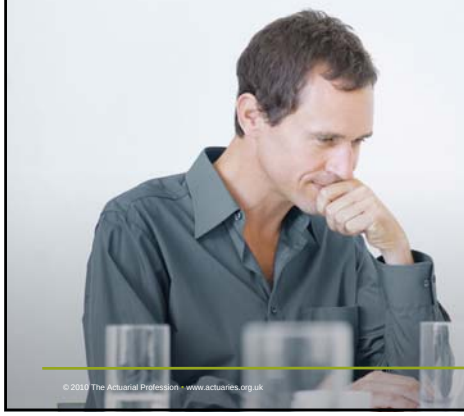
The Actuarial Profession
making financial sense of the future

Mortality and longevity seminar
Matthew Edwards



Grouping postcodes by mortality

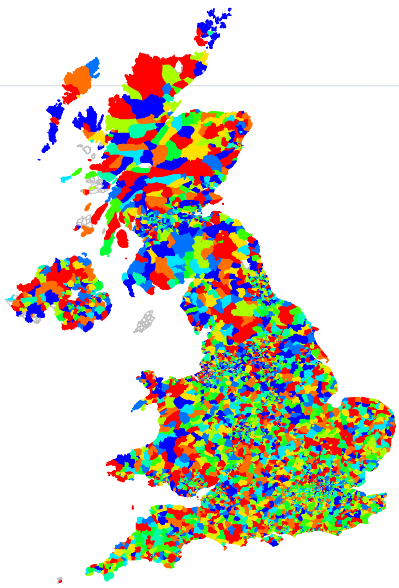
What are the challenges?

Contents

- Introduction
- Data and methodology
- Mortality clustering
- Issues
- Summary of results

Postcode mortality variations

- It has been clear for some time that mortality varies by postcode
- Map shows age-adjusted male mortality of TW pension schemes
- **Red** areas have average mortality double that of **blue** after adjusting for age, sex, and amount effect
- Financial dynamics of the bulk purchase annuity market have intensified research into this effect
- To what extent can the underlying mortality of small schemes be set by reference to postcode distribution of membership?



Contents

Multi-scheme work: data and methodology

- Multi-pension-scheme dataset totalling around 3 million people-years exposure, 100,000 deaths, from many pension schemes across sectors (total of nearly 20 schemes used for our work)
- Dataset restricted to pensioners only, ages 50+
- Dataset initially included two large insurance annuity portfolios
  - these were dropped from data after initial investigations showed their mortality characteristics to be substantially different from those of the pension scheme data, even after allowing for all possible postcode effects
- Data analysed using GLMs
- Postcode looked at by several conceptually distinct (and generally uncorrelated) 'cuts' to reduce vulnerability to any one method / categorisation / index – concentrating on postcode clusters based on mortality or similar (health) rather than purely on proxies (eg wealth, lifestyle).

Contents

Postcode clustering

- We are concerned about modelling the postcode effect on mortality. This requires grouping postcodes
- The grouping can be defined by
  - Distant proxies (financial lifestyle)
  - Close proxies (health – intuitively better than financial)
  - Mortality itself (**ideal** – if it's possible)
- Our postcode modelling has ended up based around four postcode clusters:
  - 2 mortality-based cuts
  - 1 'healthstyle' cut
  - 1 financial / lifestyle cut

Postcode clustering

- We are concerned about modelling the postcode effect on mortality. This requires grouping postcodes
- The grouping can be defined by
  - Distant proxies (financial lifestyle)
  - Close proxies (health – intuitively better than financial)
  - Mortality itself (**ideal** – if it's possible)
- Our postcode modelling based on four postcode clusters:
  - 2 mortality-based cuts
    - Clusters based on observed mortality in the dataset
    - Clusters based on observed UK population mortality
  - 1 'healthstyle' cut
  - 1 financial / lifestyle cut

Can we group by mortality?

