# Stochastic modelling of mortality and financial markets

Helena Aro
Teemu Pennanen (PhD instructor)

Department of Mathematics and Systems Analysis,
Aalto University School of Science and Technology

Northern Triangular Seminar
12th April, 2011

# Overview of research

- General objective: pricing and hedging of mortality-linked cash flows
  - Derivatives (e.g. forwards, bonds and swaps) linked to the mortality of a certain population
  - Insurance portfolios, pension fund management

  in incomplete markets
- Stochastic modelling of risk factors
  - Mortality
  - Liabilities
  - Assets
  - Other relevant factors
- Numerical techniques
  - Integration quadratures
  - Numerical optimization

# Mortality-linked cash flows

- Population age structure is shifting towards the old, rendering private insurance companies and governments even more vulnerable to risks of misestimating mortality improvements
- Means of risk management for mortality-linked cash flows is needed
- Mortality-linked securities are financial instruments with cash flows dependent on the mortality of a reference population
- Mortality/longevity markets are developing
- Quantitative methods for risk management of mortality-linked securities are instrumental for the development of such a market

# Mortality-linked cash flows

The value of financial instruments depends essentially on the following subjective factors:

1. **Probability distribution**: description of future development of claims and investment returns, both involving significant uncertainties
   - Dependence between claims and asset returns instrumental in hedging
2. **Risk preferences**: the level of risk at which assets should cover liabilities
3. **Hedging strategy**: investment strategy for the given initial capital
   - Optimal strategy generally cannot be found (cf. Black-Scholes and delta hedging )

# Stochastic modelling of risk factors

- Focus of this talk: modelling the law of a multivariate stochastic process consisting of
  - Mortality
  - Asset returns (interest rates, stock price index)
- Particular emphasis on
  - Connections between mortality and asset returns
  - Long-term development of mortality
- GDP as a link between mortality risk factors and asset returns
- Aro & Pennanen: Stochastic modelling of mortality and its connection with financial markets (manuscript)

# Mortality model

- We employ the general discrete-time stochastic framework for mortality introduced in Aro & Pennanen, *User-friendly approach to stochastic mortality modelling*. European Actuarial Journal, 2011.
- Let $E_{x,t}$ be the size of population aged $[x, x+1)$ (cohort) at the beginning of year $t$
- Denote by $D_{x,t}$ the number of deaths of people aged $[x, x+1)$ during time $[t, t+1)$
- Objective: model the values of $E_{x,t}$ over time $t = 0, 1, 2, \dots$ for a given set $X \subset \mathbb{N}$ of ages
- Assume the conditional distribution of $E_{x+1,t+1} = E_{x,t} - D_{x,t}$ given $E_{x,t}$ is binomial:

$$E_{x+1,t+1} \sim \mathrm{Bin}(E_{x,t}, p_{x,t})$$

where $p_{x,t}$ is the probability that an individual aged $x$ and alive at the beginning of year $t$ is still alive at the end of that year

# Mortality model

- We reduce the dimensionality of $(p_{x,t})_{x \in X}$ by modelling the logistic probabilities by

$$\mathrm{logit}\, p_{x,t} := \ln\left(\frac{p_{x,t}}{1-p_{x,t}}\right) = \sum_{i=1}^{n} v_t^i \phi_i(x),$$

  where $\phi_i(x)$ are user-defined basis functions across cohorts, and $v_t^i$ stochastic risk factors that vary over time
- In other words, $p_{x,t} = p_{v(t)}(x)$, where $v(t) = (v_1(t), \ldots, v_n(t))$, and $p_v : X \to (0,1)$ is the parametric function defined for each $v \in \mathbb{R}^n$ by

$$p_v(x) = \frac{\exp\left(\sum_{i=1}^n v_i \phi_i(x)\right)}{1 + \exp(\sum_{i=1}^n v_i \phi_i(x))}$$

- Modelling the logit transforms instead of $p_{x,t}$ directly guarantees that $p_{x,t} \in (0,1)$.

