Royal Statistical Society, Edinburgh

Applying survival models to life-office mortality

Stephen Richards
21$^{\text{st}}$ April 2009

Longevitas

- Used by insurers, reinsurers, investment banks and consulting actuaries
- Mainly UK clients, but some in France and Germany

Overview

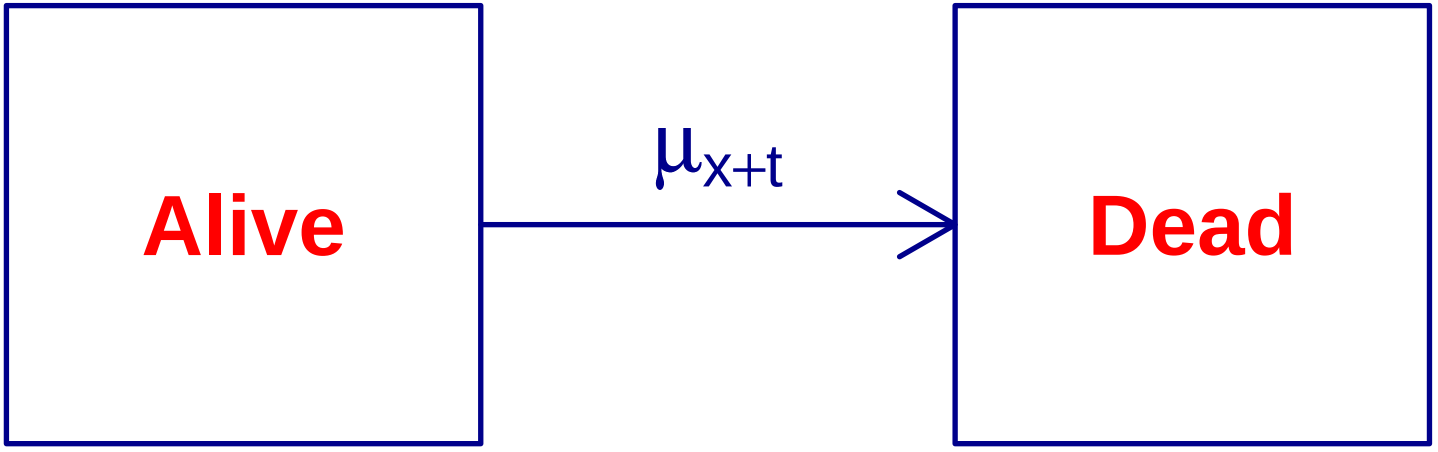
1. Model structure
2. Data
3. Risk factors
4. Time- and phase-varying factors
5. Conclusions and questions

1. Model structure

Background

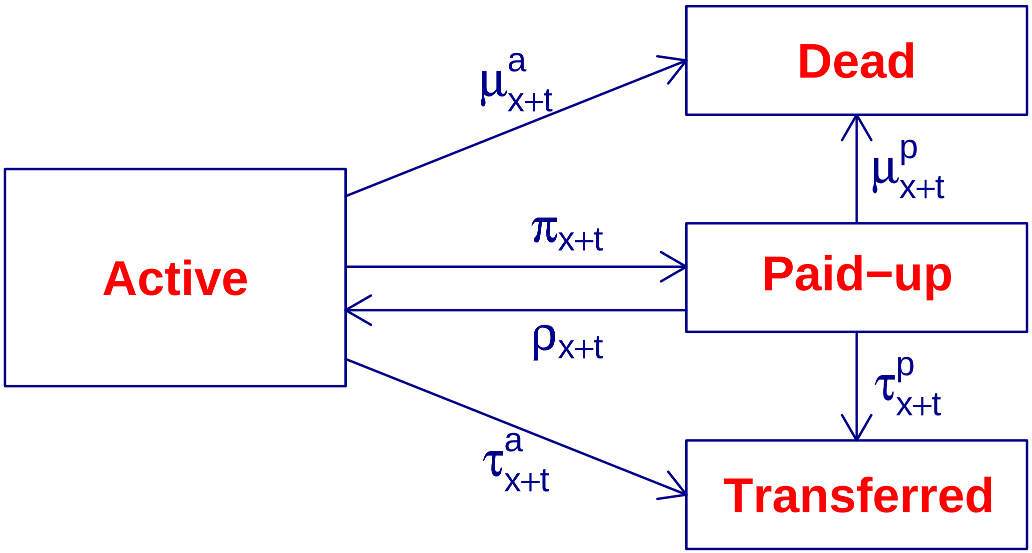
- Insurers and pension schemes have large liabilites
- Wide and varied demographic risks:
  — mortality
  — longevity
  — critical illness
  — lapse
  — credit risk and banking applications
- Prefer a single, unified approach to all these risks

