Stochastic Frailty Models

Søren Fiig Jarner, CSO

“People always live forever when there is an annuity to be paid them”

Jane Austen, Sense and Sensibility (1811)
Agenda

- Mortality modelling objectives
  - Motivating example
- Stochastic frailty models
  - Frailty theory
  - Estimation
  - Model selection
  - Forecasting
- Females vs males
  - Gender gap
  - Separate and joint modelling
  - Cointegrating parameters
- Application to international data
Motivating example – international data

- International data from 1950 to 2006
  - 18 countries: USA, Germany, UK, France, Italy, Spain, Australia, Canada, Holland, Portugal, Austria, Belgium, Switzerland, Sweden, Norway, Finland, Iceland & Denmark

- Human Mortality Database

- Data consists of death counts and exposures
  - $D(t,x) =$ number of deaths
  - $E(t,x) =$ exposure (“years lived”)
Continued increase in life expectancy

- Life expectancy in 2006
  - Females ~ 82 years
  - Males ~ 77 years

- Life expectancy gain since 1950
  - Both sexes ~ 12 years

- Average annual increase
  - Both sexes ~ 0.22 years

- Persistent gender gap between females and males of 5 to 7 years

![International life expectancy graph](image-url)
Rates of improvement decreasing with age

- **Age-specific death rate**
  \[ m(t, x) = D(t, x) / E(t, x) \]

- **Rate of improvement in year \( t \)**
  \[ \rho(t, x) = -\log \frac{m(t, x)}{m(t-1, x)} \]

- **Averate rate of improvement**
  \[ \bar{\rho}(x) = \frac{1}{56} \log \frac{m(2006, x)}{m(1950, x)} \]

- **Rates of improvement in the range from 1% to 1.5% for most ages**

- **Higher rates of improvement for females than for males**

### Average annual rates of improvement

<table>
<thead>
<tr>
<th>Age</th>
<th>Females</th>
<th>Males</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>2.5%</td>
<td>2.0%</td>
</tr>
<tr>
<td>30</td>
<td>2.0%</td>
<td>1.5%</td>
</tr>
<tr>
<td>40</td>
<td>1.5%</td>
<td>1.0%</td>
</tr>
<tr>
<td>50</td>
<td>1.0%</td>
<td>0.5%</td>
</tr>
<tr>
<td>60</td>
<td>0.5%</td>
<td>0.0%</td>
</tr>
<tr>
<td>70</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>80</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>90</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>100</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>
Rectangularization of the survival curve …
... but increasing improvements in old-age mortality

**Female average rates of improvement**

<table>
<thead>
<tr>
<th>Age</th>
<th>Rate of improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>50-59</td>
<td>0.0%</td>
</tr>
<tr>
<td>60-69</td>
<td>0.5%</td>
</tr>
<tr>
<td>70-79</td>
<td>1.0%</td>
</tr>
<tr>
<td>80-89</td>
<td>1.5%</td>
</tr>
<tr>
<td>90-99</td>
<td>2.0%</td>
</tr>
</tbody>
</table>

**Male average rates of improvement**

<table>
<thead>
<tr>
<th>Age</th>
<th>Rate of improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>50-59</td>
<td>0.0%</td>
</tr>
<tr>
<td>60-69</td>
<td>0.5%</td>
</tr>
<tr>
<td>70-79</td>
<td>1.0%</td>
</tr>
<tr>
<td>80-89</td>
<td>1.5%</td>
</tr>
<tr>
<td>90-99</td>
<td>2.0%</td>
</tr>
</tbody>
</table>
Mortality modelling objectives

- **Stylized facts**
  - Rates of improvement decreasing with age
  - Old-age rates of improvements increase over time
  - Persistent gender gap

- **Modelling objectives**
  - Simple dynamics generating stylized facts *endogenously*
  - Model covering mortality from early adulthood to the oldest-old, i.e. age 20-120
  - Good fit to historical data
  - Biologically plausible forecasts
    - mortality increase with age and decrease over time
    - forecasted rates of improvement in old-age mortality potentially *higher* than those observed historically
  - Efficient and robust estimation
Stochastic Frailty Models
**Frailty theory**

- Heterogenous populations
  - People are genetically different
  - Only the more robust will attain high ages

- Multiplicative frailty model (Vaupel, Manton & Stallard (1979))
  - Hazard for an individual with (unobserved) frailty $z$
    \[
    \mu(x; z) = z\mu_{base}(x)
    \]
    where $\mu_{base}(x)$ is a baseline hazard describing the age effect
  - Population hazard
    \[
    \mu(x) = E[Z|x]\mu_{base}(x)
    \]
    where $E[Z|x]$ is the (conditional) mean frailty in the cohort at age $x$
**Frailty at age 80**

**Conditional frailty at age 80**

Conditional mean frailty $E[Z|x=80] = 0.8313$
Calculating the mean frailty

- Integrated intensities
  \[ I(x) = \int_0^x \mu_{\text{base}}(u) du \]
  \[ H(x) = \int_0^x \mu(u) du \]

- Laplace transform of frailty distribution (at birth)
  \[ L(s) = \mathbb{E}[\exp(-sZ)] \]

Theorem, e.g. Hougaard (1984)

\[ H(x) = \nu[I(x)] \]
\[ \mathbb{E}[Z \mid x] = \nu'[I(x)] = \nu'[\nu^{-1}\{H(x)\}] \]

where \( \nu(s) = -\log L(s) \)
Gamma frailty

Let $Z$ be Gamma-distributed with mean 1 and variance $\sigma^2$

$$L(s) = \left(1 + \sigma^2 s\right)^{-1/\sigma^2}$$

$$\mathbb{E}[Z \mid x] = \left(1 + \sigma^2 I(x)\right)^{-1} = \exp\left(-\sigma^2 H(x)\right)$$

Inverse Gaussian frailty

Let $Z$ be inverse Gaussian with mean 1 and variance $\sigma^2$

$$L(s) = \exp\frac{1-\sqrt{1+2\sigma^2 s}}{\sigma^2}$$

$$\mathbb{E}[Z \mid x] = \left(\sqrt{1 + 2\sigma^2 I(x)}\right)^{-1/2} = \left(1 + \sigma^2 H(x)\right)^{-1}$$
Old-age mortality plateau

- Death rates are (approximately) exponentially increasing with age, but slower increase at very old ages

- Gompertz-Makeham
  \[ \mu(x) = \alpha \exp(\beta x) + \gamma \]

- Gompertz-Makeham with Gamma-frailty
  \[ \mu(x; z) = z \alpha \exp(\beta x) + \gamma \]
  has population hazard of logistic form
  \[ \mu(x) = \frac{\kappa \exp(\beta x)}{1 + \varphi \exp(\beta x)} + \gamma \]
Population dynamics

- Frailty model for time- and age-dependent mortality

\[ \mu(t, x; z) = z \mu_{\text{base}}(t, x) \]

where \( \mu_{\text{base}}(t, x) \) is a baseline hazard describing the time and age effect.