# Mortality model

- Certain desired properties of $p_{x,t}$, e.g. continuity with respect to $x$, are achieved by corresponding choices of $\phi_i(x)$
  - Incorporation of user preferences and/or population-specific characteristics
- Choice of basis functions assigns interpretations to risk factors
- Concrete interpretations facilitate the modelling of risk factors, which is advantageous the risk management of mortality-linked instruments
- Vector $v_t$ of risk factors is modelled as a stochastic process, based on historical values, expert opinions, or both
- Historical values of $v_t$ are constructed by maximum likelihood estimation, log-likelihood function is strictly concave

# Mortality model

- We consider female mortality dynamics of six large OECD countries: Australia, Canada, France, Japan, UK, US
- Data consists of annual values of cohort sizes $E_{x,t}$ and numbers of deaths $D_{x,t}$ for each country, covering years 1950–2006 (Source: Human mortality database)
- A model with three basis functions and three risk factors is employed

# Mortality model

- The model:

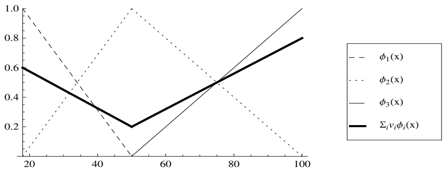
$$\text{logit } p_{x,t} = v_t^1 \phi_1(x) + v_t^2 \phi_2(x) + v_t^3 \phi_3(x),$$

where basis functions are piecewise linear:

$$\phi_1(x) = \begin{cases} 1 - \frac{x-18}{32} & \text{for } x \leq 50 \\ 0 & \text{for } x \geq 50, \end{cases} \quad \phi_2(x) = \begin{cases} \frac{1}{32}(x-18) & \text{for } x \leq 50 \\ 2 - \frac{x}{50} & \text{for } x \geq 50, \end{cases}$$

$$\phi_3(x) = \begin{cases} 0 & \text{for } x \leq 50 \\ \frac{x}{50} - 1 & \text{for } x \geq 50. \end{cases}$$

- The linear combination now also piecewise linear:




- Interpretation: values of $v_t^i$ points on the logit $p_{x,t}$ curve

# Statistical analysis of risk factors: $v_1$

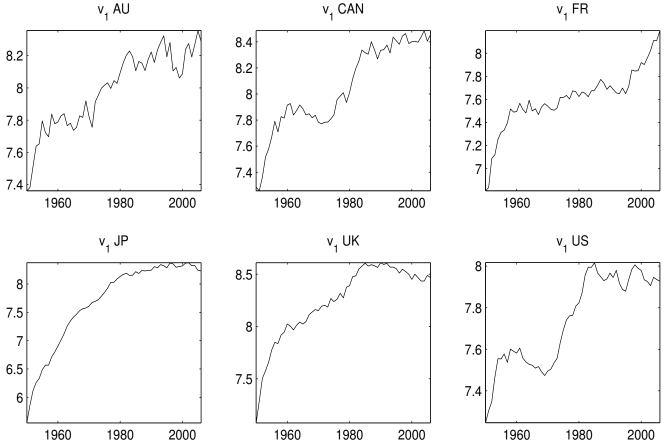



Figure: Historical values for risk factor $v_1$, females. Note the different scales.

# Statistical analysis of risk factors: $v_1$

- Fundamental question: how long can mortality keep improving?
- Mortality risk factors traditionally modelled as random walk with a negative drift, i.e. mortality will decline indefinitely
- Some experts predict that the steady decline in overall mortality during past decades will continue for the foreseeable future, while others suggest that human life expectancy might even decrease.
- In several sample countries, the historical values of $v_t^1$ display a stabilizing tendency
- In order to analyse this phenomenon, we fit the following regression into historical data:

$$\Delta v_t^1 = b + a v_{t-1}^1 + \varepsilon_t = -a\left(-\frac{b}{a} - v_{t-1}^1\right) + \varepsilon_t$$

# Statistical analysis of risk factors: $v_1$

Table: Parameter estimates, t-values and summary statistics for $\Delta v_t^1 = b + av_{t-1}^1$.