Mortality model for annuities



Source: Longevitas Ltd

Persistency model for personal pensions



Source: Longevitas Ltd

Understanding actuaries — translation table

"central exposed-to-risk" → waiting time

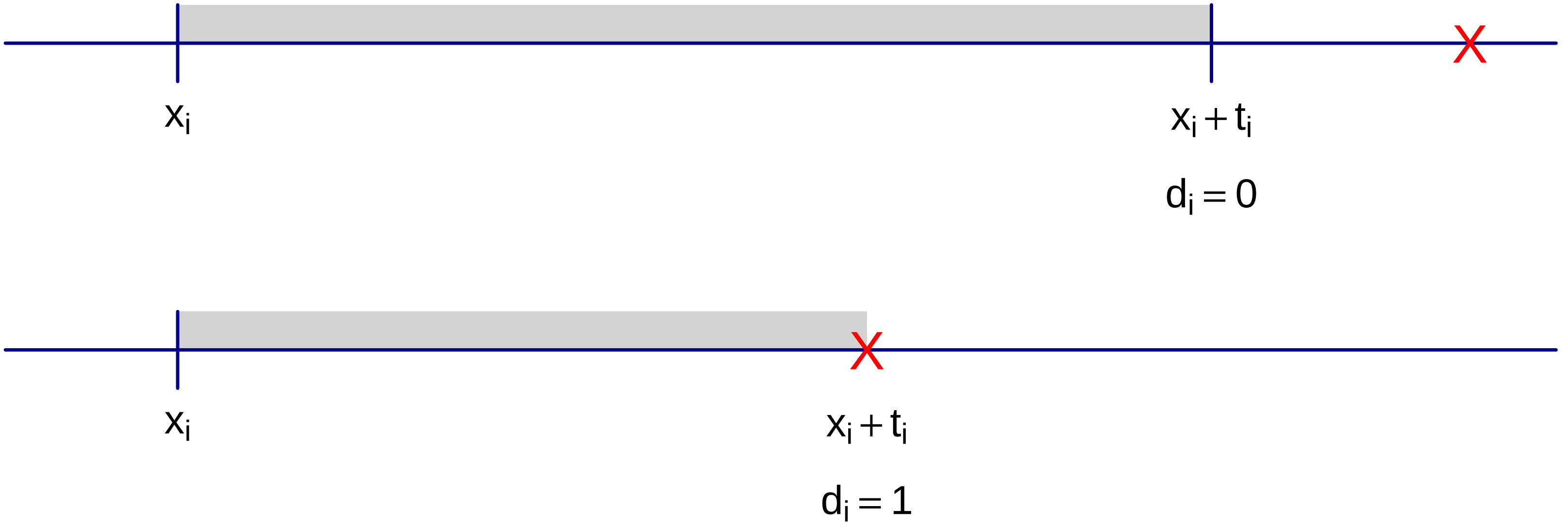
"force of mortality" → mortality hazard, usually denoted μ_x

"mortality law" → functional form for mortality hazard

$q_x \rightarrow \Pr\left(\text{death before age } x+1 \mid \text{alive aged } x\right)$

