Staple Inn, London

Applying survival models to pensioner data

Stephen Richards
25th February 2008

Plan of talk

1. The need for modelling
2. Data preparation
3. Geodemographic models
4. Selecting a model
5. Checking financial applicability of a model
6. Conclusions and questions

1. The need for modelling

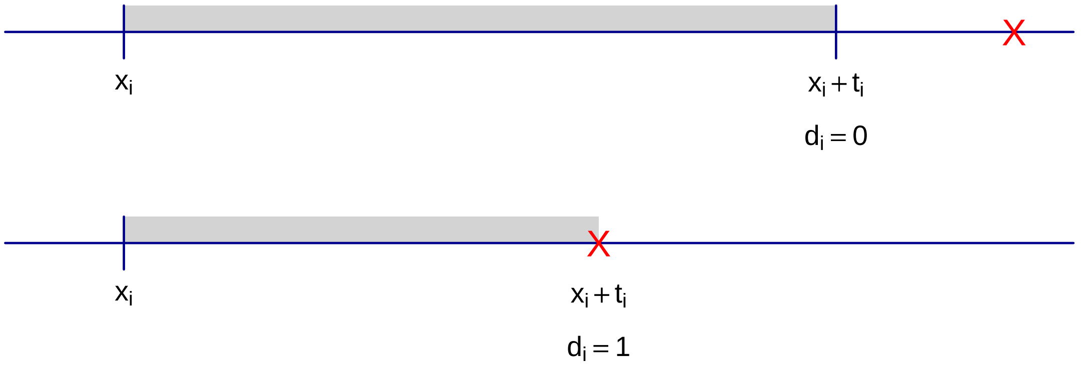
Financial impact of lifestyle

Financial impact of mortality rating factors

Factor	**Step change**	**Reserve**	**Change**
Base case	-	13.39	-
Gender	Female-male	12.14	-9.3%
Lifestyle	Top-bottom	10.94	-9.9%
Duration	Short-long	9.88	-9.7%
Pension size	Large-small	9.36	-5.2%
Region	South-North	8.90	-4.9%
Overall	-	-	-33.6%

Source: Richards and Jones (2004), page 39.

Survival models



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

Survival models

- Time observed, t_i, is *waiting time* (a.k.a. *central exposed-to-risk*)
- 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*

2. Data preparation

Four stages of data preparation

1. Extraction
2. Validation
3. Deduplication
4. Profiling

Data preparation: extraction

- Prefer data direct from payment system, *not* valuation extracts
- Dates, not ages

Data preparation: validation

- Validity, e.g. M or F for gender, *not* X or blank
- Consistency, e.g. commencement date *after* date of birth
- Sense, e.g. number of people born on 01/01/1901

Data preparation: deduplication

- Payment systems policy- or benefit-orientated
- Multiple records per person common
- Multiple records often correlated with wealth
- Must *deduplicate* to ensure independence assumption valid

Data preparation: deduplication

- How to recognize duplicates?
- System client ID typically unreliable
- NI number often not available (or unreliable)
- Create deduplication key from basic data

Data preparation: deduplication

- Date of birth
- Gender
- Surname
- Forename (first initial only)
- Postcode

Examples of matching forename fields

Surname	Forename(s)	Comment
Richards	Stephen	First initial only used.
Richards	Stephen J	First initial only used.
Richards	Steven	First initial only used.
Richards	S	First initial used.
Richards	Mr S	Title skipped, first initial used.
Richards	Rev Stephen J	Title skipped, first initial only used.