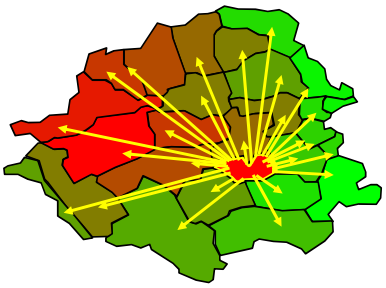
- Postcodes are tiny – typically 20-30 people
  - even fewer pensioners
  - even fewer deaths among pensioners
- Try postcode sector? (eg “RH2 0”)
  - of order of 5,000 people in population
  - in a large UK DB pensioner dataset, median may be only circa 30-40 pensioners
  - even fewer deaths
- Situation seems hopeless!

Construction of mortality postcode clusters

- Uses proprietary ‘spatial smoothing’ technique to remove random noise and find mortality patterns
- Methodology in common use in non-life claims pricing
- We construct non-contiguous postcode clusters defined with reference to similar levels of observed mortality (adjusted for age, sex, amount effect) from the dataset
- Uses optimised ‘smoothing’ to blend experience of low exposure and high exposure regions to improve goodness of fit on independent data portion
- Algorithm is credibility-based (according to exposure)

Mortality zones

- We wish to quantify the mortality effect arising from ‘micro-regions’
- Some regions have too little exposure to be at all credible, many will suffer from ‘noise’
- Solution: blend experience of one region with that of surrounding regions according to distance and credibility
- Credibility and spatial smoothing parameters are calculated from portion of data



Empirical field of work

- Vital to remember that we are working in an empirical field
- What works is what can be shown to be predictive on an independent part of the data
- The spatial smoothing algorithm will give us a potentially interesting set of postcode clusters to ‘try out’
- If they are predictive: good (and if not, bad!)
- The fact that the algorithm takes as a starting point an unfounded assumption is irrelevant

Methodology

- Derive model ignoring geography (ie model effect of age, amounts, duration etc)
- Define 'residual' r_i to be mortality effect of belonging to micro-region in question (via 'A-E' approach, with E from above model which allows for the relevant but non-geographic factors)
- Want to derive smoothed residual r_i^* representing real underlying mortality effect of micro-region
- $r_i^* = Z_i.r_i + (1 - Z_i) \times nearby\ r_i$
- … where Z is a measure of credibility

Deriving the smoothed residuals

$$r_i^* = Z(e_i).r_i + (1 - Z(e_i)) \Sigma\ e_j.r_j.f(d_{ij}) / \Sigma\ e_j.f(d_{ij})$$

where

r_i^* = smoothed residual — r_i = unsmoothed residual

$Z(e_i) = \{ e_i / (e_i + a) \}^m$ — e_i = exposure in region i

$d_{ij} = \{ (x_i - x_j)^2 + (y_i - y_j)^2 \}^{½}$

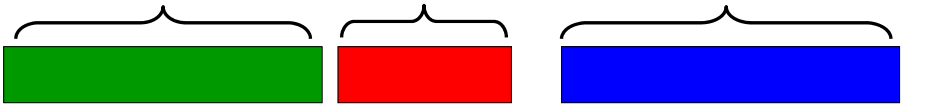
$f(d_{ij}) = 1/d_{ij}^n$ or $1/(d_{ij}^n + b^n)$ or $\exp(-n.d_{ij})$ etc

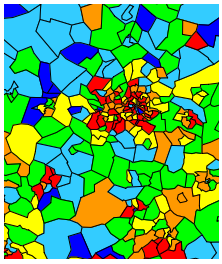
Mortality postcode clustering from actual experience: splitting the data to avoid self-prophesying

Calculate 'residuals'

