) and with age limit 67 years (red) with 95% forecast intervals.



Estimated premium rate in the pension system to maintain the pension level.



Estimated level of the average pension amount (720€ in 2007) with constant premium rate.

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.

- A. Renshaw and S. Haberman
  Lee-Carter mortality forecasting: a parallel generalized linear modelling approach for England and Wales mortality projections
  Appliued Statistics, 52:119–137, 2003.

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