A Bayesian Approach to Handle Small Population Effects in Stochastic Mortality Models

Dr. Liang Chen, FIA, FSAI, CERA, AFHEA
Prof. Andrew J.G. Cairns, Dr. Torsten Kleinow

Heriot-Watt University, Edinburgh

University College Cork, Ireland

Funded by Actuarial Research Centre, IFoA

Life Conference, Dublin, 2019

HERIOT WATT UNIVERSITY

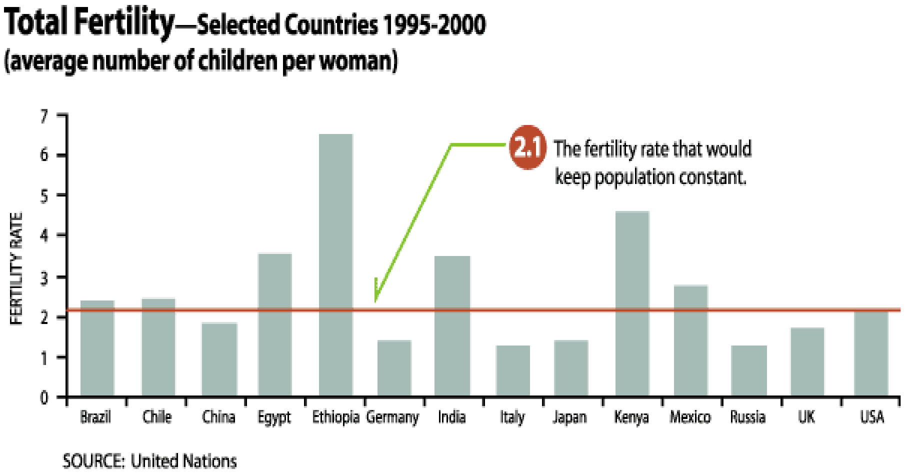
UCC
Coláiste na hOllscoile Corcaigh, Éire
University College Cork, Ireland

Actuarial Research Centre
Institute and Faculty of Actuaries

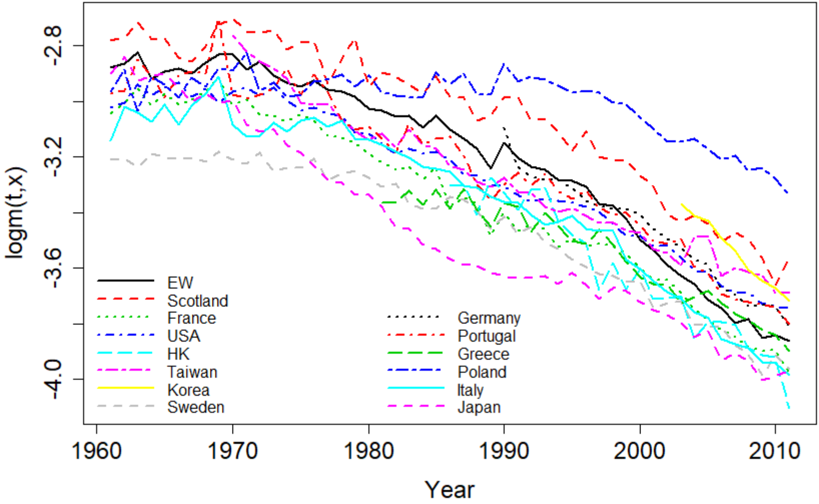
Disclaimer

Outline

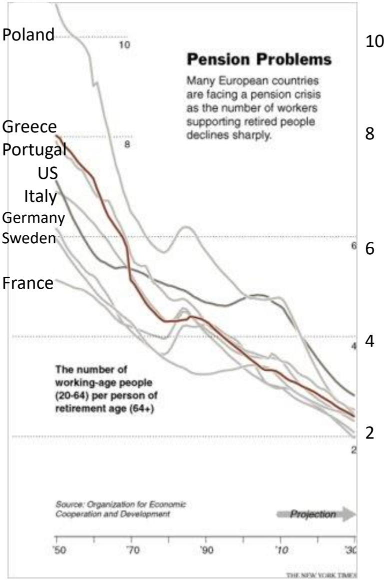
- Background and Motivations
- Impact of population size on mortality modelling based on two-stage approach
- A Bayesian approach for modelling the small population longevity risk
- Summary and conclusion
- Further research

