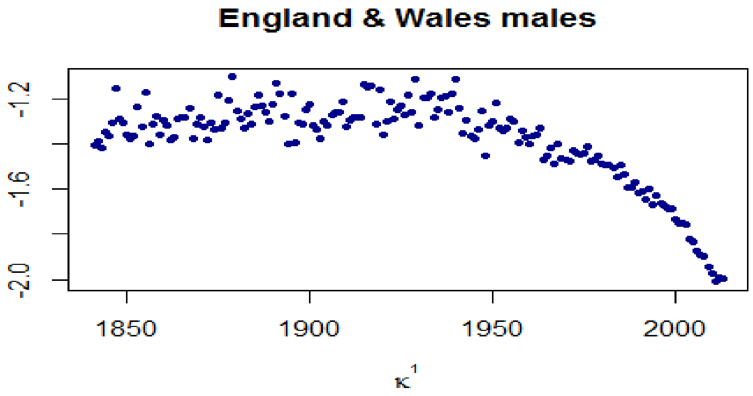
A Set of new Stochastic Trend Models

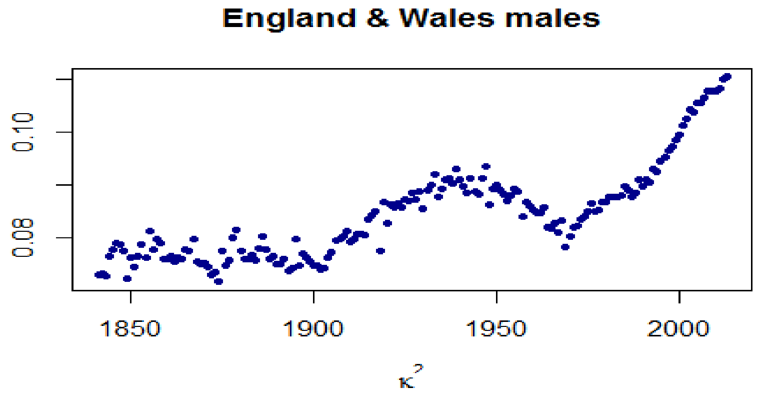
- Johannes Schupp

- Longevity 13, Taipei, 21th-22th September 2017

Introduction

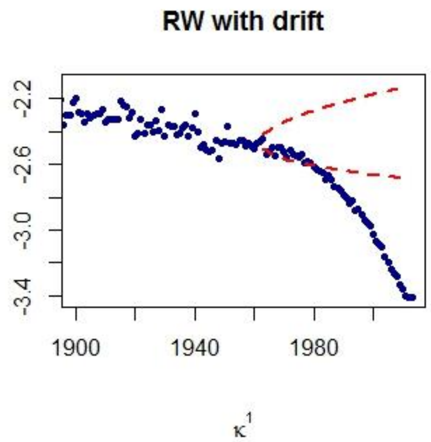
- Uncertainty about the evolution of mortality
  - Measure longevity risk in pension or annuity portfolios with stochastic mortality models
- Parametric mortality models: Lee-Carter model, Cairns-Blake-Dowd model, APC model, etc.
- Reduce the information about exposures and deaths to a few parameters:
  - CBD: Two time dependent parameter processes (Cairns et al. (2006)): $\log\left(\frac{q_{x,t}}{1-q_{x,t}}\right) = \kappa_t^1 + \kappa_t^2 \cdot (x - \bar{x})$
  - Parameter processes calibrated for English and Welsh males older than 65 years
- $L(\kappa_t^1, \kappa_t^2) \to max$ with the assumption of $D_{x,t} \sim Poi(E_{x,t} \cdot \widehat{m}_{x,t})$ or $D_{x,t} \sim Bin(E_{x,t}, \hat{q}_{x,t})$





Introduction

- Popular choice: a (multivariate) random walk with drift (RWD) for stochastic forecasts
- Backtesting in 1963 based on a 10-year calibration:
  - Future observations far outside the 99% quantile
- Historic trend **changed** once in a while
  - Only a piecewise linear trend
  - Random changes in the trends slope
  - Random fluctuation around the prevailing trend
- In principle, our approach can be applied to any parametric mortality model
- Extrapolating only the most recent trend, systematically underestimates future uncertainty, see e.g. Sweeting (2011), Li et al. (2011), Börger et al. (2014)



Agenda

- **Introduction**
- **Specification of a Stochastic model**
  - Trend component
  - Drift component
- **Parameter estimation**
  - Three alternative approaches
  - Open issues