${}_tp_x \rightarrow \Pr\left(\text{survives to age } x+t \mid \text{alive aged } x\right)$, i.e. survivor function

Survival models

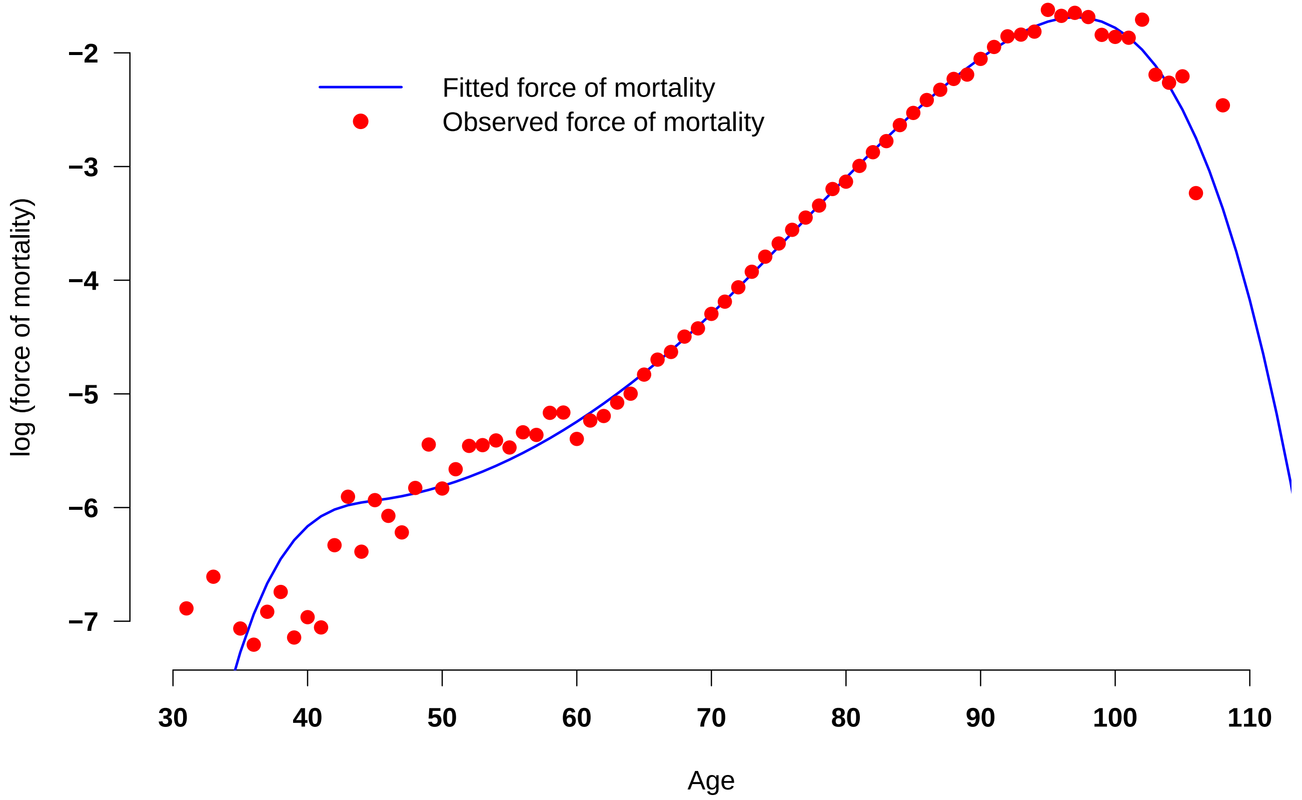


Time observed, t_i, is shown in grey, while deaths are marked $\times$.