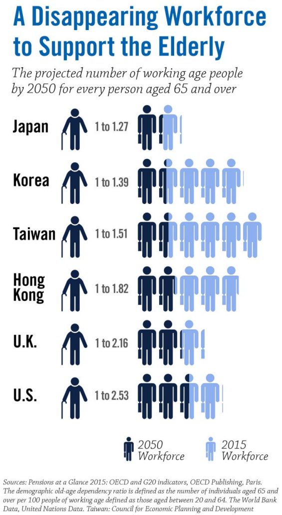

Total Fertility—Selected Countries 1995-2000
(average number of children per woman)
FERTILITY RATE
0
1
2
3
4
5
6
7
2.1
The fertility rate that would keep population constant.
Brazil
Chile
China
Egypt
Ethiopia
Germany
India
Italy
Japan
Kenya
Mexico
Russia
UK
USA
SOURCE: United Nations

Crude Mortality Improvement at age 70 (HMD)



Dependency Ratio





Measure for Mortality

- $m_c(t,x)$, the age-specific crude death rate at age x, year t, More specifically

$$m_c(t,x) = \frac{\text{Number of deaths during calendar year t, age x last birthday}}{\text{Average population during calendar year t aged x last birthday}}.$$

- $m(t,x)$, the underlying death rate, which is equal to the expected deaths divided by the exposure. More specifically

$$m(t,x) = \frac{D(t,x)}{E(t,x)}$$

- $q(t,x)$, the mortality rate, which is the probability that an individual aged exactly x at exact time t will die between t and $t+1$.
- We modelled logit $q(t,x)$ as crude $m(t,x)$ can be greater than one at advanced ages.

Why analyse small population

- Experiencing faster mortality improvement, lower interest rate, more pressure on pension funds.
- Most pension schemes are less than 1% of national population.
- Significantly more variability exhibited for mortality rates of small population
- Stochastic models might poorly fit small populations

Motivation

For small population:

- Greater sampling variation of deaths causes increased uncertainty of parameter estimates and high levels of uncertainty on projected mortality rates.
- Divergence between future realized rates and projections, future sampling variation, uncertain projection.

Stochastic Model and Data

- Let $\theta_1 = (\kappa_t^{(1)}, \kappa_t^{(2)}, \kappa_t^{(3)}, \gamma_{t-x}^{(4)})$. Stochastic Model:

$$\begin{aligned} D(t,x)|\theta_1 &\sim \text{Pois}(m(\theta_1,t,x)E(t,x)) \\ m(\theta_1,t,x) &= -\log(1-q(\theta_1,t,x)) \\ \text{logit } q(\theta_1,t,x) &= \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}) + \kappa_t^{(3)}((x-\bar{x})^2 - \hat{\sigma}_x^2) + \gamma_{t-x}^{(4)} \end{aligned}$$

- Identifiability Constraints

$$\sum_{c \in C} \gamma_c^{(4)} = 0, \qquad \sum_{c \in C} c\gamma_c^{(4)} = 0, \qquad \sum_{c \in C} c^2\gamma_c^{(4)} = 0$$

- θ_1 is estimated maximising the log-likelihood function:

$$\begin{aligned} l(\theta_1; D, E) &= \sum_{t,x} D(t,x)\log[E(t,x)m(\theta,t,x)] \\ &\quad -E(t,x)m(\theta,t,x) - \log[D(t,x)!] \end{aligned}$$

Stochastic Model and Data (cont.)

- Projecting Period Effects $\boldsymbol{\kappa} = (\kappa^{(1)}, \kappa^{(2)}, \kappa^{(3)})$ with a three dimensional normal random walk model

$$\Delta \boldsymbol{\kappa}_t = \boldsymbol{\mu} + L\epsilon_t,$$

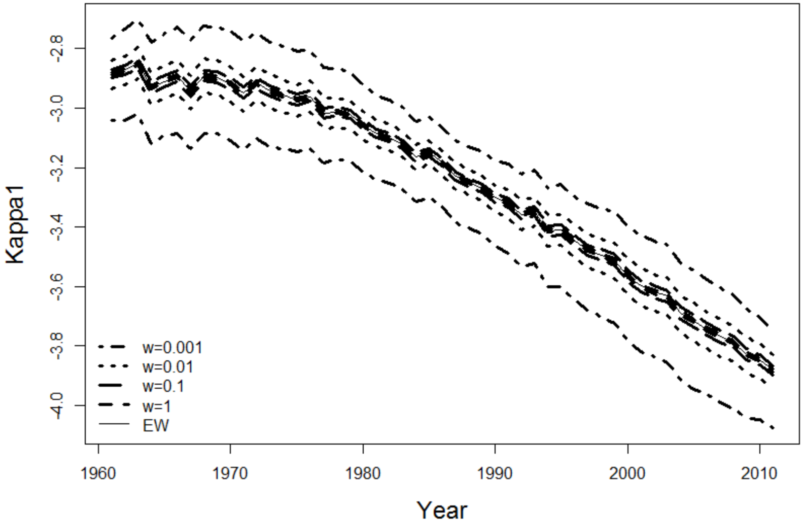
  where covariance matrix $\boldsymbol{V}_{\boldsymbol{\epsilon}} = LL'$ and the drift $\boldsymbol{\mu}$ are estimated given estimated θ_1.