| | AU | CAN | F | JP | UK | US |
|---|---|---|---|---|---|---|
| $b$ | 0.778 | 0.446 | 0.569 | 0.614 | 0.902 | 0.437 |
| p-value(t-stat.) | 0.017 | 0.028 | 0.020 | 0.000 | 0.000 | 0.015 |
| $a$ | -0.095 | -0.053 | -0.072 | -0.074 | -0.107 | -0.055 |
| p-value(t-stat.) | 0.020 | 0.036 | 0.026 | 0.000 | 0.000 | 0.018 |
| $R^2$ | 0.097 | 0.079 | 0.089 | 0.627 | 0.461 | 0.100 |
| Adj. $R^2$ | 0.080 | 0.062 | 0.072 | 0.620 | 0.451 | 0.083 |
| p-value(F-Statistic) | 0.020 | 0.036 | 0.026 | 0.000 | 0.000 | 0.018 |

# Statistical analysis of risk factors: $v_1$

Table: Residual test statistics for the regression $\Delta v_t^1 = b + a v_{t-1}^1$.

| | AU | CAN | F | JP | UK | US |
|---|---|---|---|---|---|---|
| Serial correlation (BG) | 0.329 | 0.533 | 0.123 | 0.731 | 0.084 | 0.044 |
| Normality (JB) | 0.450 | 0.364 | 0.170 | 0.484 | 0.692 | 0.481 |
| Heteroskedasticity (BP) | 0.546 | 0.307 | 0.074 | 0.013 | 0.029 | 0.263 |
| ADF t-statistic | -9.191 | -8.374 | -8.105 | -7.11 | -3.722 | -4.474 |
| p-value | $< 0.01$ | $< 0.01$ | $< 0.01$ | $< 0.01$ | $< 0.01$ | $<0.01$ |
| KPSS level | 0.219 | 0.164 | 0.394 | 0.086 | 0.255 | 0.139 |
| p-value | $> 0.1$ | $> 0.1$ | 0.0796 | $> 0.1$ | $> 0.1$ | $>0.1$ |

*Notes.* Serial correlation is tested with the Breusch–Godfrey tests for serial correlation. A small p-value suggests rejecting the zero hypothesis of serial correlation. Normality is tested with the Jarque–Bera test for normality. The Breusch-Pagan tests against heteroskedasticity, with the null hypothesis of homoskedasticity. Stationarity of the residuals is tested with the augmented Dickey-Fuller unit root test (ADF) and Kwiatkowski-Phillips-Schmidt-Shin test (KPSS).

# Statistical analysis of risk factors: $v_2$

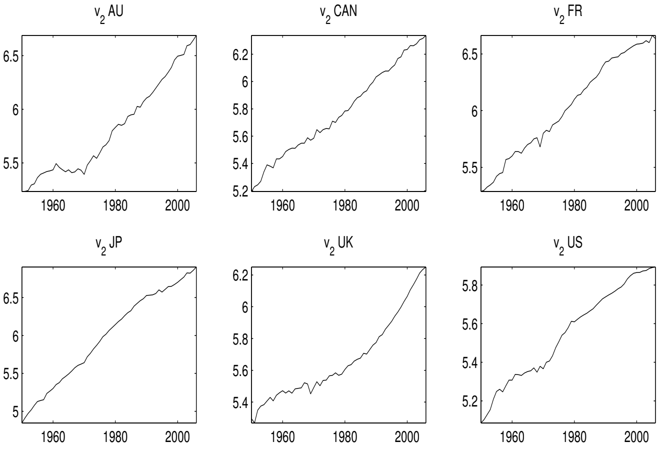



Figure: Historical values for risk factor $v_2$, females. Note the different scales.

# Statistical analysis of risk factors: $v_2$

- Dramatic reduction in coronary disease amongst the middle-aged during the past three decades shows in the rapid growth of risk factor $v_2$ reflecting the survival probability of 50-year-olds
- To which extent do improvements in treatment and possible further reductions in smoking can outweigh the detrimental effects of obesity and other lifestyle-related factors?
- Even thought the historical values (apart from Japan) do not yet show signs of levelling out, eventual stabilizing behaviour similar to that of $v_1$ may also be a future possibility for $v_2$
- To quantify the rate of improvement for $v_2$, we fit the regression