Matching surnames using double metaphone

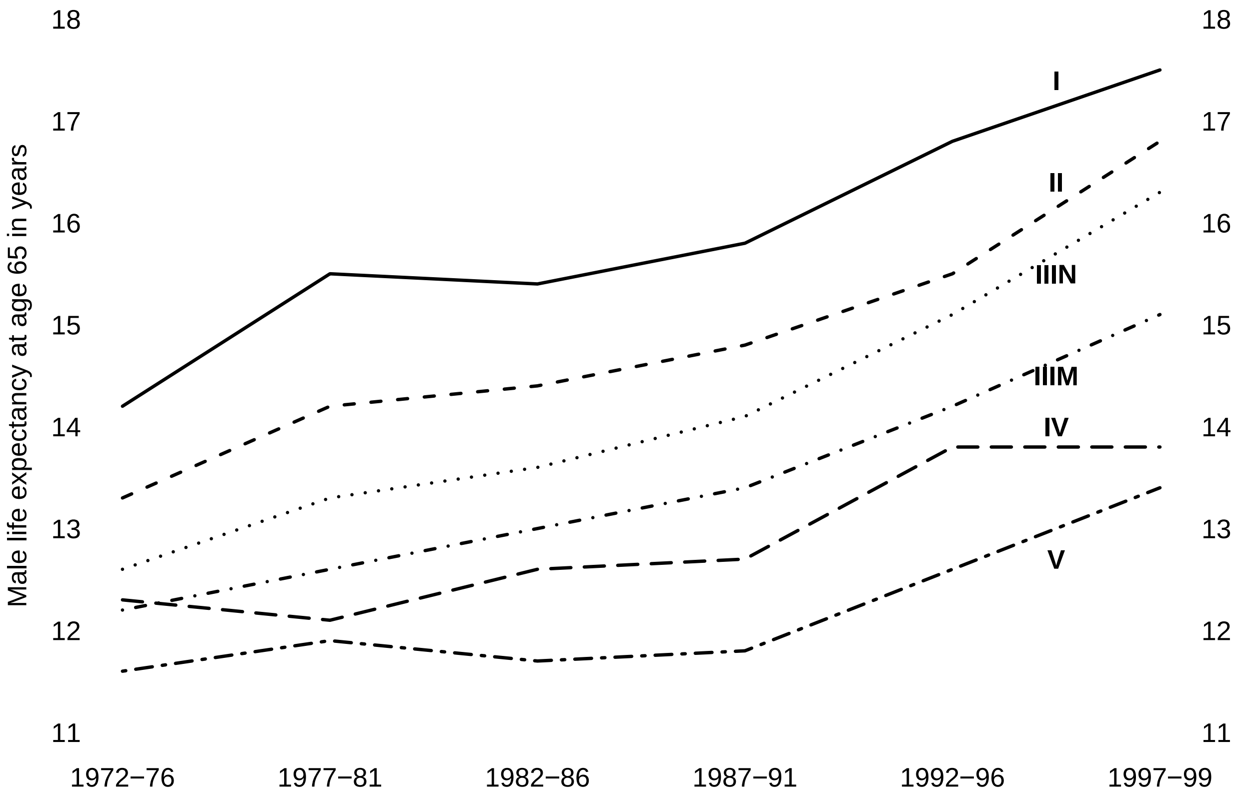
Record	Surname	Initial	Comment
1	Richie	G	
2	Ritchie	G	Match on surname in record 1.
3	Mohammed	A	
4	Muhammed	A	Match on surname in record 3.
5	Mohammad	A	Match on surname in record 3.
6	Mahamad	A	Match on surname in record 3.
7	Muammad	A	Match on surname in record 3.
8	Desantis	J	
9	D'Santis	J	Match on surname in record 8.
10	DE-SANTIS	J	Match on surname in record 8.

Source: Own examples using algorithm described in Philips (1990).

Data preparation: profiling

3. Geodemographic models

Retirement life expectancy by socio-economic group



Source: ONS Longitudinal Survey.

Why fund size is no longer reliable

- Stakeholder fund of £8,583
- Poor? Higher-mortality group?
- But AVC fund elsewhere of £42,808...
- ...giving total fund of £51,391...
- ...so not poor and likely light mortality!

3. Geodemographic models

- Use address or postcode to derive *geodemographic profile*
- Need full, two-part postcode in U.K.
- Options: Mosaic, FSS (both from Experian) or Acorn (from CACI)
- Examples:

  EH4 2AB → Mosaic Type 02 ("Cultural Leadership")

  EH4 2AB → Acorn Type 13 ("Prosperous Professionals")