- Population hazard

\[ \mu(t, x) = E[Z|t, x] \mu_{\text{base}}(t, x) \]

where \( E[Z|t, x] \) is the mean frailty in the cohort of age \( x \) at time \( t \), i.e. cohort born at time \( t-x \).

- Mean frailty (assuming same frailty distribution at birth for all cohorts)

\[
E[Z \mid t, x] = v'[I(t, x)] = v'[v^{-1}\{H(t, x)\}] \\
I(t, x) = \int_0^x \mu_{\text{base}}(u + t - x, u) du \\
H(t, x) = \int_0^x \mu(u + t - x, u) du
\]
Rates of improvement

- Assumptions
  - for fixed $x$, $\mu_{base}(t,x)$ decreases to 0 as $t$ tends to infinity
  - same frailty distribution at birth for all cohorts with mean one

- Dynamics
  - $I(t, x) = \int_0^x \mu_{base}(u + t - x, u) du$ decreases to 0 as $t$ tends to infinity
  - $E[Z | t, x] = \nu'[I(t, x)]$ increases to 1 as $t$ tends to infinity

- Rate of improvement of (frailty dependent) age-specific mortality

\[
\rho(t, x) = -\frac{\partial}{\partial t} \log \mu(t, x) = -\frac{\partial}{\partial t} \log E[Z | t, x] - \frac{\partial}{\partial t} \log \mu_{base}(t, x)
\]

Rate of improvement of old-age mortality:

- Initially, improvements in $\mu_{base}$ are (partially) offset by increase in mean frailty
- Eventually, mean frailty is close to 1 and improvements in $\mu$ and $\mu_{base}$ will be the same
A stochastic frailty model is a model for population mortality of the form

\[
\mu(t, x) = \mathbb{E}[Z|t, x] \mu_{base}(t, x)
\]

where

- frailty, \( Z \), is Gamma-distributed with mean 1 and variance \( \sigma^2 \) (same for all cohorts)
- \( \mu_{base}(t, x) = F(x; \theta_t) \), where \( F(x; \theta) \) is a parametric mortality law, e.g. Gompertz
- \( (\theta_t) \) time-series of parameters describing dynamics of individual mortality over time

Example: Time-dependent Gamma-Gompertz model

\[
\mu(t, x) = \mathbb{E}[Z|t, x; \sigma^2] \alpha_t \exp(\beta_t x)
\]
- Death counts are assumed independent with

\[ D(t, x) \sim \text{Poisson}(\mu(t, x)E(t, x)) \]

where \( \mu(t, x) = E[Z|t, x] \mu_{\text{base}}(t, x) = E[Z|t, x; \sigma^2]F(x; \theta_t) \)

- Likelihood function

\[
L(\theta, \sigma^2) \propto \prod_{t,x} \left\{ E[Z|t,x; \sigma^2]F(x; \theta_t)E(t, x) \right\}^{D(t,x)} \exp\left\{ -E[Z|t,x; \sigma^2]F(x; \theta_t)E(t, x) \right\}
\]

where \( E[Z|t,x; \sigma^2] = \left[ 1 + \sigma^2 I(t, x; (\theta_{s_{t-x<s<t}})) \right]^{-1}, \ I(t, x; (\theta_{s_{t-x<s<t}})) = \sum_{s=0}^{x-1} F(u; \theta_{u+t-x}) \)

- This likelihood function is very hard to optimize!
From individual hazard to population hazard

- Remember (property of Gamma frailty)
  \[
  E[Z \mid t, x] = \left(1 + \sigma^2 I(t, x) \right)^{-1} = \exp\left(-\sigma^2 H(t, x)\right)
  \]
  
  \[
  I(t, x) = \int_0^x \mu_{base}(u + t - x, u) du
  \]
  
  \[
  H(t, x) = \int_0^x \mu(u + t - x, u) du
  \]

- Hence

  \[
  E[Z \mid t, x; \sigma^2] = \left[1 + \sigma^2 I(t, x; (\theta_s)_{t-x<s<t}) \right]^{-1} \approx \exp\left[-\sigma^2 \tilde{H}(t, x)\right]
  \]

  where \(\tilde{H}(t, x)\) is the observed integrated population hazard

  \[
  \tilde{H}(t, x) = \sum_{u=0}^{x-1} m(t - x + u, u)
  \]

  \[
  m(t, x) = D(t, x) / E(t, x)
  \]
We will base estimation on the pseudo likelihood

\[ \tilde{L}(\theta, \sigma^2) \propto \prod_{t,x} \left\{ \exp\left[-\sigma^2 \tilde{H}(t, x)\right] F(x; \theta) E(t, x) \right\}^{D(t,x)} \exp\left[-\sigma^2 \tilde{H}(t, x)\right] F(x; \theta) E(t, x) \]

corresponding to the model

\[ D(t, x) \sim \text{Poisson}(\tilde{\mu}(t, x) E(t, x)) \]

where

\[ \tilde{\mu}(t, x) = \exp\left[-\sigma^2 \tilde{H}(t, x)\right] F(x; \theta) \]

\[ \tilde{H}(t, x) = \sum_{u=0}^{t-1} m(t - x + u, u) \]

\[ m(t, x) = D(t, x) / E(t, x) \]