$$\Delta v_t^2 = b + \varepsilon_t,$$

# Statistical analysis of risk factors: $v_3$ and GDP

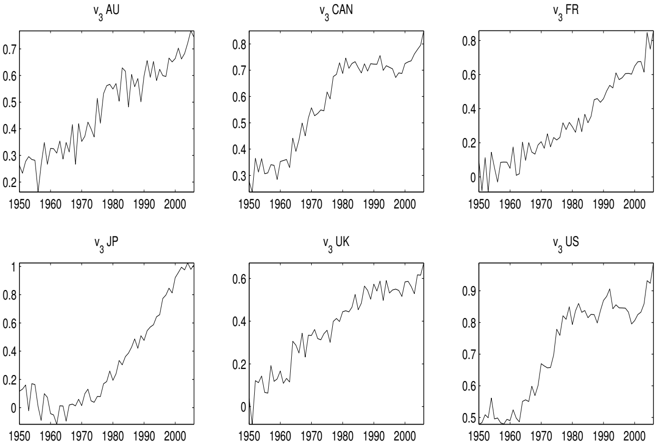



Figure: Historical values for risk factor $v_3$, females. Note the different scales.

# Statistical analysis of risk factors: $v_3$ and GDP

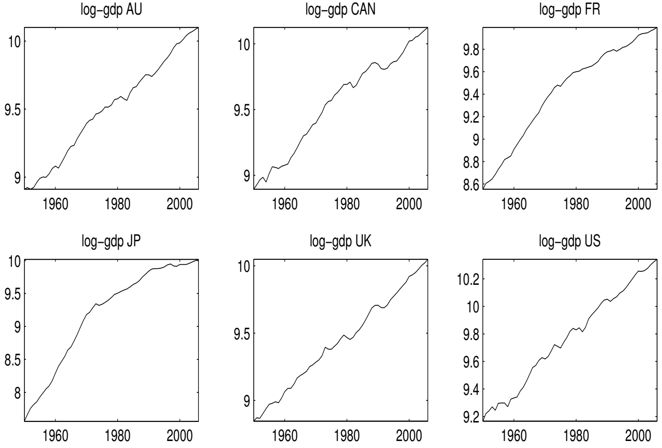



Figure: Historical values for risk factor logarithm of per capita GDP. Note the different scales.

## Statistical analysis of risk factors: $v_3$ and GDP

- Risk factor $v_3$ describes old-age mortality, with which the cash flows of mortality linked instruments are connected
- Long-term dependence between GDP and mortality has been observed (e.g. Preston 1975, Preston 2007).
- Similarities in the general shape of their plots support the observation that $v_3$ and log-per capita GDP may move together in the long run
- We analyse the dependence of $v_3$ on GDP with the regression

$$\Delta v_t^3 = b + a_1 v_{t-1}^3 + a_2 g_{t-1} + \varepsilon_t,$$

where $g_t$ is the log-per capita GDP

# Statistical analysis of risk factors: $v_3$ and GDP

Table: Parameter estimates, t-values and summary statistics for the regression $\Delta v_t^3 = b + a_1 v_{t-1}^3 + a_2 g_{t-1}$

.

| | AU | CAN | F | JAP | UK | US |
|---|---|---|---|---|---|---|
| $b$ | -3.062 | -1.686 | -1.826 | -0.332 | -2.894 | -0.664 |
| p-value (t-stat.) | 0.000 | 0.000 | 0.001 | 0.053 | 0.000 | 0.020 |
| $a_1$ | -0.877 | -0.455 | -0.388 | -0.057 | -0.656 | -0.189 |
| p-value (t-stat.) | 0.000 | 0.000 | 0.001 | 0.1941 | 0.000 | 0.015 |
| $a_2$ | 0.367 | 0.205 | 0.208 | 0.040 | 0.334 | 0.083 |
| p-value (t-stat.) | 0.000 | 0.000 | 0.001 | 0.044 | 0.000 | 0.016 |
| | | | | | | |
| $R^2$ | 0.454 | 0.255 | 0.195 | 0.076 | 0.335 | 0.109 |
| Adj. $R^2$ | 0.433 | 0.227 | 0.164 | 0.041 | 0.310 | 0.076 |
| p-value (F-Statistic) | 0.000 | 0.0004 | 0.003 | 0.124 | 0.000 | 0.046 |