- Data: Benchmark exposure $E_0(t,x)$ and corresponding deaths count $D_0(t,x)$ of the males in England and Wales (EW) in the HMD database, during year 1961 to 2011, aged 50-89 last birthday.

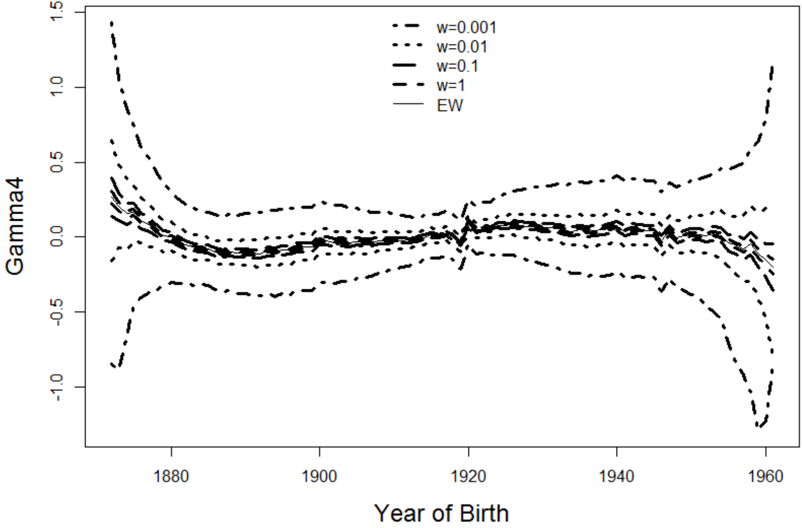
- Simulation Method
  - Estimate θ_1 for benchmark population, denoted as $\hat{\theta}_{1,0}$
  - Construct small population $E_w(t,x) = wE_0(t,x)$ for $w = 1, 0.1, 0.01, 0.001$, i.e. construct 4 scaled populations.
  - (Re-) Simulate $D_w(t,x)|\hat{\theta}_{1,0} \sim \text{Pois}(m(\hat{\theta}_{1,0}, t, x)wE_0(t,x))$
  - Estimate θ_1 for $D_w(t,x)$, denoted as $\hat{\theta}_1^w$.

Actuarial Research Centre
Institute and Faculty of Actuaries

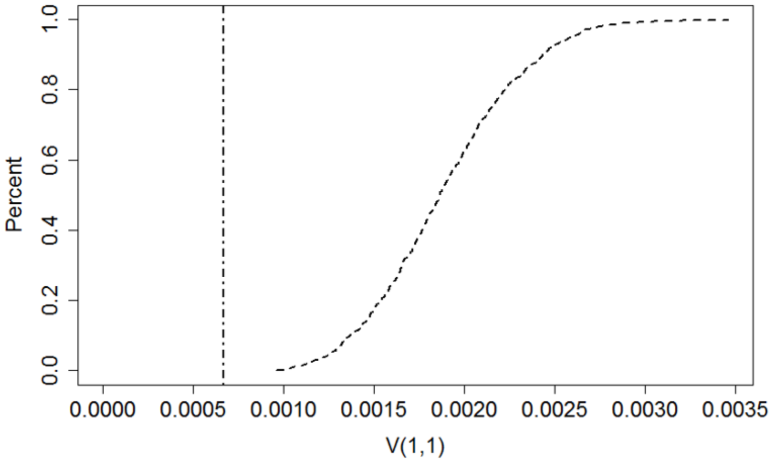
Distribution (90% CI) of the Finite Sample MLEs



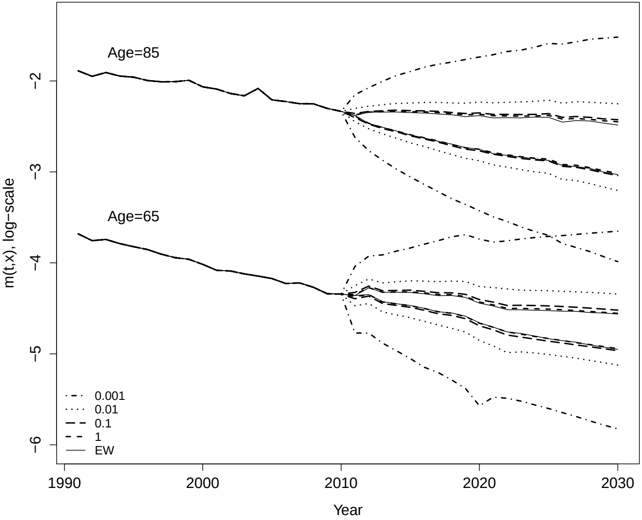
Distribution (90% CI) of the Finite Sample MLEs