This likelihood is much easier to handle!
Calculating integrated observed hazard

- Data: \( \{(D(t, x), E(t, x)) : t_{\text{min}} \leq t \leq t_{\text{max}}, x_{\text{min}} \leq x \leq x_{\text{max}}\} \)

- Integrated observed hazard depends on data outside data window

\[
\tilde{H}(t, x) = \sum_{u=0}^{x-1} m(t - x + u, u)
\]

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

- We “extend” the data window by setting

\[
m(t, x) = 0 \quad \text{for} \quad x < x_{\text{min}}
\]

\[
m(t, x) = m(t_{\text{min}}, x) \quad \text{for} \quad t < t_{\text{min}}
\]
Assume $\sigma^2$ known (fixed)

The parameters ($\theta_t$) can be estimated independently of each other

Parameter $\theta_t$ is estimated from the (partial) model

$$D(t, x) \sim \text{Poisson}\left(\exp[-\sigma^2 \tilde{H}(t, x)]E(t, x)F(x, \theta_t)\right) \text{ for } x_{\min} \leq x \leq x_{\max}$$

i.e. based only on data from year $t$

Maximum likelihood estimation can use analytical properties of $F(x; \theta)$
Estimating $\sigma^2$

- Assuming time-dependent parameters ($\theta_t$) fixed
- $\sigma^2$ is estimated from the model for all data
- Likelihood is log-concave, i.e. local maximum = global maximum

\[
\tilde{l}(\sigma^2) = \log \tilde{L}(\theta, \sigma^2) = \sum_{t,x} -\sigma^2 D(t, x)\tilde{H}(t, x) - \exp[-\sigma^2 \tilde{H}(t, x)]F(x; \theta_t)E(t, x) + \text{const}
\]

\[
\frac{\partial \tilde{l}}{\partial \sigma^2} = \sum_{t,x} - D(t, x)\tilde{H}(t, x) + \tilde{H}(t, x) \exp[-\sigma^2 \tilde{H}(t, x)]F(x; \theta_t)E(t, x)
\]

\[
\frac{\partial^2 \tilde{l}}{\partial (\sigma^2)^2} = \sum_{t,x} - \tilde{H}^2(t, x) \exp[-\sigma^2 \tilde{H}(t, x)]F(x; \theta_t)E(t, x) < 0
\]

- Likelihood equation: \[\frac{\partial \tilde{l}}{\partial \sigma^2}(\hat{\sigma}^2) = 0\]

- Efficient and robust estimation of $\sigma^2$ using monotonicity of \[\frac{\partial \tilde{l}}{\partial \sigma^2}\]
A stochastic frailty model can be estimated using a switching algorithm:

1. Calculate integrated population hazard, $\tilde{H}(t, x)$
2. Choose initial value for $\sigma^2$, e.g. $\sigma^2=0$
3. Iterate the following steps until convergence
   - Estimate time-dependent parameters $(\theta_i)$ given current value of $\sigma^2$
   - Estimate $\sigma^2$ given current value of $(\theta_i)$

It can be shown that the EM algorithm always converges to a (local) maximum of the likelihood function
A generalized stochastic frailty model is a model of the form

\[ \mu(t, x) = E[Z|t, x] \mu_{\text{base}}(t, x) + \mu_{\text{acc}}(t, x) \]

where

- frailty, Z, is Gamma-distributed with mean 1 and variance \( \sigma^2 \) (same for all cohorts)
- \( \mu_{\text{base}}(t, x) = F(x; \theta_t) \), where \( F(x; \theta) \) is a parametric mortality law, e.g. Gompertz
- \( (\theta_t) \) time-series of parameters describing dynamics of individual mortality over time
- \( \mu_{\text{acc}}(t, x) = G(x; \nu_t) \), where \( G(x; \nu) \) is a parametric mortality law describing frailty independent causes of death, e.g. accidents
- \( (\nu_t) \) time-series of parameters describing dynamics of frailty independent mortality

Example: Time-dependent Gamma-Makeham model

\[ \mu(t, x) = E[Z|t, x; \sigma^2] \alpha_t \exp(\beta_t x) + \gamma_t \]
Data model

- Death counts are assumed independent with

\[ D(t, x) \sim \text{Poisson}(\mu(t, x)E(t, x)) \]

where \( \mu(t, x) = E[Z|t, x] \mu_{\text{base}}(t, x) + \mu_{\text{acc}}(t, x) \)

- Estimation will be based on the model

\[ D(t, x) \sim \text{Poisson}(\tilde{\mu}(t, x)E(t, x)) \]

where \( \tilde{\mu}(t, x) = \exp[-\sigma^2 \tilde{H}(t, x)] \text{F}(x; \theta) + \text{G}(x; \nu, \tau) \)

\[ \tilde{H}(t, x) = \sum_{u=0}^{x-1} \{m(t-x+u, u) - G(u; \nu_{t-x+u})\} \]

(adjusted integrated population hazard)

\[ m(t, x) = D(t, x) / E(t, x) \]
Competing risks model

- Competing risks model
  \[ D(x) \sim \text{Poisson}(E(x)[\mu_1(x; \theta) + \mu_2(x; \nu)]) \]

- Interpretation: Two different, independent sources of death
- Complicated likelihood (not log-concave)

- If death counts had been recorded according to source of death then
  \[ D_1(x) \sim \text{Poisson}(E(x)\mu_1(x; \theta)) \]
  \[ D_2(x) \sim \text{Poisson}(E(x)\mu_2(x; \nu)) \]

  with \( D_1(x) \) and \( D_2(x) \) independent and \( D(x) = D_1(x) + D_2(x) \)
- Simple likelihood, easy to estimate \( \theta \) and \( \nu \)
Expectation-maximization (EM) algorithm

- EM algorithm:
  - Choose initial values $\theta^0$ and $\nu^0$
  - **E-step**: Treat $D_1$ and $D_2$ as missing data and calculate expected value of full log-likelihood given data and current value of parameters
    
    $Q(\theta, \nu) = E[l(\theta, \nu; D_1, D_2) \mid D, \theta^i, \nu^i]$ 

    where (omitting the $x$ argument)
    
    $l(\theta, \nu; D_1, D_2) = \sum_x \{D_1 \log \mu_1(\theta) - E\mu_1(\theta)\} + \sum_x \{D_2 \log \mu_2(\nu) - E\mu_2(\nu)\} + \text{const}$

    $D_1 \mid D, \theta^i, \nu^i \sim \text{Binom}(D, \mu_1(\theta^i) / [\mu_1(\theta^i) + \mu_2(\nu^i)])$

    $D_2 \mid D, \theta^i, \nu^i \sim \text{Binom}(D, \mu_2(\nu^i) / [\mu_1(\theta^i) + \mu_2(\nu^i)])$

  - **M-step**: Maximize $Q$ to obtain new estimates $\theta^{i+1}$ and $\nu^{i+1}$

  - Iterate E-step and M-step till convergence
Both the E-step and the M-step are easy to implement.