Stochastic Trend model

- Continuous piecewise linear trend, with random changes in the slope and random fluctuation around the prevailing trend
  - Model the trend process with random noise → $\kappa_t = \hat{\kappa}_t + \epsilon_t;\ \epsilon_t \sim f$
  - Extrapolate the most recent actual mortality trend → $\hat{\kappa}_t = \hat{\kappa}_{t-1} + d_t$
  - In every year, there is a possible change in the mortality trend with probability p
  - In the case of a trend change $\lambda_t = M_t \cdot S_t$
    - With absolute magnitude of the trend change $M_t \sim h$
    - Sign of the trend change S_t bernoulli distributed with values -1, 1 each with probability $\frac{1}{2}$

RWD

$\hat{\kappa}_{t-1} \rightarrow \kappa_{t-1}$

→ $d_t = d_{t-1} + \lambda_t$, where $\lambda_t = \left\{\begin{matrix} 0 & \textit{with probability } 1-p \\ M_t \cdot S_t & \textit{with probability } p \end{matrix}\right\} \sim g$

- In principle, also other distributions are possible (Pareto, Normal, t-distribution,…)
  - We propose to use: $\mathbf{f} = \mathcal{N}\left(0, \sigma^2_{\epsilon,t}\right),\ \ \mathbf{h} = \mathcal{LN}(\mu, \sigma^2)$

Parameters to be estimated for projections starting in t=0 (typically latest observation, case: $\mathbf{f} = \mathcal{N}\left(0, \sigma^2_{\epsilon,t}\right),\ \mathbf{h} = \mathcal{LN}(\mu, \sigma^2)$**):**

$p, \sigma^2_{\epsilon,t}, \mu, \sigma^2, d_0, \hat{\kappa}_0$

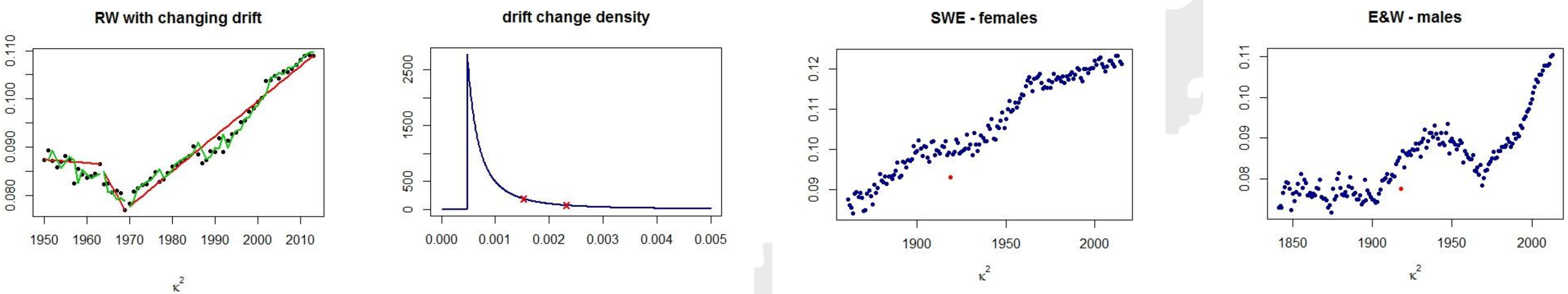
RWD

$p, \sigma^2_{\epsilon,t}, \mu, \sigma^2, d_0$

Parameter estimation
Alternative I

Calibration based on historic trends

- Use historic trends/drifts to estimate parameters (see e.g. Hunt and Blake (2014), Sweeting (2011), Börger and Schupp (2015)). Choose optimal historic trends/drifts based on some optimizing criterion (OLS, Likelihood,…). Advantage: Intuitive historic curves
  - Börger and Schupp (2015): For $k \in 0, \dots, m$ find trend process $d_0, \hat{\kappa}_0, \lambda_{-N+2}, \dots, \lambda_0$ where exactly k of $\lambda_{-N+2}, \dots, \lambda_0$ are unequal to zero (trend curve with k trend changes). Update σ_ϵ^2 iteratively. Choose optimal trend process with AIC/BIC/MBIC.
  - Example: Random Walk with changing drift (in the spirit of Hunt and Blake (2014))



Possible Problems: historic observations are unlikely to be generated with the drift change density (→inconsistent prediction possible), only few observations. Outliers can have a huge influence