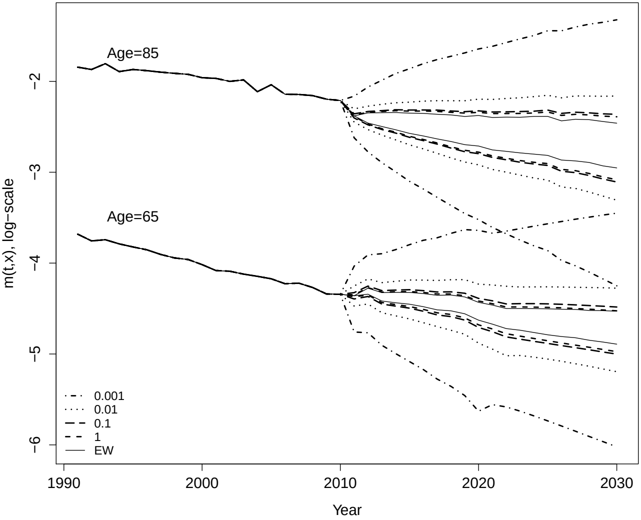
Estimated Volatility of Random Walk verse Benchmark



Projected Mortality Rate



Projected Mortality Rate with Parameter Uncertainty



Summary of the Findings

- There exists a bias in the estimated covariance matrix of the random walk fitted to the period effects when the size of the underlying population is small. As a consequence, prediction intervals are rather wide for small populations even when parameter uncertainty is ignored.
- Parameter uncertainty becomes much less important when only relatively short forecast horizons are considered. This finding aligns with Cairns et al. (2006).
- The inclusion of parameter uncertainty for the drift parameter $\boldsymbol{\mu}$ adds further uncertainty about the projected mortality rates.
- This is in line with results obtained by Kleinow and Richards (2016) who have found that the uncertainty about the drift of the period effect in a Lee-Carter model has little impact on the uncertainty of short term projections while it significantly affects the uncertainty of long-term projections.

Two-Stage and Bayesian Approaches

- Two-stage approach that fits and projects mortality rates separately, leads to biased estimates of volatility for small populations
  - Large sampling variation affects latent parameter estimation, with significant noise obscuring the true signal (Cairns et al. 2011)
  - Result in non-negligible bias to the parameter estimation of the projecting model (Chen, Cairns and Kleinow 2016)
  - Over-fit the short cohorts (Cairns et al. 2009)
- Bayesian approach offers a way to avoid or reduce this bias by:
  - Combining the Poisson likelihood with the projecting time series models
  - The estimated latent parameters are restricted to be more like proposed time series models when projecting models dominate while modelling small populations.
  - Using more informative prior distribution with the knowledge of the larger benchmark population.
  - Better estimation for short cohorts.

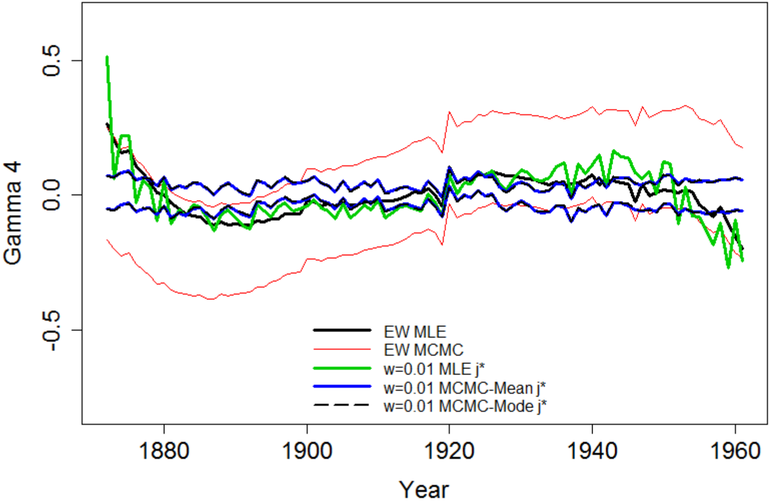
Actuarial Research Centre
Institute and Faculty of Actuaries