Formally the EM algorithm corresponds to iteratively estimating the models:

\[ D_1^i(x) \sim \text{Poisson}(E(x)\mu_1(x; \theta)) \]

\[ D_2^i(x) \sim \text{Poisson}(E(x)\mu_2(x; \nu)) \]

where

\[ D_1^i(x) = D(x) \frac{\mu_1(x;\theta^i)}{\mu_1(x;\theta^i) + \mu_2(x;\nu^i)} \]

\[ D_2^i(x) = D(x) \frac{\mu_2(x;\nu^i)}{\mu_1(x;\theta^i) + \mu_2(x;\nu^i)} \]

Note: This only holds formally, since \( D_1^i(x) \) and \( D_2^i(x) \) are not integer-valued.
Estimating time-dependent parameters

- **Model:** \( D(t, x) \sim \text{Poisson}(E(t, x)[e^{-\sigma^2 \tilde{H}^{(t,x)}(t, x)} F(x; \theta_t) + G(x; \nu_t)]) \)

- **EM algorithm for fixed value of \( \sigma^2 \):**
  - Choose initial values \((\theta_0^0)\) and \((\nu_0^0)\)
  - Calculate adjusted integrated population hazard (on extended data window)
    \[
    \tilde{H}_i^{(t, x)}(t, x) = \sum_{u=0}^{x-1} \left\{ n(t - x + u, u) - G(u; \nu_{t-x+u}^i) \right\}
    \]
  - Obtain new estimates \((\theta_{i+1}^t)\) and \((\nu_{i+1}^t)\) by estimating the (formal) models
    \[
    D_{\text{base}}^i(t, x) \sim \text{Poisson}(E(t, x)e^{-\sigma^2 \tilde{H}_i^{(t, x)}(t, x)} F(x; \theta_t)) \quad D_{\text{base}}^i(t, x) = D(t, x) \frac{\tilde{\mu}_{\text{base}}^i(t, x)}{\tilde{\mu}_{\text{base}}^i(t, x) + \tilde{\mu}_{\text{acc}}^i(t, x)},
    \]
    \[
    D_{\text{acc}}^i(t, x) \sim \text{Poisson}(E(t, x)G(x; \nu_t)) \quad D_{\text{acc}}^i(t, x) = D(t, x) \frac{\tilde{\mu}_{\text{acc}}^i(t, x)}{\tilde{\mu}_{\text{base}}^i(t, x) + \tilde{\mu}_{\text{acc}}^i(t, x)},
    \]

  where \( \tilde{\mu}_{\text{base}}^i(t, x) = e^{-\sigma^2 \tilde{H}_i^{(t, x)}(t, x)} F(x; \theta_t^i) \) and \( \tilde{\mu}_{\text{acc}}^i(t, x) = G(x; \nu_t^i) \)

- Iterate steps 1 and 2 till convergence
The EM algorithm is reliable but slow
- guaranteed to converge to (local) maximum likelihood estimates
- depending on model complexity run times vary from seconds to hours!

A switching algorithm with a nested EM algorithm is very inefficient

Instead we optimize the (univariate) profile log-likelihood

\[ l(\sigma^2) = l(\sigma^2, \hat{\theta}_t(\sigma^2), \hat{\nu}_t(\sigma^2)) = \sum_x \{D(t,x) \log \bar{\mu}(t,x) - E(t,x)\bar{\mu}(t,x)\} \]

where \( \hat{\theta}_t(\sigma^2) \) and \( \hat{\nu}_t(\sigma^2) \) are the maximum likelihood estimates from the EM algorithm
and \( \bar{\mu}(t,x) = e^{-\sigma^2\bar{H}(t,x)} F(x; \hat{\theta}_t(\sigma^2)) + G(x; \hat{\nu}_t(\sigma^2)) \)

Numeric optimization may be achieved using only few evaluations of \( l(\sigma^2) \)
Summary

- Stochastic frailty model

\[ \mu(t, x) = E[Z|t,x] \mu_{base}(t, x) \]

- Estimation by switching algorithm:
  1. Estimate time-dependent parameters \( \theta_i \) for current value of \( \sigma^2 \)
  2. Estimate \( \sigma^2 \) for current value of \( \theta_i \) exploiting log-concavity

- Generalized stochastic frailty model

\[ \mu(t, x) = E[Z|t,x] \mu_{base}(t, x) + \mu_{acc}(t, x) \]

- Competing risks model
- Maximization of profile log-likelihood for \( \sigma^2 \)
- Profile log-likelihood calculated by EM-algorithm
Statistical analysis
Goodness of fit

- **Estimated model:** \( D(t, x) \sim \text{Poisson}(\hat{\mu}(t, x)E(t, x)) \)

- **Residuals**
  - **Pearson residual**
    \[
    r_p(t, x) = \frac{D(t, x) - \hat{\mu}(t, x)E(t, x)}{\sqrt{\hat{\mu}(t, x)E(t, x)}} = \frac{E(t, x)}{\sqrt{\hat{\mu}(t, x)}} \left( \frac{m(t, x) - \hat{\mu}(t, x)}{\hat{\mu}(t, x)} \right) \sim_{approx} N(0,1)
    \]
  - **Anscombe residual for Possion distribution**
    \[
    r_A(t, x) = 3 \sqrt{\frac{D(t, x)/\hat{\mu}(t, x)E(t, x)}{2}} - \left[ \frac{\hat{\mu}(t, x)E(t, x)}{2} \right]^{2/3} = \sqrt{\frac{E(t, x)}{\hat{\mu}(t, x)}} \left( \frac{m(t, x) - \hat{\mu}(t, x)}{\hat{\mu}(t, x)} \right) \sim_{approx} N(0,1)
    \]

