A Regression Based Approach for Valuing Longevity Measures

Pietro Millossovich
Bayes Business School & University of Trieste

joint work with Fabio Viviano & Anna Rita Bacinello

Longevity 18 Conference - London 2023

Objective

- Evaluate future longevity metrics using (extrapolative) forecasting mortality models $\rightsquigarrow$ longevity risk
  - cohort approach vs period approach $\rightsquigarrow$ conditional expectations
  - multiple populations $\rightsquigarrow$ biometric constraints
  - simulation $\rightsquigarrow$ empirical distribution
  - main tool: Least-Square Monte Carlo (LSMC) method

Cohort Based Valuations

- Importance of cohort based valuations
  - Cairns et al. 2011a, 2011b (annuities and longevity metrics) $\rightsquigarrow$ Taylor expansion (re-simulation)
  - Boyer and Stentoft, 2013, 2017 (survival probs and longevity derivatives)
  - Feng et al. 2022 $\rightsquigarrow$ green nested simulations
- LSMC:
  - American options: Tilley 1993, Carriere 1996, Tsitsiklis & Van Roy 2001, Longstaff & Schwarz 2001, ...
  - surrender option: Andreatta & Corradin 2003, Bacinello et al. 2008, 2009, 2011
  - solvency requirements: Floryszczak, et al 2016, Bauer and Ha 2018

Extrapolative Stochastic Mortality Models

- Multiple populations: $m_{x;t}^{(p)}$ = central death rate at age x in year t for population p

$$m_{x,t}^{(p)} = f(t, x, p, X_t)$$

  - populations: M/F, set of countries, smokers/non smokers, national population/pension scheme, ...
  - X_t = (vector of Markov) state variables
  - most extrapolative multi-population models included here: CF, ACF, Common Age Effect, joint/relative, ... (Villegas et al. 2017, Enchev et al. 2017, Li et al. 2015)
  - latent or explanatory variables (Boonen & Li, 2017)

Extrapolative Stochastic Mortality Models

- Example: Augmented Common Factor (ACF) model, (Li & Lee 2005), $p = M, F$
  - death counts $D_{x,t}^{(p)} \sim \text{Poisson}\big(E_{x,t}^{(p)} m_{x,t}^{(p)}\big)$,

$$\log m_{x,t}^{(p)} = \alpha_x^{(p)} + B_x K_t + \beta_x^{(p)} \kappa_t^{(p)}, \qquad p = M,\ F$$

  - K_t common factor, $\kappa_t^{(p)}$, $p = M, F$ specific factors
  - $X_t = (K_t, \kappa_t^{(M)}, \kappa_t^{(F)})$

Cohort vs Period

- Future calculations with a stochastic mortality model: period or cohort?
- $0 =$ today, $T > 0$ future time; x age of the individual at time T
- Φ: a function(al) defining the metric to be calculated
  - period
$$\underbrace{\Phi(m_{x,T}, m_{x+1,T}, m_{x+2,T}, \ldots)}_{\text{known at } T}$$
  - cohort
$$E_T[\underbrace{\Phi(m_{x,T}, m_{x+1,T+1}, m_{x+2,T+2}, \ldots)}_{\text{not known at } T}]$$

Cohort vs Period

- Formally
  - $\mathcal{F}_t$ information available at time t $\rightsquigarrow$ includes information on mortality rates up to time t
    $$\sigma(X_u,\ u \leq t) \subset \mathcal{F}_t$$
  - $\tau_x(t)$ residual lifetime of a (representative) individual aged x at time t $\rightsquigarrow$ stopping time in $(\mathcal{F}_{t+u})_{u\geq 0}$
  - biometric variables:
    $$\mathcal{M} = \sigma(X_t,\ t \geq 0)$$
  - (implicit) assumption:
    $$P(\tau_x(t) \geq l | \mathcal{F}_t \vee \mathcal{M}) = \exp\left\{-\sum_{k=0}^{l-1} m_{x+k,T+k}\right\}$$

Cohort vs Period

- Example: l-years survival prob. for an individual aged x at time T
  - period:
$$ {}_l p_{x,T} = \exp\left\{ -\sum_{k=0}^{l-1} m_{x+k,T} \right\} $$
  - cohort:
$$ {}_l p_x(T) = P_T(\tau_x(T) \geq l) = \mathbb{E}_T\left[\exp\left\{ -\sum_{k=0}^{l-1} m_{x+k,T+k} \right\} \right] $$