Prior Distributions

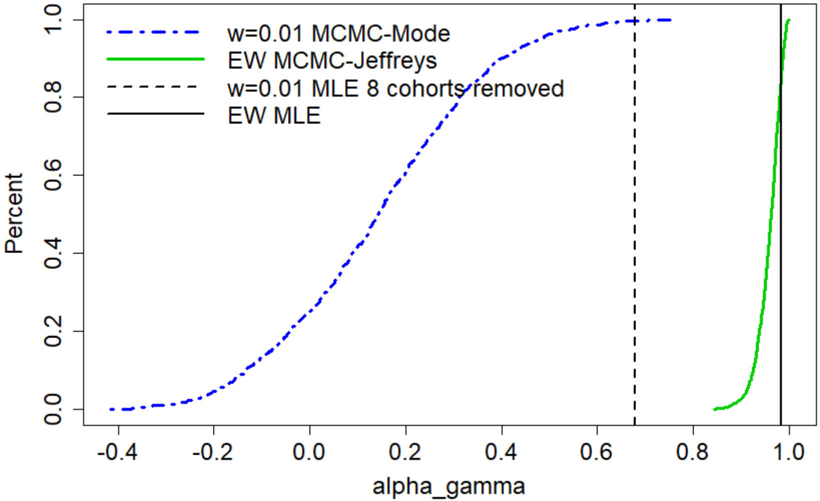
- $(\kappa_{t_1}^{(1)}, \kappa_{t_1}^{(2)}, \kappa_{t_1}^{(3)}) \propto 1$,
- $\boldsymbol{\kappa}_t = \boldsymbol{\kappa}_{t-1} + \boldsymbol{\mu} + \boldsymbol{\epsilon}_t$ for $t \leq t_2$,
  - $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu) \propto 1$,
  - $\boldsymbol{\epsilon}_t \sim MVN(0, \boldsymbol{V}_{\boldsymbol{\epsilon}})$, i.i.d three dimensional multi-variate normal distribution independent of t,
- $\boldsymbol{V}_{\boldsymbol{\epsilon}} \sim InverseWishart(\nu, \boldsymbol{\Sigma})$
  - MCMC-Mean: Fix the mean of the prior to $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}$
  - MCMC-Mode: Fix the mode of the prior to $\hat{\boldsymbol{V}}_{\boldsymbol{\epsilon}}^{\text{EW}}$
- $\gamma_c^{(4)} = \alpha_\gamma \gamma_{c-1}^{(4)} + \epsilon_c$ for $c > t_1 - x_{n_a}$,
  - i.i.d $\epsilon_c \sim N(0, \sigma_\gamma^2)$,
  - $\alpha_\gamma \propto (1 - \alpha_\gamma^2)^g$ for $|\alpha_\gamma| < 1$,
  - $\sigma_\gamma^2 \sim$ Inverse Gamma (a_γ, b_γ)
- $\gamma_{c_1}^{(4)} \sim N(0, \frac{\sigma_\gamma^2}{1-\alpha_\gamma^2})$