- **Residuals are generally too large due to large exposures**

- **Relative error**
  \[
  \frac{m(t, x) - \hat{\mu}(t, x)}{\hat{\mu}(t, x)}
  \]

- **Summary measure of goodness of fit**
  \[
  \text{deviance} = -2[\log L(\hat{\mu}) - \log L(m)], \quad m(t, x) = D(t, x) / E(t, x)
  \]
Model selection

A good simple model is better than an excellent complex model

- Trade-off
  - Complexity: good fit to historical data
  - Simplicity: ease of forecasting

- Fit can always be improved by adding more parameters

- Information criteria, e.g. AIC or BIC, balance improved fit against complexity
  - but they tend to prefer more parameters to fewer parameters

- Our primary aim is forecasting
Model 1: Gamma-Makeham

\[ \mu(t, x) = \mathbb{E}[Z|t, x; \sigma^2] \exp(\alpha + \beta x + \gamma) \]

Model 2: Log-quadratic

\[ \mu(t, x) = \mathbb{E}[Z|t, x; \sigma^2] \exp(\alpha + \beta \gamma x + \eta x^2) + \gamma \]

Fitted to international male mortality for age 20-100 in the period 1950-2006

Deviance
- Model 1: 415097
- Model 2: 224348

Model 2 is clearly superior in terms of overall fit (deviance)
QQ-plot of Anscombe residuals

Normal QQ plot for Model 1

Normal QQ plot for Model 2
Time-dependent parameters of baseline hazard

Model 1: level ($\alpha$)

Model 2: level ($\alpha$)

Model 1: slope ($\beta$)

Model 2: slope ($\beta$)

Model 2: quadratic ($\eta/100$)

Model 1: Gamma-Makeham

$\mu_{base}(t, x) = \exp(\alpha + \beta_x)$

Model 2: Log-quadratic

$\mu_{base}(t, x) = \exp(\alpha_x + \beta_x + \eta_x x^2)$
Assume model has been chosen and estimated

\[ \hat{\mu}(t, x) = e^{-\sigma^2 \tilde{H}(t, x)} F(x; \hat{\theta}_t) + G(x; \hat{\nu}_t) \]

Next, choose and estimate time-series models for \((\theta_t)\) and \((\nu_t)\)

In many cases a random walk with drift is adequate

\[ \theta_t = \theta_{t-1} + \mu_\theta + \sigma_\theta \varepsilon_t \]
\[ \nu_t = \nu_{t-1} + \mu_\nu + \sigma_\nu \xi_t \]

where \((\varepsilon_t)\) and \((\xi_t)\) are independent (normal) random variates

In other cases more elaborate time-series models may be used
Rewrite intensity in terms of “observed” integrated individual intensity

\[ \hat{\mu}(t, x) = \left(1 + \hat{\sigma}^2 \tilde{I}(t, x) \right)^{-1} F(x; \hat{\theta}, \hat{\nu}) + G(x; \hat{\nu}), \quad \text{where} \quad \tilde{I}(t, x) = \left( e^{\hat{\sigma}^2 \tilde{H}(t, x)} - 1 \right)/\hat{\sigma}^2 \]

Forecasting from jump-off year \( T \)

\[ \mu(t, x) = \left(1 + \hat{\sigma}^2 \tilde{I}(t, x) \right)^{-1} F(x; \bar{\theta}, \bar{\nu}) + G(x; \bar{\nu}), \quad \text{for} \quad t > T \]

where \( \bar{\theta} \) and \( \bar{\nu} \) are either deterministic (e.g. mean) forecasts, or stochastic realizations from the time-series model, and \( \tilde{I}(t, x) \) is given by the recursion

For \( x = x_{\min} \):
\[ \tilde{I}(t, x_{\min}) = 0 \]

For \( x > x_{\min} \):
\[ \tilde{I}(t, x) = \tilde{I}(t-1, x-1) + F(x-1, \hat{\theta}_{t-1}) \]
Joint modelling of females and males
The gender gap

- Female life expectancy higher than male life expectancy
- Gender gap varies over time, but is believed to persist
- Separate analyses of females and males lead to diverging forecasts
- We need to model females and males jointly to ensure non-diverging forecasts

![Graph showing life expectancy difference between females and males over time](image-url)

- Green line: Life expectancy difference at age 0
- Orange line: Life expectancy difference at age 60
- Purple line: Life expectancy difference at age 100

Year:
- 1950
- 1960
- 1970
- 1980
- 1990
- 2000

Difference in years:
- 0
- 1
- 2
- 3
- 4
- 5
- 6
- 7

Life expectancy difference between females and males
Separate analyses

- Assume the same model has been estimated for females and males

\[ \hat{\mu}_f(t, x) = e^{-\hat{\sigma}_f^2 \hat{H}_f(t,x)} F(x; \hat{\Theta}_f(t)) + G(x; \hat{V}_f(t)) \]
\[ \hat{\mu}_m(t, x) = e^{-\hat{\sigma}_m^2 \hat{H}_m(t,x)} F(x; \hat{\Theta}_m(t)) + G(x; \hat{V}_m(t)) \]

- Assume time-dependent parameters are modelled by random walks with drift

\[ \theta_f^f = \theta_{f-1}^f + \hat{\mu}_f^f + \hat{\sigma}_f^f \epsilon_i^f \quad (\epsilon_i^f \text{ iid normal variates}) \]
\[ \theta_m^m = \theta_{m-1}^m + \hat{\mu}_m^m + \hat{\sigma}_m^m \epsilon_i^m \quad (\epsilon_i^m \text{ iid normal variates}) \]

- Difference is also a random walk with drift

\[ d_t = \hat{\theta}_f^f - \hat{\theta}_m^m \]
\[ d_t = d_{t-1} + (\hat{\mu}_f^f - \hat{\mu}_m^m) + \sigma_d u_t \quad (u_t \text{ iid normal variates}) \]