Survival models

- Time observed, t_i, is *waiting time* (*central exposed-to-risk* to actuaries)
- d_i is the event indicator
- t_i and d_i not independent, so considered as a pair $\{t_i, d_i\}$
- Not all lives are dead, so survival times are *right-censored*
- Lives enter at age $x_i > 0$, so data is *left-truncated*

Shape of mortality hazard



Source: Richards (2008)

Traditional survival models

$$\text{Cox } \mu_x = e^{\alpha}$$

$$\text{Weibull } \mu_x = e^{\alpha} x^{\sigma-1}$$

$$\text{Log} - \text{Logistic } \mu_x = \frac{e^{\alpha+\sigma} x^{e^{\sigma}-1}}{1 + e^{\alpha} x^{e^{\sigma}}}$$

$$\text{Lognormal } \mu_x = \frac{\frac{1}{xe^{\sigma}\sqrt{2\pi}} \exp\left(-\frac{(\log x - \alpha)^2}{2e^{2\sigma}}\right)}{1 - \Phi\left(\frac{\log x - \alpha}{e^{\sigma}}\right)}$$

Actuarial mortality laws

$$\text{Gompertz } \mu_x = e^{\alpha+\beta x}$$

$$\text{Perks } \mu_x = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}$$

$$\text{Beard } \mu_x = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\rho+\beta x}}$$

$$\text{Makeham } \mu_x = e^{\epsilon} + e^{\alpha+\beta x}$$

$$\text{Makeham} - \text{Perks } \mu_x = \frac{e^{\epsilon} + e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}$$

$$\text{Makeham} - \text{Beard } \mu_x = \frac{e^{\epsilon} + e^{\alpha+\beta x}}{1+e^{\alpha+\rho+\beta x}}$$

Problems with left-truncation

- Standard survival models don't cope well with left-truncation
- Revert to first principle with likelihood function, L:

$$L \propto \Pr\left(\text{survival to age } x+t \middle| \text{alive aged } x\right) = \exp\left(-H_x(t)\right)\mu_{x+t}^d$$

where $H_x(t)$ is the *integrated hazard function* and d is a binary indicator variable for the event of interest.

- Easier to work with log-likelihood, ℓ:

$$\ell = -H_x(t) + d\log\mu_{x+t}$$

Simple analytics

- Work out $H_x(t)$
- Parameter estimates from maximising log-likelihood, ℓ
- Approximate standard errors from inverting information matrix