Actuarial Research Centre
Institute and Faculty of Actuaries

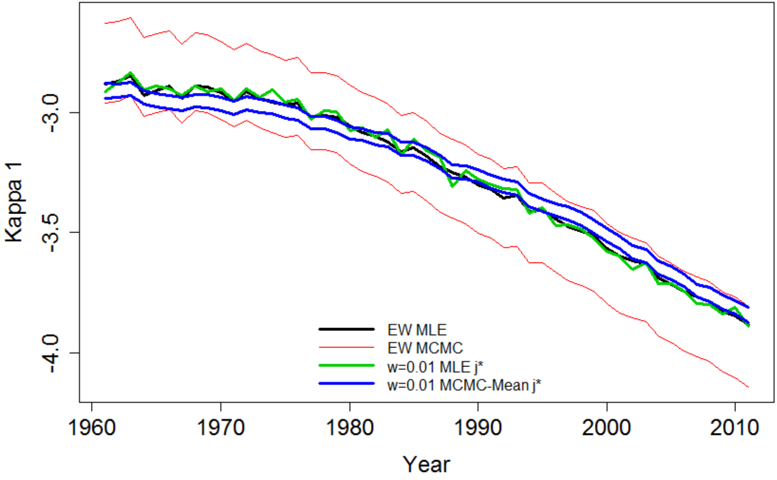
Results: γ MCMC



Results: α_γ MCMC

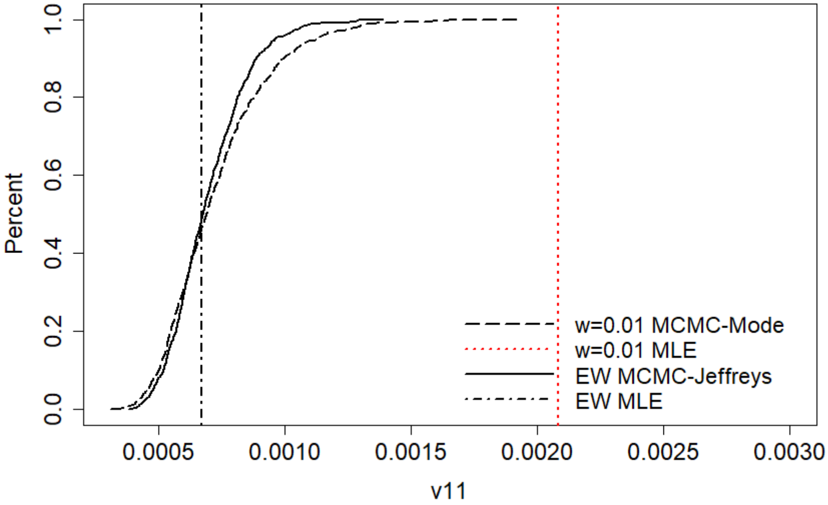


Results $\boldsymbol{\kappa}^{(1)}$

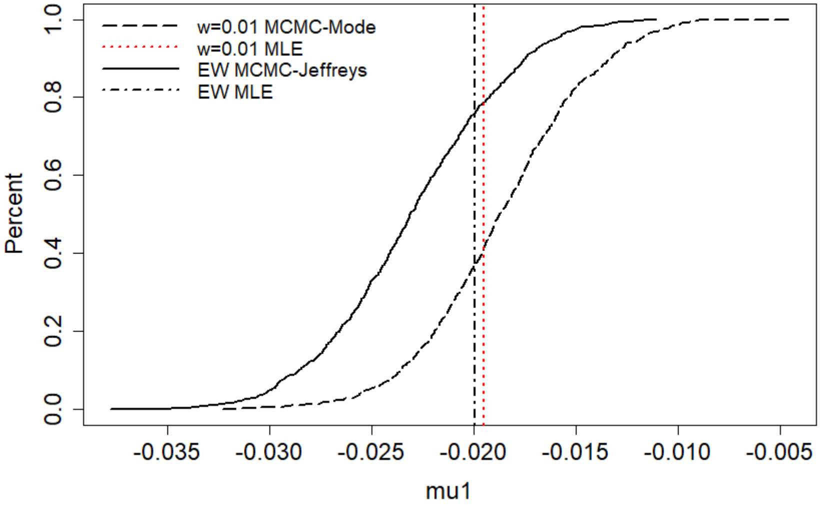


Actuarial Research Centre
Institute and Faculty of Actuaries

Results: $\boldsymbol{V}_\epsilon(1,1)$ MCMC



Results: $\mu(1)$ MCMC



Longevity Risk of A 25-year Temporary Annuity

A temporary annuity of £1 per annum payable annually in arrears to a life aged 65 exactly, starting at the beginning of year 2012 with term of 25 years. We define the Longevity Risk (LR) at 99.5% level as:

$$LR_{99.5\%} = \left(\frac{a_{99.5\%}}{a_{50\%}} - 1 \right) \times 100,$$

where $a_{.\%}$ is the percentile of the distribution of the annuity price.

	$i =$	4%	$i =$	2%	$i =$	0%
	Mean	LR (%)	Mean	LR (%)	Mean	LR (%)
EW-MCMC	12.2631	5.27	14.8394	6.28	18.3365	7.47
w-MCMC	12.1220	5.72	14.6420	6.76	18.0556	7.95
EW-MLE	12.2166	4.24	14.7720	5.04	18.2371	5.98
w-MLE	12.2052	5.12	14.7441	6.08	18.1805	7.09

Actuarial Research Centre
Institute and Faculty of Actuaries

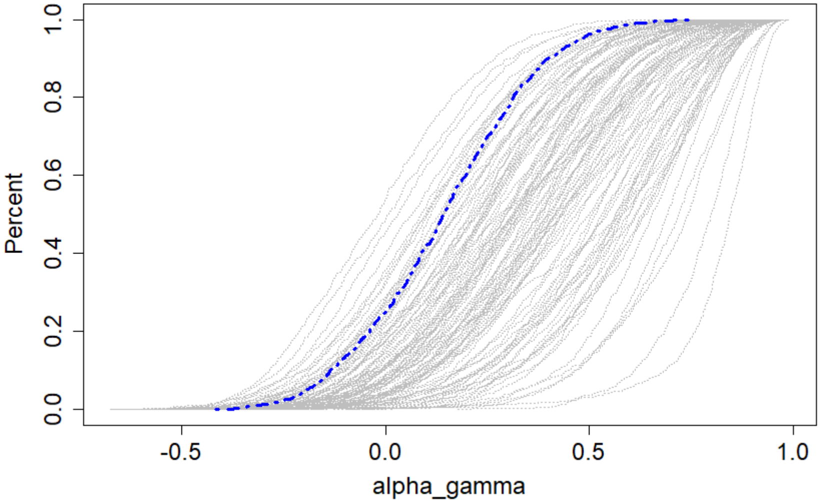
Longevity Risk of A 25-year Temporary Annuity Deferred by 10 Years

A temporary annuity of £1 per annum payable annually in arrears to a life aged 55 exactly, deferred for 10 years, starting at the beginning of year 2012 with term of 25 years.