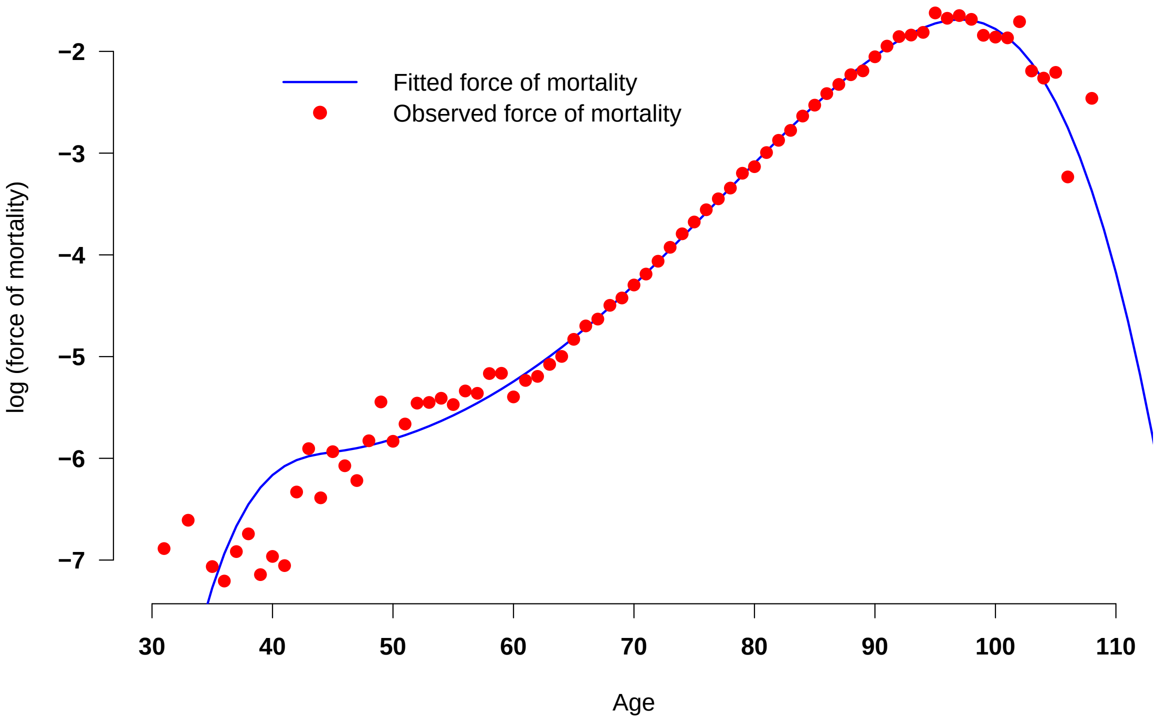
Final checks on data: Cramer's V

	Gender	**Region code**	**Size**	**Status**	**Type**
Birth year	21.6	3.1	11.4	54.4	4.0
Gender		4.8	16.1	12.4	5.6
Region code			5.9	6.4	20.6
Size				17.4	9.7
Status					10.4

Source: Own calculations of Cramer's V statistic for life-office pensioner data set, all ages. "Type" is the Experian Postcode Mosaic type code. "Status" is a boolean flag for whether death has occured (1) or not (0).

4. Selecting a model

Crude force of mortality



Source: Observed force of mortality (●) together with P-spline regression results in blue. Only the mortality between ages 60 and 100 shows regular behaviour suitable for a mortality law. Longevitas Ltd calculations using mortality experience of a portfolio of life-office pensioners.

Some mortality laws

Gompertz (1825)	$\mu_x = e^{\alpha+\beta x}$
Makeham (1859)	$\mu_x = e^{\epsilon} + e^{\alpha+\beta x}$
Perks (1932)	$\mu_x = \dfrac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}$
Beard (1959)	$\mu_x = \dfrac{e^{\alpha+\beta x}}{1+e^{\alpha+\rho+\beta x}}$
Makeham-Perks (1932)	$\mu_x = \dfrac{e^{\epsilon}+e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}$
Makeham-Beard (1932)	$\mu_x = \dfrac{e^{\epsilon}+e^{\alpha+\beta x}}{1+e^{\alpha+\rho+\beta x}}$

Model structure

$$\alpha_i = a_{baseline} + \sum_{j=1}^{m} z_{ij} a_j$$

$$\beta_i = b_{baseline} + \sum_{j=1}^{m} z_{ij} b_j$$

m components (factors) to the overall risk

a_j is a parameter for main effect of risk j

b_j is a parameter for the interaction of risk j with age

z_{ij} takes the value 1 when life i has risk factor j and the value 0 otherwise

Example model structure

Model with risk factors for both gender and smoker status:

$$\alpha_i = a_{baseline} + z_{i,male}a_{male} + z_{i,smoker}a_{smoker}$$
$$\beta_i = b_{baseline} + z_{i,male}b_{male} + z_{i,smoker}b_{smoker}$$

Choosing between models

Minimise Akaike's Information Criterion (Akaike, 1987):

$$AIC = -2\ell + 2n$$

where n is the number of parameters used in fitting the model and ℓ is the log-likelihood function evaluated at the joint maximum-likelihood estimate.