2. Data

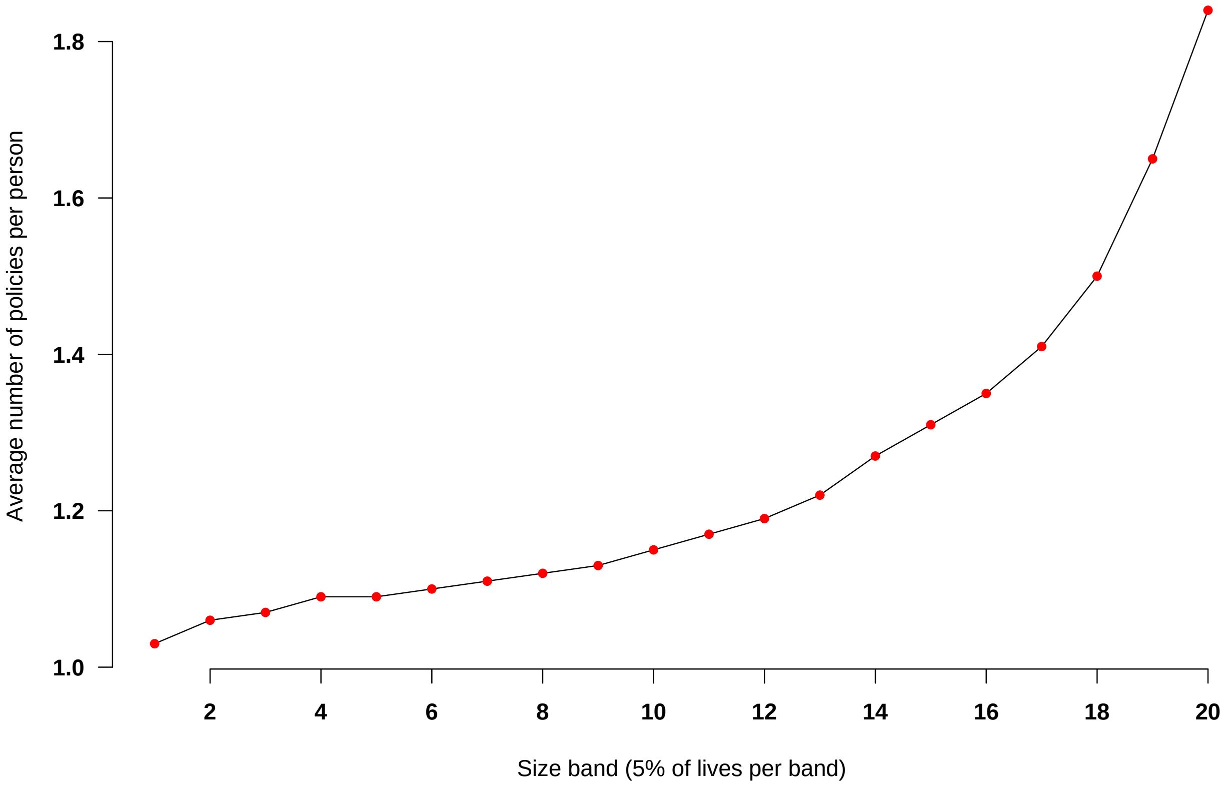
Life-insurance policies

- Longitudinal study with continual recruitment
- Detailed personal data
- High-quality: role of money and legal liability!
- Large-scale: typically tens or hundreds of thousands of policies
- Left-truncated: only adults buy insurance policies

Data preparation

- Data is policy-oriented
- People have multiple policies
- Need to ensure independence assumption
- Need to find n independent lives behind p dependent policies ($p \geq n$)
- Process of *deduplication*

Wealth and duplicates



Source: Richards and Currie (2009)

Deduplication challenges

Problem: client identifier rarely reliable

Solution: use combination key made up from reliable fields, e.g.

— Date of birth

— Gender

— Surname

— First initial

— Postcode

What's in a name?

Problem: teleserviced data contains mis-spellings of same surname, e.g.

— Ritchie

— Richie

— Richey

— Richey

Solution: use metaphone encoding of names

What's in a name?

Problem: metaphone structured for Anglo-Saxon names. What about

— Muhammed

— Muhammad

— Mohammed?