Cohort vs Period

- Period vs cohort
  - period only considers mortality improvements up to time T, neglects further improvements after T
  - cohort: more sound approach
- Closed form expressions
  - available under some special cases (eg affine processes, Biffis 2005)
  - not available under most common stochastic mortality models (LC, CBD with Gaussian time indices: sum of lognormals)
- Evaluate the conditional expectations under the cohort approach via LSMC

Pros/Cons of the LSMC

- Pros
  - universality wrt model choice
  - flexibility wrt contract structure
  - consistent assessment wrt multiple values of x and T
- Cons
  - needs to store all simulations
  - number of simulations vs number of basis functions? Moreno and Navas 2003, Stentoft 2004

Longevity Measures

- Focus on
  - life expectancy $\rightsquigarrow$ location
  - lifespan disparity $\rightsquigarrow$ dispersion
  - can be extended to other metrics
- Serious misspecification due to
  - cohort vs period approach $\rightsquigarrow$ effect of rolling improvements
  - single vs multi-population $\rightsquigarrow$ interaction between groups, coherent assessment

Life Expectancy

- For an individual aged x at time $T > 0$
  - period:

$$e^p_{x,T} = \frac{1}{2} + \sum_{l=1}^{\infty} {}_l p_{x,T}$$

  - cohort:

$$\begin{aligned} e^c_x(T) =& \frac{1}{2} + \sum_{l=1}^{\infty} {}_l p_x(T) \\ =& \frac{1}{2} + \mathbb{E}_T\left[\sum_{l=1}^{\infty} \exp\{-\left(m_{x,T} + \ldots + m_{x+l-1,T+l-1}\right)\}\right] \end{aligned}$$

Lifespan Disparity

- *Life expectancy lost due to death by an individual aged x at time t* (Vaupel, 1986), aka *numbers of years of life lost*

- For an individual aged x at time $T > 0$

  - period:

$$e^{\dagger,p}_{x,T} = \sum_{k=0}^{\infty} e^{p}_{x+k,T} \cdot {}_{k}p_{x,T} \cdot (1 - e^{-m_{x+k,T}})$$

Lifespan Disparity

- *Life expectancy lost due to death by an individual aged x at time t* (Vaupel, 1986), aka *numbers of years of life lost*

- For an individual aged x at time $T > 0$
  - cohort:

$$
\begin{aligned}
e_x^{\dagger,\,c}(T) =& \mathbb{E}_T\left[e^c_{x+\tau_x(T)}(T+\tau_x(T))\right] \\
=& \mathbb{E}_T\left[\sum_{k=0}^{\infty}(1-e^{-m_{x+k,T+k}})\sum_{h=1}^{\infty} e^{-\sum_{l=0}^{k+h-1} m_{x+l,T+l}}\right] \\
=& \mathbb{E}_T\left[\sum_{k=0}^{\infty} e^c_{x+k}(T+k)\cdot e^{-\sum_{l=0}^{k-1} m_{x+l,T+l}}\cdot\left(1-e^{-m_{x+k,T+k}}\right)\right]
\end{aligned}
$$

$\rightsquigarrow$ "double" LSMC!

Longevity Metrics

- More generally, for any functional Φ

$$Z_{T,x} = \mathbb{E}_T[\Phi(m_{x,T}, m_{x+1,T+1}, m_{x+2,T+2}, \ldots)]$$

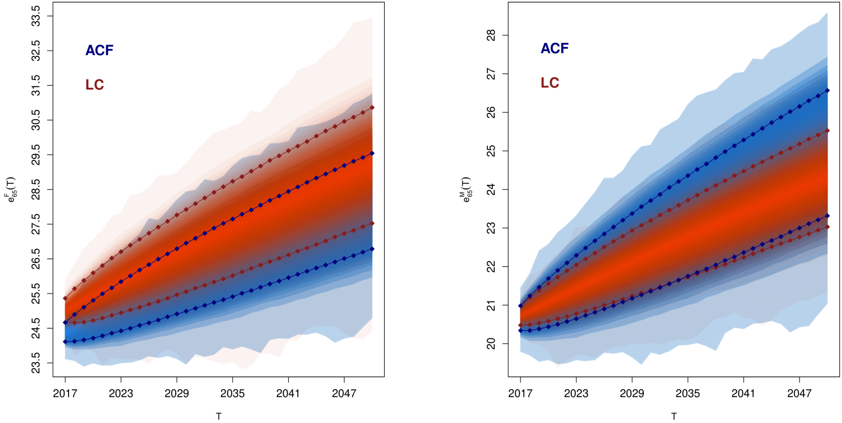
then a double LSMC can be used

$$\mathbb{E}_T[Z_{T+\tau_x(T),x+\tau_x(T)}] =$$
$$= \mathbb{E}_T\left[\sum_{k=0}^{\infty} Z_{T+k,x+k} \cdot e^{-\sum_{l=0}^{k-1} m_{x+l,T+l}} \cdot \left(1 - e^{-m_{x+k,T+k}}\right)\right]$$