- Generally, diverging forecasts using standard estimates

\[ \hat{\mu}_f^f = \frac{1}{n} \sum_{t} \Delta \hat{\theta}_f^f \neq \frac{1}{n} \sum_{t} \Delta \hat{\theta}_m^m = \hat{\mu}_m^m \]
\[ \sigma_f^f = \sqrt{ \frac{ \sum (\Delta \hat{\theta}_f^f - \hat{\mu}_f^f)^2 }{n-1} } \neq \sqrt{ \frac{ \sum (\Delta \hat{\theta}_m^m - \hat{\mu}_m^m)^2 }{n-1} } = \sigma_m^m \]
\[ \sigma_d = \sqrt{ (\sigma_f^f)^2 + (\hat{\sigma}_m^m)^2 } \]

where \( \Delta \hat{\theta}_f^f = \hat{\theta}_f^f - \hat{\theta}_{f-1}^f \), \( \Delta \hat{\theta}_m^m = \hat{\theta}_m^m - \hat{\theta}_{m-1}^m \) and \( n \) is number of differences in data.
Converging *mean* forecasts can be achieved by setting

\[
\hat{\mu}_t^f = \hat{\mu}_t^m = \frac{1}{2n} \sum_i \left( \Delta \hat{\theta}_t^f + \Delta \hat{\theta}_t^m \right)
\]

but the difference is still a random walk (without drift)

\[
d_t = \hat{\theta}_t^f - \hat{\theta}_t^m
\]

\[
d_t = d_{t-1} + \sigma_d u_t
\]

- Mean and variance given value at jump-off year \( T \)
  - \( E[d_{T+h} | d_T] = d_T \)
  - \( \text{Var}[d_{T+h} | d_T] = h \sigma_d^2 \to \infty \quad \text{for} \quad h \to \infty \)
- Probability that females and males will stay close
  - \( P(\mid d_{T+h} - d_T \mid \leq K) \to 0 \quad \text{for} \quad h \to \infty \)

for any \( K \)!
Joint modelling objectives

- Female and male parameters should evolve similar to random walks
- The difference should possess randomness but stay bounded (in probability)

Formally, we want female and male parameters to *cointegrate*

Definitions

- A process $X_t$ is called *stationary* if there exists a distribution $\pi$ such that $X_t \xrightarrow{D} \pi$

- A process $X_t$ is called *first-order integrated* if $\Delta(X_t - E(X_t))$ is stationary, e.g. $X_t$ is a random walk (with drift) or similar to a random walk

- A multivariate, first-order integrated process $X_t$ is called *cointegrated* with cointegration vector $\beta \neq 0$ if $\beta'X_t$ is stationary
Error correction model

- Assume (for the moment) that $\theta_t^f$ and $\theta_t^m$ are univariate
- Form the bivariate process
  \[ X_t = \begin{pmatrix} \theta_t^f \\ \theta_t^m \end{pmatrix} \]
- **Error correction** model
  \[ \Delta X_t = \alpha \beta' X_{t-1} + \mu + \epsilon_t \]
  where $\alpha = (\alpha_1 \alpha_2)'$, $\beta = (\beta_1 \beta_2)'$, $\mu = (\mu_1 \mu_2)'$, and $\epsilon_t \sim N_2(0, \Omega)$
- Interpretation
  - Consider $\beta' X_t = E(\beta' X_t) = c$ as defining the relation between males and females
  - $X_t$ is updated in response to the disequilibrium error $\beta' X_t - c$ through the adjustment $\alpha$
    \[ \beta' X_t = \beta' X_{t-1} + \beta' \alpha (\beta' X_{t-1} - c) + \beta' \epsilon_t \]
We are interested in the model with $\beta = (1 -1)'$

$$\Delta X_i = \alpha d_{i-1} + \mu + \epsilon_i$$

where $d_i = \theta_i^f - \theta_i^m$

Difference process

$$d_i = (1 + \alpha_1 - \alpha_2) d_{i-1} + (\mu_1 - \mu_2) + (\epsilon_i^1 - \epsilon_i^2)$$

In stationarity (requires $|1+\alpha_1 - \alpha_2| < 1$), i.e. when females and males cointegrate

$$E[d_i] = \frac{\mu_1 - \mu_2}{\alpha_2 - \alpha_1}$$

$$E[\Delta X_i] = E\left(\begin{pmatrix} \theta_i^f - \theta_{i-1}^f \\ \theta_i^m - \theta_{i-1}^m \end{pmatrix}\right) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \frac{\alpha_2 \mu_1 - \alpha_1 \mu_2}{\alpha_2 - \alpha_1}$$
Rewrite model in standard regression form

\[ Y_t = B'Z_t + \epsilon_t \]

where \( Y_t = \Delta X_t \), \( B' = \begin{pmatrix} \alpha_1 & \mu_1 \\ \alpha_2 & \mu_2 \end{pmatrix} \), \( Z_t = \begin{pmatrix} d_{t-1} \\ 1 \end{pmatrix} \), and \( \epsilon_t \sim N_2(0, \Omega) \)

Regression estimators (maximum likelihood estimators)

\[
\hat{B} = \left( \sum_i Z_i Z_i' \right)^{-1} \left( \sum_i Z_i X_i' \right) = S_{xz}^{-1} S_{xx} \\
\hat{\Omega} = \frac{1}{n} \sum_t (X_t - \hat{B}'Z_t)(X_t - \hat{B}'Z_t)' = S_{xx} - S_{xz}S_{xz}^{-1} S_{zx}
\]

where for any two processes \( X_t \) and \( Z_t \) we use the notation

\[
S_{zx} = \frac{1}{n} \sum_{t=1}^n Z_t X_t'
\]
Estimated, stationary mean difference and mean drift will not necessarily equal empirical values.

