Chapter 13
Analysis of the Stochasticity of Mortality Using Variance Decomposition

Erland Ekheden and Ola Hössjer

Abstract We analyse the stochasticity in mortality data from the USA, the UK and Sweden, and in particular to which extent mortality rates are explained by systematic variation, due to various risk factors, and random noise. We formalise this in terms of a mixed regression model with a logistic link function, and decompose the variance of the observations into three parts: binomial risk, the variance due to random mortality variation in a finite population, systematic risk explained by the covariates and unexplained systematic risk, variance that comes from real changes in mortality rates, not captured by the covariates. The fraction of unexplained variance caused by binomial risk provides a limit in terms of the resolution that can be achieved by a model. This can be used as a model selection tool for selecting the number of covariates and regression parameters of the deterministic part of the regression function, and for testing whether unexplained systematic variation should be explicitly modelled or not. We use a two-factor model with age and calendar year as covariates, and perform the variance decomposition for a simple model with a linear time trend on the logit scale. The population size turns out to be crucial, and for Swedish data, the simple model works surprisingly well, leaving only a small fraction of unexplained systematic risk, whereas for the UK and the USA, the amount of unexplained systematic risk is larger, so that more elaborate models might work better.

13.1 Introduction

Decreasing mortality rates is not a new phenomenon. This trend has been evident for over a century in countries like Sweden, the United Kingdom and the United States. "Longevity" is an often used term for this trend, especially when the trend is

E. Ekheden (✉) · O. Hössjer
Stockholm University, Stockholm, Sweden
e-mail: erland@math.su.se

O. Hössjer
e-mail: ola@math.su.se

D. Silvestrov and A. Martin-Löf (eds.), *Modern Problems in Insurance Mathematics*, EAA Series, DOI: 10.1007/978-3-319-06653-0_13,
© Springer International Publishing Switzerland 2014

viewed as an economic risk for society, pension funds and insurers. Actuaries and demographers have a long tradition of making life tables and models for mortality. Thirty years ago Osmond [31] introduced the Age-Period-Cohort model within the medical statistics literature, but the interest in stochastic modelling of mortality first took off with the paper by Lee and Carter [27] in which a principal components approach of Bozik and Bell [6] and Bell and Monsell [4] was modified. Since then a variety of models have been proposed, see [2, 5, 11, 12, 14] for recent overviews with further references. They differ in the way in which the covariates; age x, calendar year t and cohort $t-x$, are included, and whether the one year death risk, q_{tx}, or the closely related expected number of deaths per individual and unit of time, the death intensity $\mu_{tx} \approx -\log(1-q_{tx})$, is modelled.

The richness of proposed models shows that the problem is non-trivial, with a high dimensional data set. There are more than a hundred observed age specific mortalities, often for males and females, collected for over thirty, fifty and even a hundred calendar years. Still there are substantial correlations in data, since mortality in general increases with age. The improvements of mortalities seem, however, to be non-stationary, in that the rates vary over ages and time. On top of this we have random noise, caused by individual variation in a finite population.

When evaluating models, some seem to be too simple. This may either be assessed in an explorative data analysis which may reveal marked patterns in residual plots that signify features of historical data not explained by the model, or formally by some model selection criterion such as BIC [12]. Other models seem to be too complex. Even though they fit historical data well, they might be sensitive to varying indata and have less robust forecasts, see [12, 14]. Bell [3] showed that a simple model, where the logged death rates constitute a random walk with drift, separately for each age, can sometimes outperform more complex models in terms of forecasting. Bell's work has received relatively little attention in the literature, and it seems there is still work to be done in terms of selecting models for mortality and forecasting.

In the Lee–Carter model and many of its successors, it is often taken for granted that either the *observed* death rates $\hat{\mu}_{xt}$ or mortality rates $\hat{q}_{xt}$ are stochastic processes. It is, however, seldom explicitly pointed out that what we observe is a finite population and that the randomness is partially caused by this. Brouhns et al. [10] used Poisson regression, where instead, the *actual* death rates are non-random, whereas the randomness from the finite population manifests itself in terms of a Poisson distributed number of deaths (see also [1, 35]). This source of variation has been referred to as Poisson risk [15], and analogously, we speak of binomial risk if the number of deaths is assumed be binomially distributed. Both Poisson and binomial risks are examples of diversifiable risks.