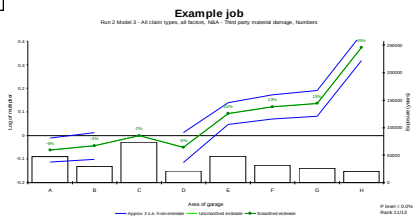
Seek parameters to optimise fit

Use to determine the cluster relativities

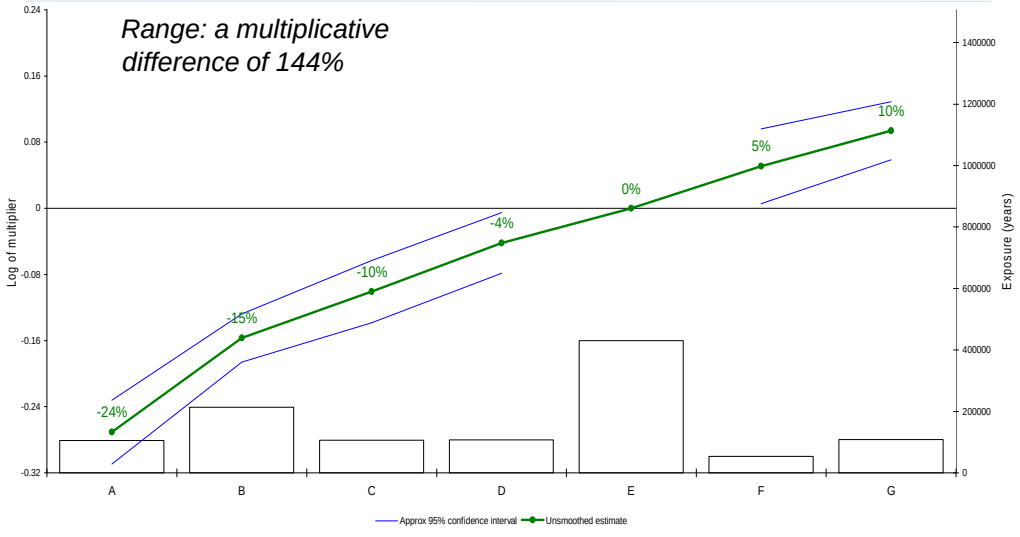




a, m, n, b



Scheme mortality spatial clusters



Contents

Issues with postcode mortality clustering: micro-regions and distance

- What level of 'micro-region' works?
  - Postcode sector can give good results
  - Best results with lower layer super output area
  - This is an NHS grouping, typical population 1,500
- Refining the distance metric
  - In the algorithm shown above, the effect of distance is assumed constant in all areas
  - This does not seem a reasonable assumption: compare one mile of variation in London with one mile in Yorkshire
  - Can introduce extra elements in the distance metric to accommodate this effect
  - For instance, alter the distance between any two areas according to relative density of those two areas

Different metrics

$r_i^* = Z(e_i).r_i + (1 - Z(e_i)) \Sigma_j e_j.r_j.f(d_{ij}) / \Sigma_j e_j.f(d_{ij})$ i

where

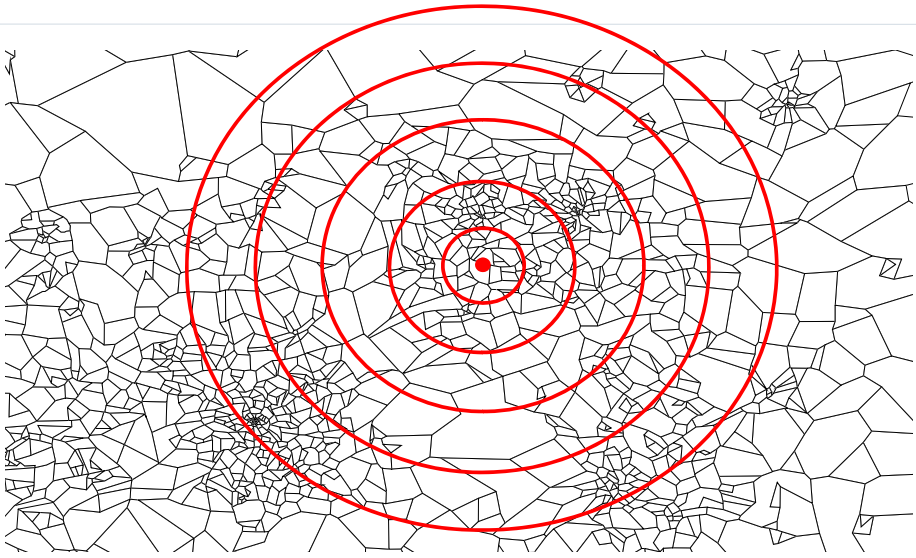
r_i^*= smoothed residual r_i = unsmoothed residual