Parameter estimation

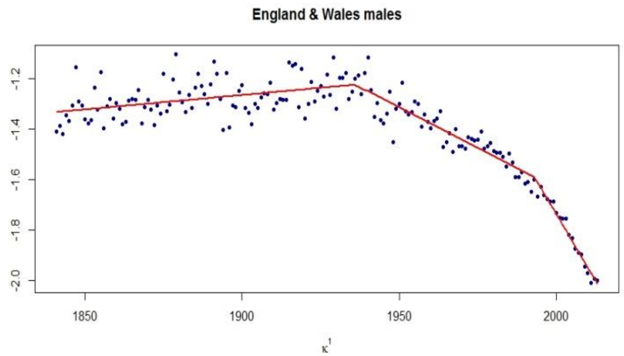
Alternative II

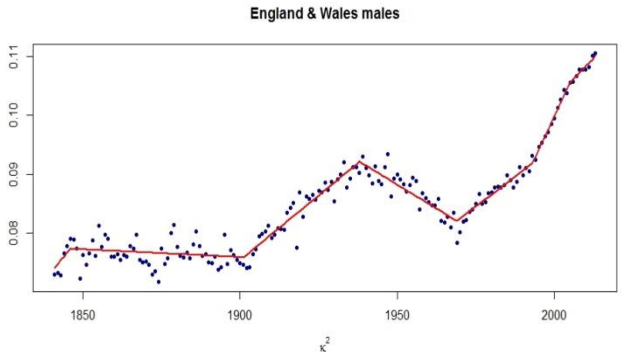
Calibration based on historic trends with a combined likelihood

- Include the distribution of the trend changes used for simulations in the optimization criterion
- Calibrate optimal historic trends based on $f_{\mathcal{N}}\left(\kappa^{i}_{-N,\dots,0} \middle| \sigma_{\epsilon}^{2}, d_0, \hat{\kappa}_0, \lambda_{-N+2}, \dots, \lambda_0\right) \cdot g(\lambda_{-N+2}, \dots, \lambda_0 | \mu, \sigma^2, p)$
  - For $k \in 1, \dots, m$ find trend process $d_0, \hat{\kappa}_0, \lambda_{-N+2}, \dots, \lambda_0$ that maximizes

  $f_{\mathcal{N}}\left(\kappa^{i}_{-N,\dots,0} \middle| \sigma_{\epsilon}^{2}, d_0, \hat{\kappa}_0, \lambda_{-N+2}, \dots, \lambda_0\right) \cdot g(\lambda_{-N+2}, \dots, \lambda_0 | \mu, \sigma^2, p)$, where exactly k of $\lambda_{-N+2}, \dots, \lambda_0$ are unequal to zero (trend curve with k trend changes). Update $\sigma_{\epsilon}^{2}, p, \mu, \sigma^2$ iteratively.

  - Based on optimal goodness of fit $\left(f_{\mathcal{N}}\left(\kappa^{i}_{-N,\dots,0} \middle| \sigma_{\epsilon}^{2}, d_0, \hat{\kappa}_0, \lambda_{-N+2}, \dots, \lambda_0\right)\right)$ choose optimal historic trend
- Advantages: Consistency between historic trends and stochastic simulation, avoid rather subjective selection with information criteria
- The parameters required for stochastic forecasts are part of the calibration: p, σ_{ϵ}^{2}, $\mu, \sigma^2, d_0, \hat{\kappa}_0$





Parameter estimation

Alternative III

Calibration based on MLE

- Stochastic forecasts require: $\mu, \sigma^2, p, \sigma_\epsilon^2, \widehat{\kappa_0^i}, d_0$. Not necessarily a historic trend required. The focus here will be solely on forecasts!
- Idea: Classic MLE: $L\left(\mu, \sigma^2, p, \sigma_\epsilon^2, \widehat{\kappa_0^i}, d_0 \,\middle|\, \kappa^i\right) \to \max \quad i = 1,2$
- Example: Consider last three years and one index:

$$\hat{\kappa}_t = \hat{\kappa}_{t-1} + d_t$$
$$d_t = d_{t-1} + \lambda_t$$

  - Known trend in 0, unknown trend in -1 (possible trend change λ_0)
  - $L(\mu, \sigma^2, p, \sigma_\epsilon^2, \hat{\kappa}_0, d_0 | \kappa_{-2}, \kappa_{-1}, \kappa_0)$
  - $= f_{\mathcal{N}}(\kappa_0 - \widehat{\kappa_0} | \sigma_\epsilon^2, \widehat{\kappa_0}) \cdot f_{\mathcal{N}}(\kappa_{-1} - (\widehat{\kappa_0} - d_0) | \sigma_\epsilon^2, \widehat{\kappa_0}, d_0) \cdot (f_{\mathcal{N}} * g)(\kappa_{-2} | \mu, \sigma^2, p, \sigma_\epsilon^2, \widehat{\kappa_0}, d_0)$
  - $= f_{\mathcal{N}}(\epsilon_0 | \sigma_\epsilon^2, \widehat{\kappa_0}) \cdot f_{\mathcal{N}}(\epsilon_{-1} | \sigma_\epsilon^2, \widehat{\kappa_0}, d_0) \cdot \int_{\mathbb{R}} g(\lambda_0 | \mu, \sigma^2, p) \cdot f_{\mathcal{N}}(\kappa_{-2} - (\widehat{\kappa_0} - d_0 - d_{-1}) | \sigma_\epsilon^2, \widehat{\kappa_0}, d_0) d\lambda_0 \to \max$
  - $= f_{\mathcal{N}}(\epsilon_0 | \sigma_\epsilon^2, \widehat{\kappa_0}) \cdot f_{\mathcal{N}}(\epsilon_{-1} | \sigma_\epsilon^2, \widehat{\kappa_0}, d_0) \cdot \int_{\mathbb{R}} g(\lambda_0 | \mu, \sigma^2, p) \cdot f_{\mathcal{N}}(\kappa_{-2} - (\widehat{\kappa_0} - d_0 - (d_0 - \lambda_0)) | \sigma_\epsilon^2, \widehat{\kappa_0}, d_0) d\lambda_0 \to \max$
- Knowing $\mu, \sigma^2, p, \sigma_\epsilon^2, \widehat{\kappa_0^i}, d_0$, we can give a likelihood function for the historic data

Parameter estimation
Alternatives III

- Consider the complete history:
- $L(\theta | \kappa^i_{-N,\ldots,0}) \to \max$ with $\theta \coloneqq \mu, \sigma^2, p, \sigma_\epsilon^2, \hat{\kappa}_0, d_0$
- We can calculate the trend process recursively $\widehat{\kappa_{-s}} = \widehat{\kappa_0} - sd_0 + \sum_{l=1}^{s-1} l \cdot \lambda_{-(s-1-l)}$, $0 \leq s$
  - $L\left(\mu, \sigma^2, p, \sigma_\epsilon^2, \widehat{\kappa_0^i}, d_0 \middle| \kappa^i_{-N,\ldots,0}\right) = f_{\mathcal{N}}(\epsilon_0 | \sigma_\epsilon^2, \widehat{\kappa_0}) \cdot f_{\mathcal{N}}(\epsilon_{-1} | \sigma_\epsilon^2, \widehat{\kappa_0}, d_0)$

$$\cdot \int_{\mathbb{R}^{N-1}} \prod_{s=2}^{N} g(\lambda_{-(s-2)} | \theta) \cdot f_{\mathcal{N}}\left(\kappa^i_{-s} - (\widehat{\kappa_0} - sd_0 + \sum_{l=1}^{s-1} l \cdot \lambda_{-(s-1-l)}) | \theta\right) d\lambda_{-N+2,\ldots,0} \to max$$

- Challenge: In parameter calibration, we need to solve and optimize this N-1 dimensional integral

RWD

$$f_{\mathcal{N}}\left(\kappa^i_{-s} - (\widehat{\kappa_0} - sd_0 + \sum_{l=1}^{s-1} l \cdot \lambda_{-(s-1-l)}) | \theta\right) \to f_{\mathcal{N}}\left(\kappa^i_{-s} - (\kappa^i_{-s+1} - d_0 - \sum_{l=1}^{s-1} \lambda_{-(s-1-l)}) | \theta\right)$$

Parameter estimation
Alternatives III

- Likelihood of the trend model
  - $L\left(\mu, \sigma^2, p, \sigma_\epsilon^2, \widehat{\kappa_0^i}, d_0 \,\middle|\, \kappa_{-N,\ldots,0}^i\right) = f_{\mathcal{N}}(\epsilon_0 | \sigma_\epsilon^2, \widehat{\kappa_0}) \cdot f_{\mathcal{N}}(\epsilon_{-1} | \sigma_\epsilon^2, \widehat{\kappa_0}, d_0)$

$$\cdot \int_{\mathbb{R}^{N-1}} \prod_{s=2}^{N} g\left(\lambda_{-(s-2)} | \theta\right) \cdot f_{\mathcal{N}}\left(\kappa_{-s}^i - (\widehat{\kappa_0} - s d_0 + \sum_{l=1}^{s-1} l \cdot \lambda_{-(s-1-l)} \,| \theta\right) d\lambda_{-N+2,\ldots,0} \rightarrow max$$