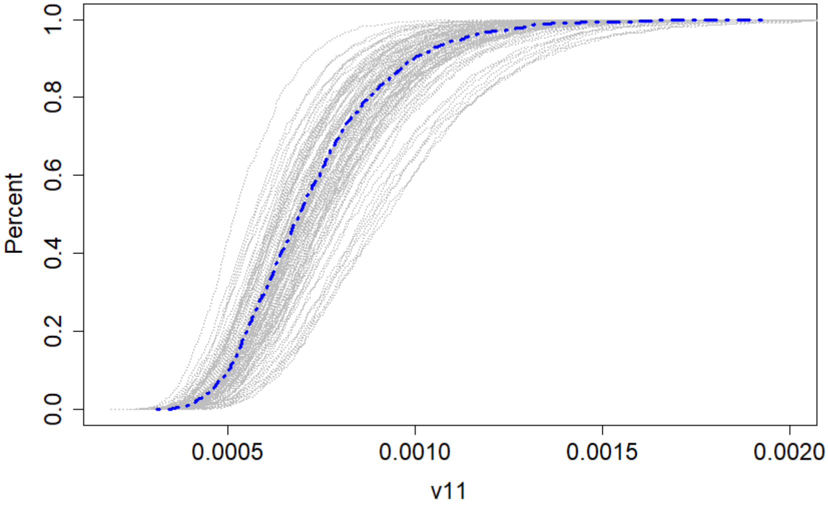
	i =4%		i =2%		$i = 0\%$	
	Mean	LR (%)	Mean	LR (%)	Mean	LR (%)
EW-MCMC	8.2744	7.02	12.2519	8.14	18.6117	9.45
w-MCMC	7.9292	8.23	11.6832	9.45	17.6509	10.87
EW-MLE	8.1928	5.50	12.1150	6.35	18.3759	7.32
w-MLE	8.2539	7.57	12.2039	8.80	18.5059	10.17

Actuarial Research Centre
Institute and Faculty of Actuaries

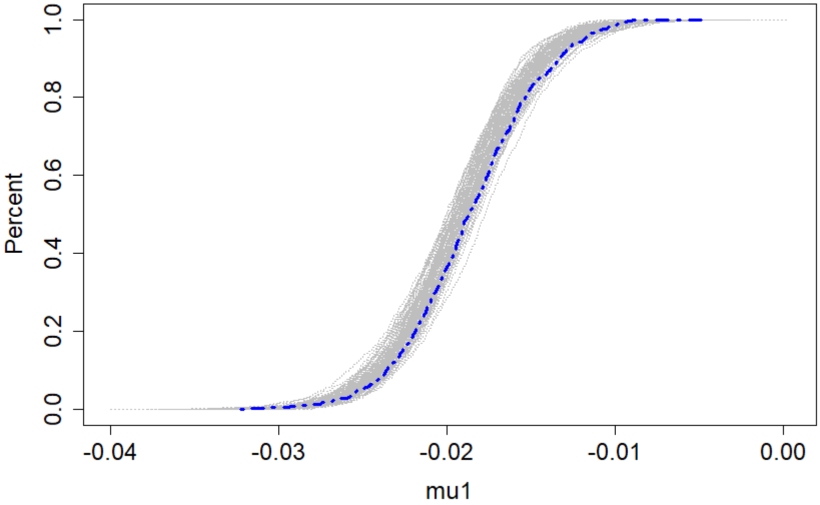
Sensitivity Test: α_γ MCMC



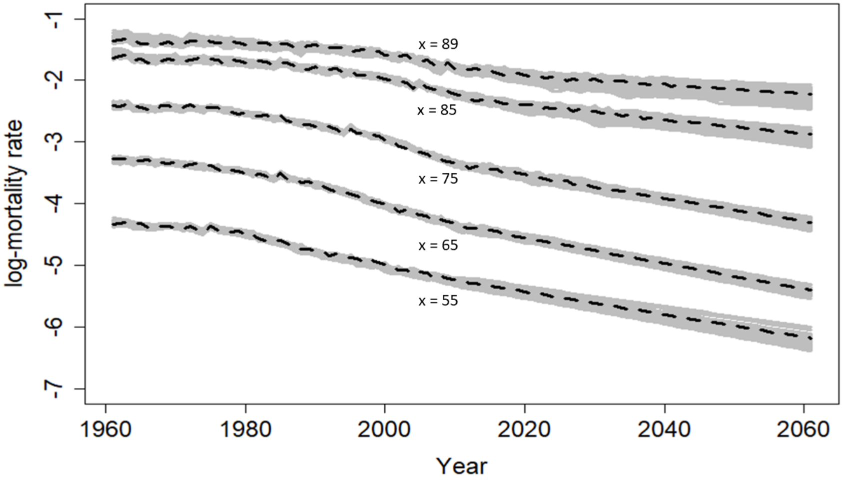
Sensitivity Test: $\boldsymbol{V}_{\epsilon}(1,1)$ MCMC