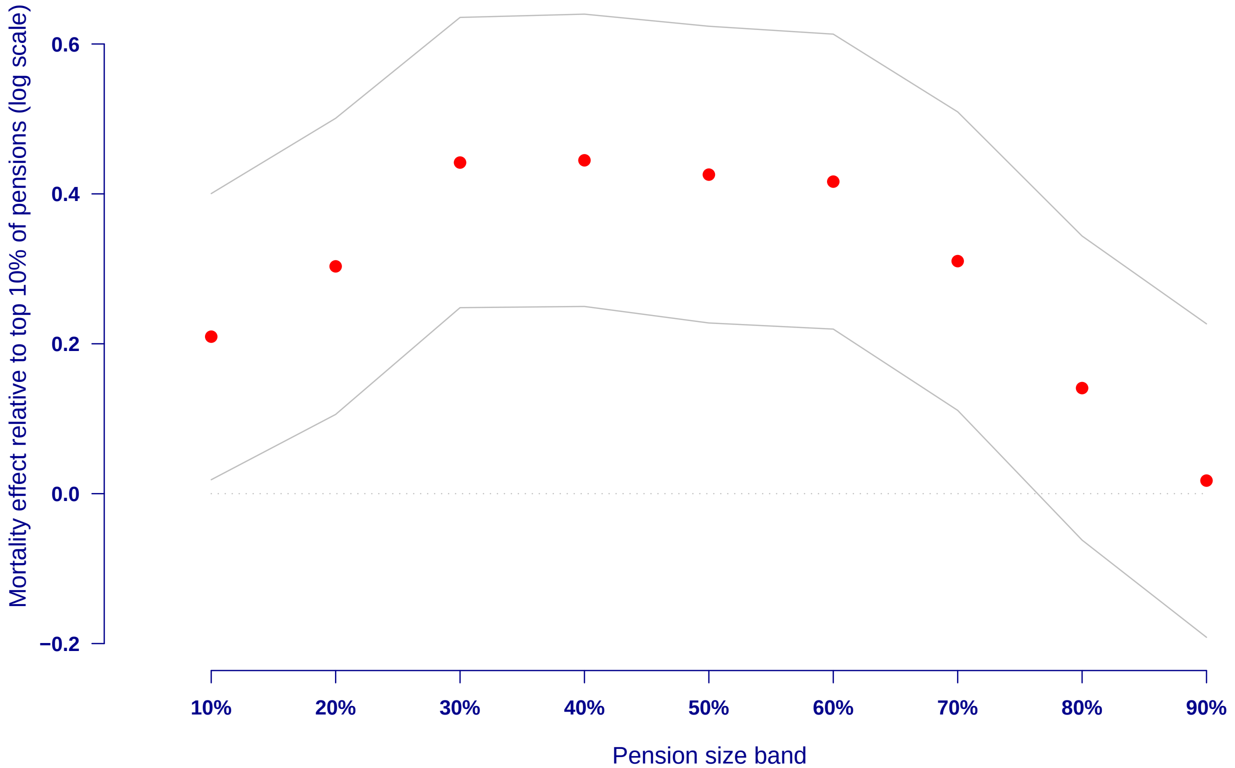
Solution: use double metaphone encoding of Philips (1990)

3. Risk factors

Traditional risk factors

- Age and gender universally used
- Pension size as proxy for wealth and income

Weakness of pension size



Modern risk factors

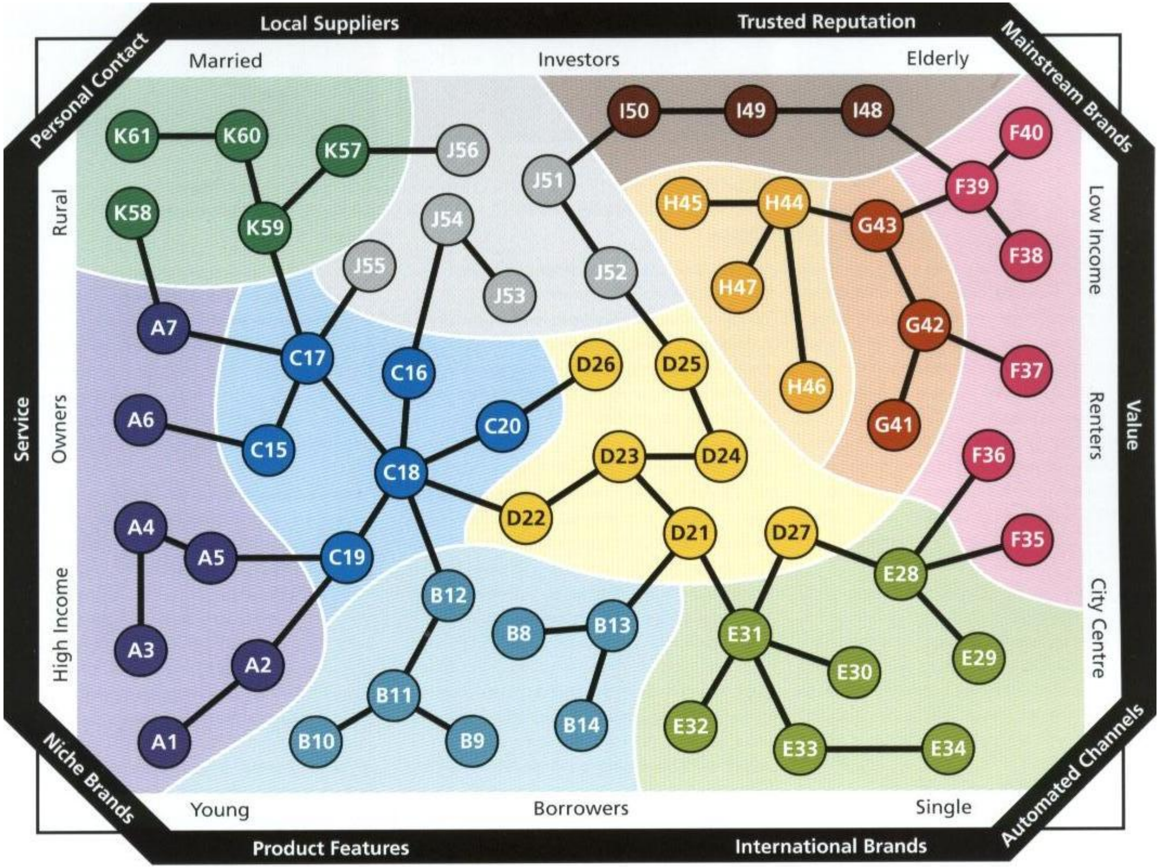
- Pension size imperfect proxy for wealth or income
- Postcode used to augment picture
- Postcodes now routinely used for pricing annuities