# Statistical analysis of risk factors: $v_3$ and GDP

Table: Residual test statistics for $\Delta v_t^3 = b + a_1 v_{t-1}^3 + a_2 g_{t-1}$.

| | 1950–2006 | | | | | |
|---|---|---|---|---|---|---|
| | AU | CAN | F | JAP | UK | US |
| Serial correlation (BG) | 0.000 | 0.002 | 0.000 | 0.000 | 0.090 | 0.26 |
| Normality (JB) | 0.536 | 0.279 | 0.009 | 0.203 | 0.880 | 0.533 |
| Heteroskedasticity (BP) | 0.330 | 0.038 | 0.101 | 0.001 | 0.012 | 0.734 |
| ADF t-statistic | -7.356 | -3.921 | -1.172 | -10.599 | -6.843 | -8.499 |
| p-value | $< 0.01$ | $< 0.01$ | $> 0.1$ | $< 0.01$ | $< 0.01$ | $< 0.01$ |
| KPSS level | 0.120 | 0.311 | 0.463 | 0.084 | 0.263 | 0.150 |
| p-value | $> 0.1$ | $> 0.1$ | 0.050 | $> 0.1$ | $> 0.1$ | $> 0.1$ |

*Notes.* Serial correlation is tested with the Breusch–Godfrey tests for serial correlation. A small p-value suggests rejecting the zero hypothesis of serial correlation. Normality is tested with the Jarque–Bera test for normality. The Breusch-Pagan tests against heteroskedasticity, with the null hypothesis of homoskedasticity. Stationarity of the residuals is tested with the augmented Dickey-Fuller unit root test (ADF) and Kwiatkowski-Phillips-Schmidt-Shin test (KPSS).

# Statistical analysis of risk factors: GDP and financial markets

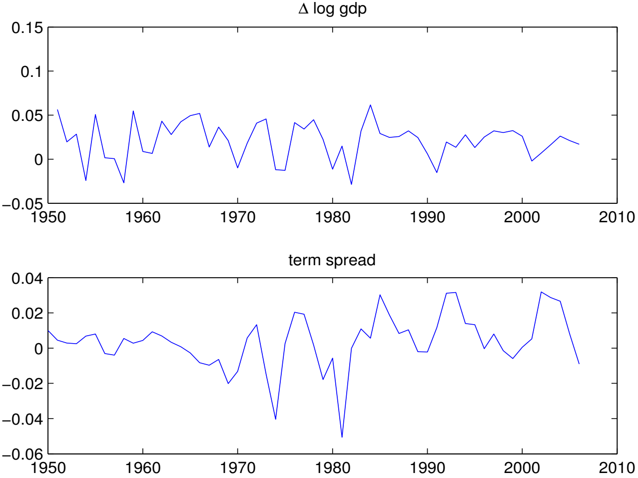



Figure: Differences of US log per capita GDP and term spread

# Statistical analysis of risk factors: GDP and financial markets

- GDP is linked with financial markets (extensive review article: Stock&Watson 2003)
  - Yield curve/term spread
  - Credit spread
  - Stock prices
- As in e.g. Wheelock&Wohar(2009), we analyse the connection between $g_t$ and term spread $s_t$ with the regression

$$\Delta g_t = b + as_{t-1} + \varepsilon_t,$$

where $s_t$ is the difference between the 10-year Treasury rate and the 6-month Certificate of Deposit rate

# Statistical analysis of risk factors: GDP and financial markets

Table: Parameter estimates, t-values and summary statistics for the regression $\Delta g_t = b + as_{t-1}$.

| | | | |
|---|---|---|---|
| $b$ | 0.0185 | Serial Correlation (BG) | 0.4498 |
| p-value(t-statistic) | 0.0000 | Normality (JB) | 0.7417 |
| $a$ | 0.6460 | Heteroskedasticity (BP) | 0.9257 |
| p-value(t-statistic) | 0.0004 | | |
| | | ADF t-statistic | -6.85 |
| $R^2$ | 0.2117 | p-value | $< 0.01$ |
| Adj. $R^2$ | 0.1971 | KPSS level | 0.3374 |
| p-value(F-Statistic) | 0.0004 | p-value | $> 0.1$ |