Sensitivity Test: μ_1 MCMC



Sensitivity Test: $m(t, x)$ MCMC



Summary

- We have demonstrated to the users of the stochastic mortality models how the information of a larger population could be embedded for parameter estimation and forecasts by a Bayesian model.
- Studied to what extent the parameter estimation could be improved compared with the two-stage approach and the financial implication in the context of annuity pricing.
- The users should be informed how the importance of the prior information dominates the parameter estimation of a much smaller population and in what way the sampling variation affects the parameter estimation and mortality forecasts.
- We find that our Bayesian model and methodology of using the information of large referencing population provide an improved estimation for the volatility of small population.
- The (central) projections of small populations are not "significantly" different from the "true" projections (of the larger reference population).

What is next?

We will pursue our research on developing a methodology with the advantages of both the two-stage and the Bayesian methods for modelling the mortality experience of small populations. We aim to develop a likelihood based Bayesian approach that can be easily implemented by both the industry practitioners and faculty researchers.

Appendix: Markov Chain Monte Carlo (MCMC)

The joint posterior density for $\boldsymbol{\theta} = (\boldsymbol{\kappa}, \boldsymbol{\gamma}, \boldsymbol{\mu}, \boldsymbol{V}_{\boldsymbol{\epsilon}}, \alpha_\gamma, \sigma_\gamma)$ conditional on $\boldsymbol{D}, \boldsymbol{E}$ can be expressed as:

$$p(\boldsymbol{\theta}|\boldsymbol{D}, \boldsymbol{E}) \propto p(\boldsymbol{D}, \boldsymbol{E}|\boldsymbol{\kappa}, \boldsymbol{\gamma}) p(\boldsymbol{\kappa}, \boldsymbol{\gamma}|\boldsymbol{\mu}, \boldsymbol{V}_{\boldsymbol{\epsilon}}, \alpha_\gamma, \sigma_\gamma) p(\boldsymbol{\mu}, \boldsymbol{V}_{\boldsymbol{\epsilon}}, \alpha_\gamma, \sigma_\gamma)$$

257-dimensional parameter vector. Drawing samples directly from the above? Hmm...

The aim of MCMC, based on Gelman et al. (2014), is to simulate a random walk path in the space of $\boldsymbol{\theta}$ that eventually converges to our target distribution $p(\boldsymbol{\theta}|\boldsymbol{D}, \boldsymbol{E})$.

Gibbs sampler and Metropolis-Hastings algorithm have been widely used in previous studies, e.g. Cairns et al. (2011), Czado et al. (2005), Pedroza (2006).

Hamiltonian Monte Carlo (HMC)

See Neal (1993) and De Almeida (1990) for the principle of Hamiltonian dynamic system. Updating Algorithm is as follows:

- Given the current value for $\boldsymbol{\theta}$ and $\boldsymbol{p}$ at iteration τ, denoted as $(\hat{\boldsymbol{\theta}}(\tau), \hat{\boldsymbol{p}}(\tau))$, sample a new momentum variable through the canonical function $\tilde{\boldsymbol{p}}(\tau) \sim f_{\boldsymbol{P}}(\hat{\boldsymbol{p}}(\tau))$.
- From state τ to $\tau + 1$, perform leapfrog discretization by L times with step size δ. Denote as $(\boldsymbol{\theta}^*, \boldsymbol{p}^*)$ the ending value.
- Calculate the Metropolis acceptance probability:

$$\alpha = \min\Big\{1, \exp\Big(- U_{\boldsymbol{\theta}}(\boldsymbol{\theta}^*) + U_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) - K_{\boldsymbol{P}}(\boldsymbol{p}^*) + K_{\boldsymbol{P}}(\hat{\boldsymbol{p}})\Big)\Big\} \quad (1)$$

- Draw a random number $u \sim U(0, 1)$ where $U(\cdot)$ represents a uniform distribution. If $u \leq \alpha$, accept the new state value $(\boldsymbol{\theta}^*, \boldsymbol{p}^*)$ as the current value for iteration $\tau + 1$, else $(\hat{\boldsymbol{\theta}}(\tau), \hat{\boldsymbol{p}}(\tau))$ is kept for state $\tau + 1$.

Thank You!

Questions?

HERIOT WATT UNIVERSITY

UCC
Coláiste na hOllscoile Corcaigh, Éire
University College Cork, Ireland

Actuarial Research Centre
Institute and Faculty of Actuaries