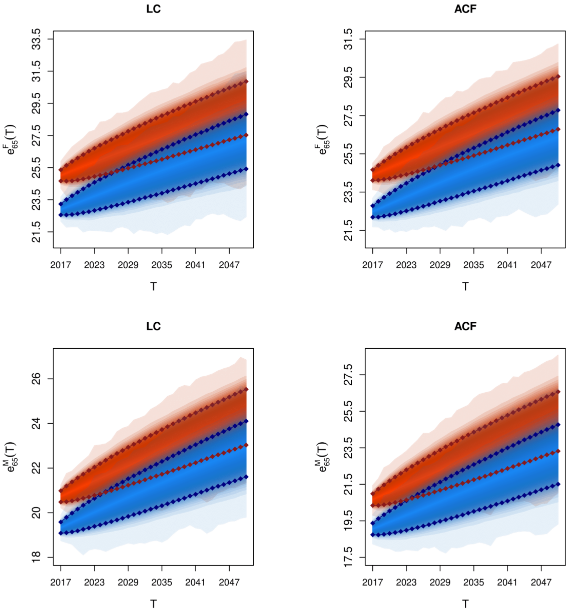
Data

- Italian F & M population deaths from HMD
  - period: 1965-2016
  - age: 35-89
  - Fit:
    - ACF (K_t: RWD, $\kappa_t^{(M)}, \kappa_t^{(F)}$: AR(1))
    - LC model $\rightsquigarrow$ independent modelling
  - log-linear closure up to the ultimate age 120
  - 20000 simulations
  - basis functions: raw polynomials of degree 3
  - simulate future life expectancy and lifespan disparity for M and F aged $x = 65$ at future horizons $T \in \{2017, \ldots, 2050\}$.

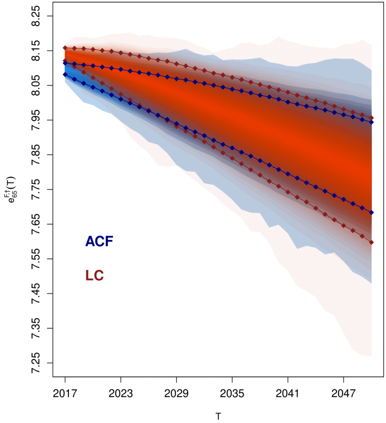
Life Expectancy - LC vs ACF

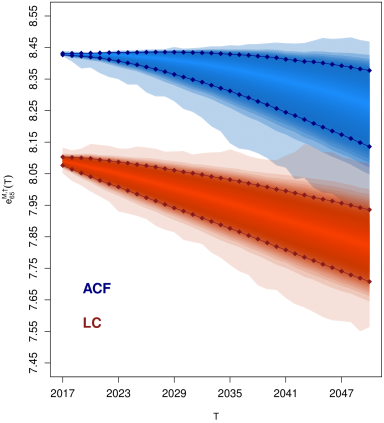


Life Expectancy - Period (blue) vs Cohort (red)

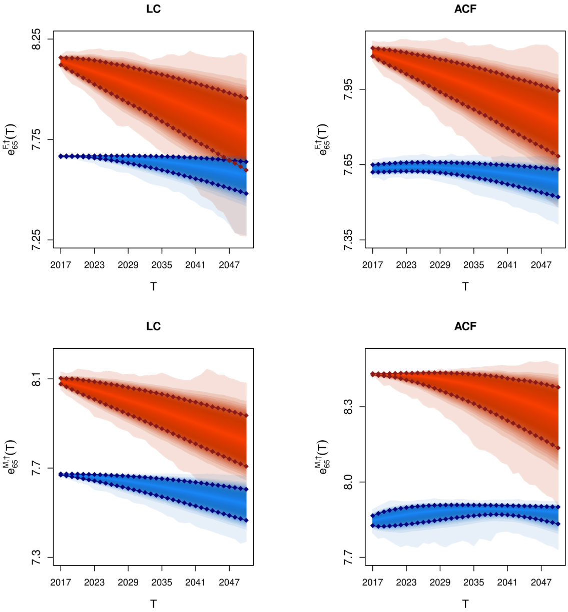


Lifespan Disparity - LC vs ACF





Lifespan Disparity - Period (blue) vs Cohort (red)



Conclusion

- Simulation + regression approach to calculate cohort based future annuity values and other longevity metrics
  - flexibility
  - wide range of applications
  - joint longevity metrics

Thank you!