Desired mean values of

\[ E[d_t] = \frac{\mu_1 - \mu_2}{\alpha_2 - \alpha_1} = d_{target} \]

\[ E[\theta_t^f - \theta_{t-1}^f] = E[\theta_t^m - \theta_{t-1}^m] = \frac{\alpha_2 \mu_1 - \alpha_1 \mu_2}{\alpha_2 - \alpha_1} = \Delta_{target} \]

can be achieved by adjusting drift parameters (and retaining \( \alpha \))

\[ \mu_1 = \Delta_{target} - \alpha_1 d_{target} \]

\[ \mu_2 = \Delta_{target} - \alpha_2 d_{target} \]
Application to international data
- International data from Human Mortality Database
  - 18 countries
  - Data exists for age 0-110
- Data window used for estimation
  - Time period: 1950-2006
  - Age span: 20-100
- Model selection
  - Find common model for females and males
- Separate and joint modelling of females and males
  - Random walk and error correction model
- Mean forecast and stochastic forecast
  - Cohort and period life expectancy
- Analysis performed in R
Model selection

- Generalized stochastic frailty model

\[
\mu(t, x) = \mathbb{E}[Z|t, x] \mu_{\text{base}}(t, x) + \mu_{\text{acc}}(t, x) = e^{-\sigma^2 \tilde{\rho}(t, x)} F(x; \theta_t) + G(x; \nu_t)
\]

- Candidate models

<table>
<thead>
<tr>
<th></th>
<th>( F(x; \theta_t) )</th>
<th>( G(x; \nu_t) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>( - )</td>
<td>( \exp(\alpha_i + \beta_i x) )</td>
</tr>
<tr>
<td>Model 2</td>
<td>( - )</td>
<td>( \exp(\alpha_i + \beta_i x + \gamma_t) )</td>
</tr>
<tr>
<td>Model 3</td>
<td>( - )</td>
<td>( \exp(\alpha_i + \beta_i x + \eta_i x^2) + \gamma_t )</td>
</tr>
<tr>
<td>Model 4</td>
<td>( - )</td>
<td>( \frac{\kappa_i \exp(\beta_i x)}{1 + \phi_i \exp(\beta_i x)} + \gamma_t )</td>
</tr>
<tr>
<td>Model 5</td>
<td>( \exp(\alpha_i + \beta_i x) )</td>
<td>( \gamma_t )</td>
</tr>
<tr>
<td>Model 6</td>
<td>( \exp(\alpha_i + \beta_i x + \eta_i x^2) )</td>
<td>( \gamma_t )</td>
</tr>
</tbody>
</table>
Frailty models are generally better

Model 6 is by far the best overall
  - but we know from the example that its parameters are hard to forecast

Consider model “in between” Model 5 and Model 6 with $\eta_i = \eta$

<table>
<thead>
<tr>
<th>Model</th>
<th>$F(x; \theta_i)$</th>
<th>$G(x; \nu_i)$</th>
<th>Females</th>
<th>Males</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>$-$</td>
<td>$\exp(\alpha_i + \beta_i x)$</td>
<td>1765791</td>
<td>1731634</td>
</tr>
<tr>
<td>Model 2</td>
<td>$-$</td>
<td>$\exp(\alpha_i + \beta_i x) + \gamma_i$</td>
<td>534533</td>
<td>572201</td>
</tr>
<tr>
<td>Model 3</td>
<td>$-$</td>
<td>$\exp(\alpha_i + \beta_i x + \eta_i x^2) + \gamma_i$</td>
<td>405989</td>
<td>224315</td>
</tr>
<tr>
<td>Model 4</td>
<td>$-$</td>
<td>$\frac{\kappa_i \exp(\beta_i x)}{1 + \varphi_i \exp(\beta_i x)} + \gamma_i$</td>
<td>512072</td>
<td>334767</td>
</tr>
<tr>
<td>Model 5</td>
<td>$\exp(\alpha_i + \beta_i x)$</td>
<td>$\gamma_i$</td>
<td>534158</td>
<td>415097</td>
</tr>
<tr>
<td>Model 6</td>
<td>$\exp(\alpha_i + \beta_i x + \eta_i x^2)$</td>
<td>$\gamma_i$</td>
<td>217076</td>
<td>224348</td>
</tr>
</tbody>
</table>
Females and males do not agree on the best model

- Females prefer constant value of $\eta \sim 6 \cdot 10^{-4}$
- Males prefer constant value of $\eta \sim 0$ (or negative)

- Value with lowest total deviance, $\eta = 2 \cdot 10^{-4}$, is chosen as common value

<table>
<thead>
<tr>
<th>Model</th>
<th>$F(x; \theta_i)$</th>
<th>$G(x; \nu_i)$</th>
<th>Females</th>
<th>Males</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 5</td>
<td>$\exp(\alpha_i + \beta_i x)$</td>
<td>$\gamma_i$</td>
<td>534158</td>
<td>415097</td>
<td>949255</td>
</tr>
<tr>
<td></td>
<td>$\exp(\alpha_i + \beta_i x + 1x^210^{-4})$</td>
<td>$\gamma_i$</td>
<td>463798</td>
<td>461744</td>
<td>925541</td>
</tr>
<tr>
<td></td>
<td>$\exp(\alpha_i + \beta_i x + 2x^210^{-4})$</td>
<td>$\gamma_i$</td>
<td>403532</td>
<td>514086</td>
<td>917618</td>
</tr>
<tr>
<td></td>
<td>$\exp(\alpha_i + \beta_i x + 3x^210^{-4})$</td>
<td>$\gamma_i$</td>
<td>353287</td>
<td>570513</td>
<td>923801</td>
</tr>
<tr>
<td></td>
<td>$\exp(\alpha_i + \beta_i x + 6x^210^{-4})$</td>
<td>$\gamma_i$</td>
<td>268205</td>
<td>781940</td>
<td>1050145</td>
</tr>
<tr>
<td>Model 6</td>
<td>$\exp(\alpha_i + \beta_i x + \eta_i x^2)$</td>
<td>$\gamma_i$</td>
<td>217076</td>
<td>224348</td>
<td>441424</td>
</tr>
</tbody>
</table>
- Model fitted to ages 20 to 100

- Complex structure of mortality below age 20
  - high infant mortality
  - decreasing to age 10
  - increasing after age 10
  - hump around age 20

- At very old ages fit is too high for females and too low for males
  - comprise/common model