Anatomy of a UK postcode

Postcodes

- 1.6 million residential postcodes
- Each maps to a *geodemographic type*

Geodemographic example — Mosaic



Source: Experian Ltd.

Geodemographic examples

EH4 2AB → Mosaic Type A02 (“Cultural Leadership”)

EH4 2AB → Acorn Type 13 (“Prosperous Professionals”)

EH4 2AB → P^2 Type C07 (“Contented Families”)

Relative strength of risk factors

Parameter	Estimate	Z-value	p-value
Age	0.117945	145.29	0
Gender.M	0.403402	31.32	0
Intercept	-12.6977	-186.12	0
Mosaic.B	0.166925	4.97	0
Mosaic.C	0.121779	5.11	0
Mosaic.D	0.341533	13.53	0
Mosaic.E	0.269638	6.35	0
Mosaic.F	0.559107	15.41	0
Mosaic.G	0.52112	17.17	0
Mosaic.H	0.414819	16.25	0
Mosaic.I	0.355807	12.11	0
Mosaic.J	0.0731409	2.84	0.0045
Mosaic.K	0.0901384	2.41	0.0159

4. Time- and phase-varying factors

Extending basic model structure

Basic mortality law is static apart from age, e.g. Gompertz Law is:

$$\mu_x = \exp\left(\alpha + \beta x\right)$$

Extending basic model structure — I

- Can add duration since contract start, r, e.g.