Simplifying complex factors

- Mosaic Type has 61 levels
- Acorn Type has 57 levels

$\longrightarrow$ neither convenient nor parsimonious!

- Consider various assignments to (say) three broad groups
- Use AIC to choose optimal assignment

Frailty models

- Gompertz model is $\mu_x = e^{\alpha+\beta x}$
- Re-write as $\mu_x = ze^{\beta x}$, where $z = e^\alpha$
- If z has gamma distribution, then *population* law is:

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

  even when each individual i follows $\mu_x = e^{\alpha_i+\beta x}$.
- Horiuchi and Coale (1990)

Model with age only

Age

Mortality law	AIC	AIC relative to Gompertz	Parameters
Gompertz	386742	0	2
Makeham	386744	2	3
Perks	386618	-124	2
Beard	386560	-182	3
Makeham-Perks	386620	-122	3
Makeham-Beard	386559	-183	4

Source: Own calculations using mortality experience of life-office pensioners aged between 60 and 95 between 2000-2006.

Model with age and gender

Age*Gender

Mortality law	AIC	AIC relative to Gompertz	Parameters
Gompertz	384824	0	4
Makeham	384826	2	5
Perks	384765	-59	4
Beard	384761	-63	5
Makeham-Perks	384762	-62	5
Makeham-Beard	384728	-96	6

Source: Own calculations using mortality experience of life-office pensioners aged between 60 and 95 between 2000-2006.

Model with age, gender and pension size-band

Age*(Gender+SizeBand)

Mortality law	AIC	AIC relative to Gompertz	Parameters
Gompertz	383562	0	8
Makeham	383564	2	9
Perks	383515	-47	8
Beard	383513	-49	9
Makeham-Perks	383510	-52	9
Makeham-Beard	383486	-76	10

Source: Own calculations using mortality experience of life-office pensioners aged between 60 and 95 between 2000-2006.

Model with age, gender and lifestyle

Age*(Gender+Lifestyle)

Mortality law	AIC	AIC relative to Gompertz	Parameters
Gompertz	383537	0	8
Makeham	383539	2	9
Perks	383518	-19	8
Beard	383520	-17	9
Makeham-Perks	383513	-24	9
Makeham-Beard	383509	-28	10

Source: Own calculations using mortality experience of life-office pensioners aged between 60 and 95 between 2000-2006.

Model with age, gender, lifestyle and pension size

Age*(Gender+Lifestyle+SizeBand)

Mortality law	**AIC**	**AIC relative to Gompertz**	**Parameters**
Gompertz	382597	0	12
Makeham	382599	2	13
Perks	382583	-14	12
Beard	382583	-14	13
Makeham-Perks	382575	-22	13
Makeham-Beard	382576	-21	14

Source: Own calculations using mortality experience of life-office pensioners aged between 60 and 95 between 2000-2006.