# Statistical analysis of risk factors: GDP and financial markets

- We apply the established drifting Brownian motion approach, and fit the process

$$\Delta p_t = b + \varepsilon_t$$

into the logarithm of S&P Composite stock price index data $p_t$ for the same observation period

- Residuals of the above equation and of

$$\Delta g_t = b + a s_{t-1} + \varepsilon_t$$

have a correlation coefficient of 0.4843 (significance below 0.001)

# Modelling the risk factors

- Based on our observations, we propose the linear stochastic difference equation model for US:

$$\Delta x = Ax + b + \varepsilon_t$$

for $x = [v_t^1, v_t^2, v_t^3, g_t, s_t, p_t]$ and multivariate Gaussian $\varepsilon_t$,

- The equations are

$$\begin{aligned}
\Delta v_t^1 &= a_{11} v_t^1 + b_1 + \varepsilon_t^1 \\
\Delta v_t^2 &= b_2 + \varepsilon_t^2 \\
\Delta v_t^3 &= a_{33} v_t^3 + a_{34} g_t + b_3 + \varepsilon_t^3 \\
\Delta g_t &= a_{45} s_t + b_4 + \varepsilon_t^4 \\
\Delta s_t &= a_{55} s_t + b_5 + \varepsilon_t^5 \\
\Delta p_t &= b_6 + \varepsilon_t^6
\end{aligned}$$

where $a_{i,j}$ and $b_j$ are components of matrix $A$ and vector $b$

# Simulations

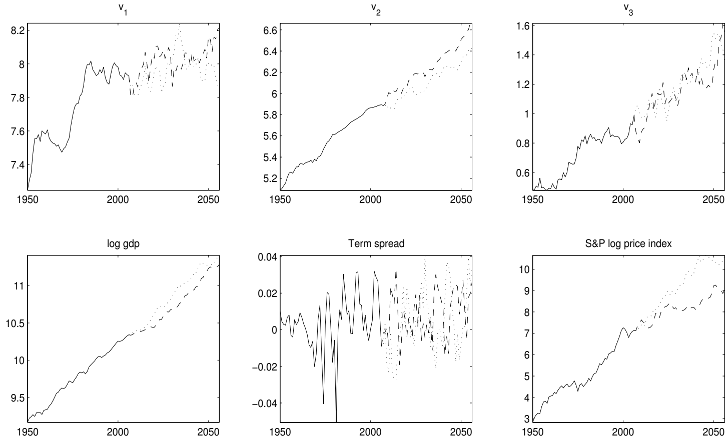



Figure: Simulation sample paths, US females.

# Simulations

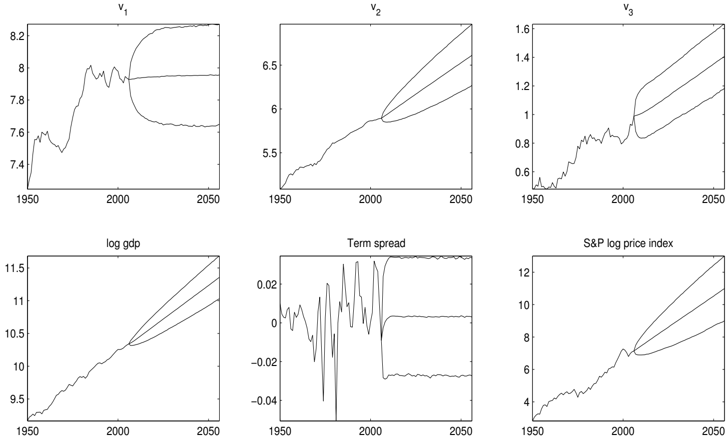



Figure: Medians and 95% confidence intervals for MC simulations (N=10000), US females.

# Conclusions

- Recent development of markets for mortality-linked securities has created a need for quantitative methods for their risk management
- We present a relatively simple model connecting mortality and economic conditions, which provides a solid starting point for solving pricing and hedging problems of mortality-linked cash flows
- Next step: how to choose the investment strategy, given the model and its set of assets?