$$\mu_x = \exp\left(\alpha + \beta x + \gamma r\right)$$

Extending basic model structure — II

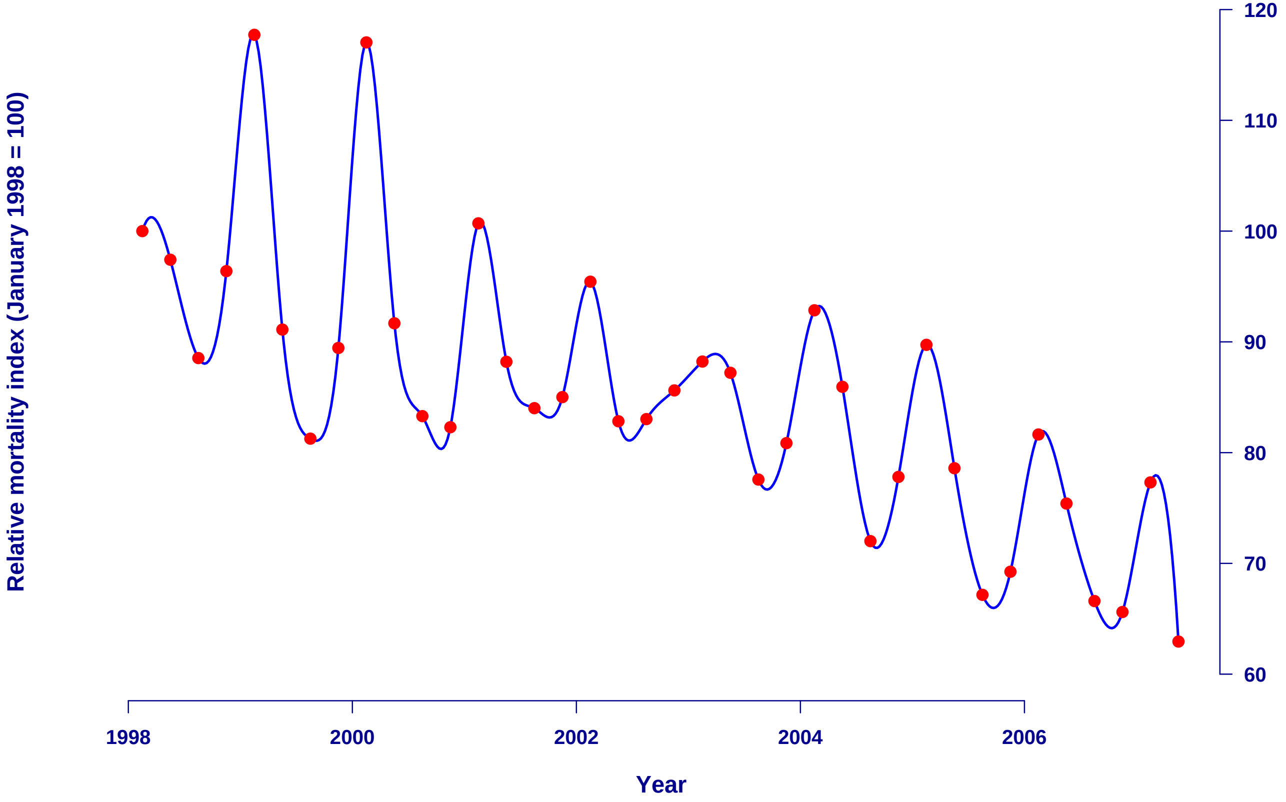
- Can add calendar time, y, e.g.

$$\mu_x = \exp\left(\alpha + \beta x + \gamma r + \delta y\right)$$

Extending basic model structure — III

- Can vary α piece-wise, e.g. for seasonal effects

Example phase risk factor: seasonal mortality



Source: Longevitas Ltd calculations using mortality experience between ages 60–95 for an annuity portfolio. Cox survival model with age, gender and calendar period (season).

5. Conclusions and questions

- Single unified procedure for mortality, longevity and persistency
- Data preparation is important, especially deduplication
- Insured data is ideally suited for survival models
- Geodemographic group is a powerful predictor of mortality
- Preprints available at the front

LONGEVITAS

References

Akaike, H. **1987** *Factor analysis and AIC*, Psychometrica, **52**, 317–333

Beard, R. E. **1959** *Note on some mathematical mortality models. In: The Lifespan of Animals, G. E. W. Wolstenholme and M. O'Connor (eds.)*, Little, Brown, Boston, 302–311

Cox, D. R. **1972** *Regression models and life tables*, Journal of the Royal Statistical Society, Series B, **24**, 187–220 (with discussion)

Gompertz, B. **1825** *The nature of the function expressive of the law of human mortality*, Philosophical Transactions of the Royal Society, **115**, 513–585

Perks, W. **1932** *On some experiments in the graduation of mortality statistics*, Journal of the Institute of Actuaries, **63**, 12–40

Philips, L. **1990** *Hanging on the metaphone*, Computer Language, 1990, **7** (12), 39–43

Richards, S. J. **2008** *Applying survival models to pensioner mortality data*, British Actuarial Journal (to appear)