5. Checking financial applicability

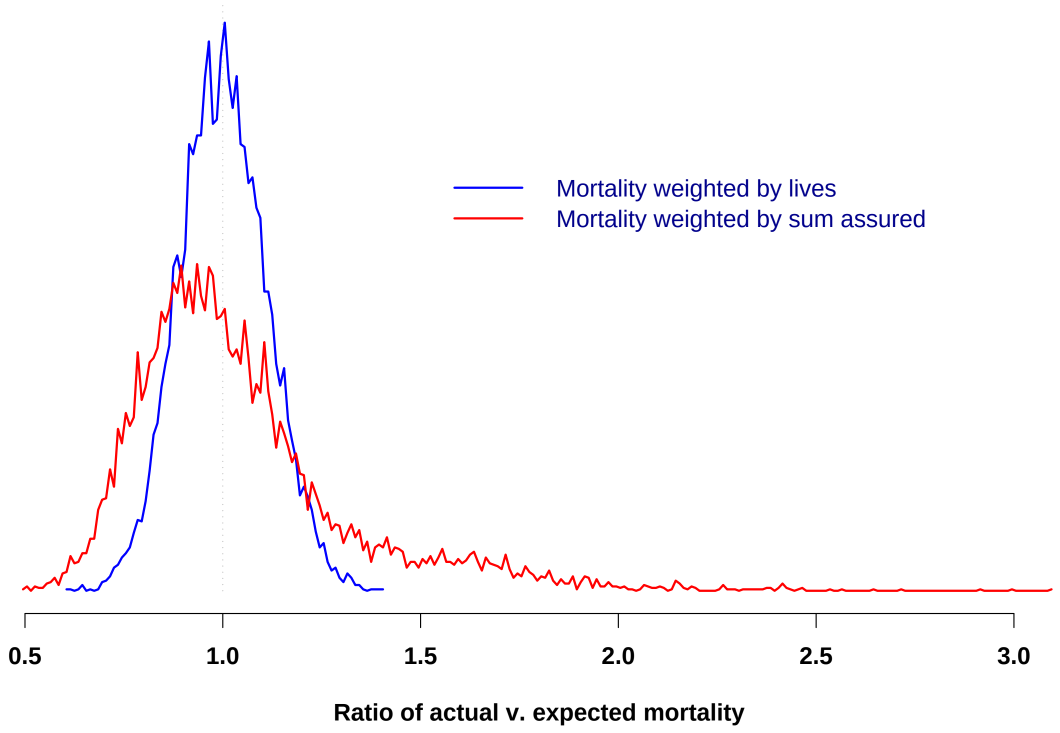
- Statistical models are lives-based…
- …whereas financial liabilities are not

Membership decile	Percentage of portfolio pension: (i) Life office	(ii) Pension schemes
1	54.3%	46.3%
2	15.2%	17.8%
3	9.4%	11.4%
4	6.6%	8.0%
5	4.9%	5.8%
6	3.6%	4.1%
7	2.7%	2.9%
8	1.8%	2.0%
9	1.1%	1.2%
10	0.4%	0.5%
Total	100.0%	100.0%

5. Checking financial applicability

- Can use bootstrapping to check model
- Sample randomly from portfolio
- Use model to predict mortality
- Compare with what actually happened
- Repeat sampling 10,000 times (say)

Bootstrapping a term assurance model



Source: Bootstrapped experience for portfolio of 50,000 term assurances, repeated 10,000 times. Longevitas Ltd calculations using model for mortality experience of a portfolio of nearly 1 million term-assurance policies between 2002 and end-2006. Model is Age + Gender + SelectPeriod + Smoker + JointLife + Product + Size.

6. Conclusions and questions

- Insured data is a natural fit for survival models
- Careful data preparation is *critical*
- Geodemographic models substantially enhance fit
- Models combining postcode and pension size usually better than either one alone
- Beard parameter, ρ, signals further variation to be explained
- Bootstrapping checks financial applicability of statistical models

Acknowledgements

- Gavin Ritchie
- Justin Armsworth, Experian Ltd

References I

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)

Cramer, H. **1999** *Mathematical Methods of Statistics*, Princeton University Press, ISBN13: 978-0-691-00547-8

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

Horiuchi, S. and Coale, A. J. **1990** *Age patterns of mortality for older women: an analysis using the age-specific rate of mortality change with age*, Mathematical Population Studies, **2**(4), 245–267

References II

LONGEVITAS DEVELOPMENT TEAM **2007** *Longevitas v2.2*, Longevitas Ltd, Edinburgh, United Kingdom. URL http://www.longevitas.co.uk

MACDONALD, A. S. **1996a** *An actuarial survey of statistical models for decrement and transition data, I: multiple state, Poisson and Binomial models*, British Actuarial Journal, **2** (1) 129–155

MACDONALD, A. S. **1996b** *An actuarial survey of statistical models for decrement and transition data, II: competing risks, non-parametric and regression models*, British Actuarial Journal, **2**(2) 429–448

MACDONALD, A. S. **1996c** *An actuarial survey of statistical models for decrement and transition data, III: counting process models*, British Actuarial Journal, **2**(3) 703–726

MAKEHAM, W. M. **1859** *On the law of mortality and the construction of annuity tables*, Journal of the Institute of Actuaries, **8**, 301–310

References III

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

R Development Core Team **2004** *R: a language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.r-project.org

Richards, S. J. and Jones, G. L. **2004** *Financial aspects of longevity risk*, SIAS

Richards, S. J. **2008** *Applying survival models to pensioner mortality data*, Institute Sessional Meeting Paper, June 2008