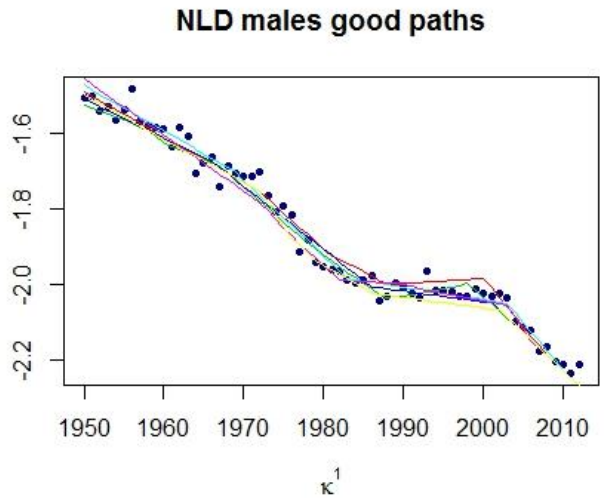
- Use Monte-Carlo integration to calculate and optimize the N-1 dimensional integral. Basic idea:
  - $\mathrm{I} = \int f(x) g(x) dx$
  - Simulate $x^1, \ldots, x^m$ with $x^i \sim g$
  - $\hat{I} = \frac{1}{m} \sum_{i=1}^m f(x^i)$
  - Here: Simulate $x^1, \ldots, x^m$ trends according to $x^l = (\lambda_{-N+2}, \ldots, \lambda_0)^l$ with $\lambda_j \sim g$
  - Calculate $\hat{I} = \frac{1}{m} \sum_{l=1}^m \prod_{j=-N}^{0} f_{\mathcal{N}}(\epsilon_j^i | \theta, x^l)$ for $i = 1,2$
  - Starting in $\mathrm{t} = 0$ we simulate historic trends. The estimated parameters can be used for projections directly.

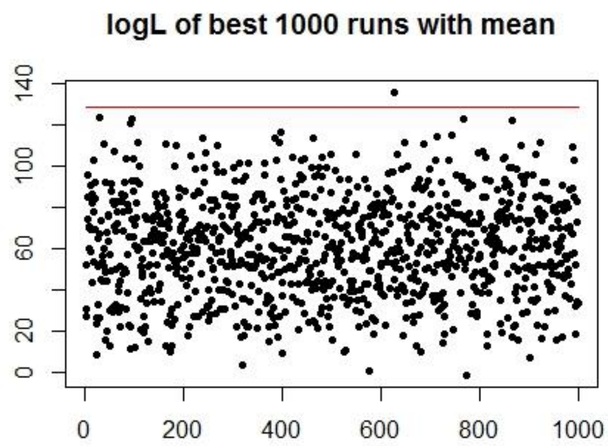
Parameter estimation

Alternatives III

Preliminary ideas

- A first example: NLD-males (constant volatility) with 1.4 Mio trials
  - $\mu = -5, \sigma^2 = 0.7, p = 0.0635, \sigma_\epsilon^2 = 0.0005, \hat{\kappa}_0 = -2.266, d_0 = -0.01977$. Starting in 2012 we simulate historic paths
- Advantages: Maximum of consistency in forecasts, flexibility on distributional assumptions
- Disadvantages and open issues:
  - No historic trends
  - Trends x^l with a high likelihood $\left(\prod_{j=-N}^{0} f_{\mathcal{N}}\left(\epsilon_j^i \middle| \theta, x^l\right)\right)$ are extremely rare
  - Huge number of simulations necessary
  - Dominated by very few simulations





Literature

- Börger, M., Fleischer, D., Kuksin, N., 2014. Modeling Mortality Trend under Modern Solvency Regimes. ASTIN Bulletin, 44: 1–38.
- Börger, M., Schupp, J., 2015. Modeling Trend Processes in Parametric Mortality. Working Paper, Ulm University.
- Cairns, A., Blake, D., Dowd, K., 2006. A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration. Journal of Risk and Insurance, 73: 687–718.
- Hunt, A. and Blake, D. (2014). Consistent mortality projections allowing for trend changes and cohort effects. Working Paper, Cass Business School
- Li, J. S.-H., Chan, W.-S., and Cheung, S.-H. (2011). Structural changes in the Lee-Carter indexes: detection and implications. North American Actuarial Journal, 15(1): 13–31.
- Sweeting, P., 2011. A Trend-Change Extension of the Cairns-Blake-Dowd Model. Annals of Actuarial Science, 5: 143–162.

Contact

Johannes Schupp(M.Sc.)
+49 (731) 20 644-241
j.schupp@ifa-ulm.de