---

### Data and fit in 2006

![Graph showing data and fit in 2006](image)
- Common model for females and males

\[ \mu(t, x) = e^{-\sigma^2 \tilde{H}(t, x)} \exp(\alpha_t + \beta_t x + 2x^2 10^{-4}) + \gamma_t \]

- Estimated time-dependent parameters

![Level (\(\alpha\))] ![Slope (\(\beta\))] ![Accident rate (\(\gamma\))]
Separate and joint modelling

- Separate modelling by random walks (and no further improvements in accident rates)

\[
\alpha^f_t = \alpha^f_{t-1} - 0.02662 + 0.04588 \epsilon^f_t \\
\beta^f_t = \beta^f_{t-1} + 0.0001548 + 0.0008109 u^f_t \\
\gamma^f_t = \gamma^f_{2006} = 0.03440\%
\]

\[
\alpha^m_t = \alpha^m_{t-1} - 0.02345 + 0.04015 \epsilon^m_t \\
\beta^m_t = \beta^m_{t-1} + 0.0001411 + 0.0006624 u^m_t \\
\gamma^m_t = \gamma^m_{2006} = 0.06882\%
\]

- Joint modelling by error correction model adjusted to empirical difference and drift

\[
\Delta \left( \begin{array}{c}
\alpha^f_t \\
\alpha^m_t
\end{array} \right) = \left( \begin{array}{c}
-0.1057 \\
-0.0436
\end{array} \right) (\alpha^f_{t-1} - \alpha^m_{t-1}) + \left( \begin{array}{c}
-0.1273 \\
-0.0672
\end{array} \right) + \epsilon^\alpha_t \\
E(\Delta \alpha^f) = E(\Delta \alpha^m) = -0.9674 \\
E(\alpha^f - \alpha^m) = -0.02504
\]

\[
\Delta \left( \begin{array}{c}
\beta^f_t \\
\beta^m_t
\end{array} \right) = \left( \begin{array}{c}
-0.1161 \\
-0.0858
\end{array} \right) (\beta^f_{t-1} - \beta^m_{t-1}) + \left( \begin{array}{c}
0.0007682 \\
0.0006062
\end{array} \right) + \epsilon^\beta_t \\
E(\Delta \beta^f) = E(\Delta \beta^m) = 0.005340 \\
E(\beta^f - \beta^m) = 0.0001480
\]

and no further improvements in accident rates
Mean forecast from jump-off year $T$

- Separate modelling by random walks (accident rates kept at jump-off year value)

$$E(\alpha^f_t) = \alpha^f_T - 0.02662(t - T)$$

$$E(\alpha^m_t) = \alpha^m_T - 0.02345(t - T)$$

$$E(\beta^f_t) = \beta^f_T + 0.0001548(t - T)$$

$$E(\beta^m_t) = \beta^m_T + 0.0001411(t - T)$$

- Joint modelling by error correction model (accident rates kept at jump-off year value)

Mean forecasts are given by the recursion ($t > T$):

$$\begin{pmatrix} E(\alpha^f_t) \\ E(\alpha^m_t) \end{pmatrix} = C_\alpha \begin{pmatrix} E(\alpha^f_{t-1}) \\ E(\alpha^m_{t-1}) \end{pmatrix} + \begin{pmatrix} -0.1273 \\ -0.0672 \end{pmatrix}, \quad C_\alpha = \begin{pmatrix} 1 - 0.1057 & 0.1057 \\ -0.0436 & 1 + 0.0436 \end{pmatrix}$$

$$\begin{pmatrix} E(\beta^f_t) \\ E(\beta^m_t) \end{pmatrix} = C_\beta \begin{pmatrix} E(\beta^f_{t-1}) \\ E(\beta^m_{t-1}) \end{pmatrix} + \begin{pmatrix} 0.0007682 \\ 0.0006062 \end{pmatrix}, \quad C_\beta = \begin{pmatrix} 1 - 0.1161 & 0.1161 \\ -0.0858 & 1 + 0.0858 \end{pmatrix}$$
Forecasted time-dependent parameters

Estimated and forecasted alpha

Estimated and forecasted beta
Same fit but (slightly) different forecasts

**International female mortality**

**International male mortality**
## Period and cohort life expectancy

### Period life expectancy (without future improvements in mortality)

<table>
<thead>
<tr>
<th></th>
<th>2006 Data</th>
<th>RW</th>
<th>EC</th>
<th>2030 RW</th>
<th>EC</th>
<th>2050 RW</th>
<th>EC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>82.1</td>
<td>82.1</td>
<td>82.1</td>
<td>85.2</td>
<td>85.3</td>
<td>87.5</td>
<td>87.6</td>
</tr>
<tr>
<td>Male</td>
<td>76.7</td>
<td>76.5</td>
<td>76.5</td>
<td>79.3</td>
<td>79.7</td>
<td>81.6</td>
<td>82.1</td>
</tr>
<tr>
<td>Difference</td>
<td>5.4</td>
<td>5.6</td>
<td>5.6</td>
<td>5.8</td>
<td>5.6</td>
<td>6.0</td>
<td>5.5</td>
</tr>
</tbody>
</table>

### Cohort life expectancy (with future improvements in mortality)

<table>
<thead>
<tr>
<th></th>
<th>2006 Data</th>
<th>RW</th>
<th>EC</th>
<th>2030 RW</th>
<th>EC</th>
<th>2050 RW</th>
<th>EC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>-</td>
<td>91.5</td>
<td>91.3</td>
<td>94.2</td>
<td>93.8</td>
<td>96.3</td>
<td>95.7</td>
</tr>
<tr>
<td>Male</td>
<td>-</td>
<td>84.6</td>
<td>85.5</td>
<td>87.2</td>
<td>88.3</td>
<td>89.2</td>
<td>90.4</td>
</tr>
<tr>
<td>Difference</td>
<td>-</td>
<td>6.9</td>
<td>5.8</td>
<td>7.0</td>
<td>5.5</td>
<td>7.1</td>
<td>5.3</td>
</tr>
</tbody>
</table>
Different stochastic structure of RW and EC model...
...leads to large differences in gender gap distribution.

Separate modelling (RW)

Joint modelling (EC)