$Z(e_i) = \{ e_i / (e_i + a) \}^m$ e_i = exposure in region i

$d_{ij} = \{ (x_i - x_j)^2 + (y_i - y_j)^2 + (s.q_i - s.q_j)^2 \}^{½}$

$f(d_{ij}) = 1/d_{ij}^n$ or $1/(d_{ij}^n + b^n)$ or $\exp(-n.d_{ij})$ etc

Limiting the definition of “neighbouring”



Contents

Recap of postcode clustering methods used

- 'ACORN' and 'HealthACORN' factors are off-the-shelf demographic clusters provided by CACI. Each of these groups postcodes into a number of different clusters according to perceived lifestyle (ACORN) and 'healthstyle' (HealthACORN).
- 'Scheme mortality clusters' are clusters of postcodes which have similar mortality characteristics from the schemes' data, using a proprietary algorithm to blend together low exposure areas to ensure sufficient statistical credibility.
- 'Population mortality clusters' are clusters of postcodes which have similar mortality characteristics (ages 50+) according to general population data.
- This allows us to rank groups of postcodes (from lower layer super output areas) by similar levels of population mortality (ages 50+)

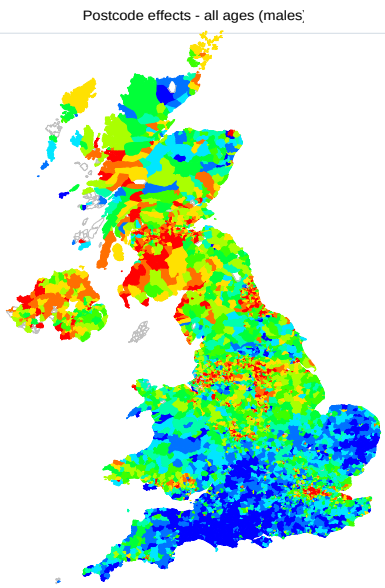
Summary of relative factor powers

Clustering method	Power	Contribution to total
Healthstyle	157%	33%
Pension amount	148%	29%
Scheme mortality	144%	27%
Population mortality	111%	7%
Lifestyle	105%	4%
Total effect	387%	100%

Note that the above summary is 'unfair' to population mortality clusters and lifestyle clusters, which would be substantially more predictive if looked at in isolation

Summary of postcode results via map

- Map shows combined effect of all four postcode clustering factors used (males, all ages)
- Map summary at postcode sector level, aggregating the postcode level results
- **Red** areas have average mortality 200% that of **blue**



Aggregating postcode-level mortality multipliers by post town

Post town	Average multiplier
KILBIRNIE	**154%**
BOOTLE (LIVERPOOL)	**153%**
LOCHGELLY	**153%**
KYLE	**151%**
CORBY	125%
BOLTON	125%
BANGOR	110%
ILFORD	110%
STAFFORD	110%
SIDCUP	100%
STOURBRIDGE	100%
SURBITON	**100%**
NEWMARKET	100%
LEIGHTON BUZZARD	100%
RYE	90%
WELLINGTON	90%
CHIPPING CAMPDEN	75%
SHERBORNE	75%
LYMINGTON	**69%**
LYME REGIS	**69%**
CHURCH STRETTON	**68%**
BROCKENHURST	**67%**
MONTACUTE	**64%**