In this paper we will take a closer look at the randomness of observed mortality rates $\hat{q}_{xt}$. The aim is to get a better understanding of the underlying processes and to get new means for model selection and/or model validation. This close look will start with an explorative data analysis, where some stochastic behaviour of data in a finite population is expected regardless of the stochastic nature of the underlying mortality rates q_{xt}.

As a next more formal step we divide variation in observed mortalities into three components by means of a certain variance decomposition for mixed regression models [24] that has previously been applied to non-life insurance data [25]. The first two components is systematic risk (variation in true mortalities) that is either explained or unexplained by the covariates age and year, and the third component, binomial risk, is due to the finiteness of the population. The novel feature, in the life insurance context, is that unexplained variation can be divided into binomial risk and unexplained systematic risk. We develop a test where the size of these two risk components are compared, and show that the unexplained systematic risk can/should be excluded from the model for a small population, or at low ages. In that case a simple logistic regression analysis can be employed. This test can also be interpreted as a test of over-dispersion of the annual death counts compared to what would be expected for a binomial distribution with non-random mortality rates.

In our analysis, we will use data sets for Swedish, UK, and United States populations. Rather than finding a multi-population model that fits all three data sets [13, 29], we build a single model separately for each country. There is a danger of using a single data set, since it may contain something specific that one takes to be general. However, the three chosen countries have a broad range of population sizes and are popular in the literature for their economic importance and size (UK and USA) or admittedly good data quality (Sweden), and therefore constitute a fairly broad range of Western populations. We use the latest available data at the time of download, ranging from 1979 to 2011 (Sweden), 2009 (UK) and 2007 (USA) respectively, with males and females handled separately. The data comes from the Human Mortality database, see mortality.org for further documentation.

13.1.1 Preliminaries

We study a population of ages $x = x_l, \ldots, x_u$ spanning between lower and upper limits x_l and x_u, during calendar years $t = t_1, \ldots, t_T$, where t_T is the latest year of observations and T is the length of the time window. Assuming that N_{xt} is the number of individuals of age x alive at the beginning of calendar year t (or more generally the exposure-to-risk E_{xt}, rounded to the nearest integer), the number of deaths

$$D_{xt}|q_{xt} \sim \text{Bin}(N_{xt}, q_{xt})$$

among them within one year is assumed to have a binomial distribution, with a death probability or mortality rate q_{xt} that can be estimated as

$$\hat{q}_{xt} = \frac{D_{xt}}{N_{xt}}. \tag{13.1}$$

As mentioned in the introduction, a Poisson approximation

$$D_{xt}|\mu_{xt} \sim \text{Po}(E_{xt}\mu_{xt})$$

to death counts is often employed, see for instance [9, 10]. This is a useful approximation for most ages, but for higher ages, over 80, the Poisson distribution increasingly overestimates the variance, making it less suitable for our purposes.

We will work with logit transformed (LM) mortality data

$$Y_{xt}^{\text{LM}} = \text{logit}\hat{q}_{xt} = \log \frac{\hat{q}_{xt}}{1-\hat{q}_{xt}}, \tag{13.2}$$

and in order to remove linear age trends, we also study the logit transformed increments (LMI)

$$Y_{xt}^{\text{LMI}} = \Delta\text{logit}\hat{q}_{xt} = \text{logit}\hat{q}_{xt} - \text{logit}\hat{q}_{x,t-1} \tag{13.3}$$

in time, regarding data as a time series for every fixed age x.

13.2 Explorative Data Analysis

We wish to get a better understanding of the probabilistic properties of mortality data. We perform an explorative data analysis in order to achieve this. We know that there are year to year variations in mortalities. First we ask if the logit transformed increments (13.3) are normally distributed. We inspect quantile-quantile plots, some of which are shown in Fig. 13.1, and find the normal distribution to be a reasonable assumption.

We then proceed to investigate the correlation structure of the LMI data. In order to get a picture of how the correlation between nearby ages varies with age, we study in Fig. 13.2, for each age x, the average

$$\frac{1}{4}\sum_{\substack{|h|\le 2\\ h\neq 0}} \widehat{\text{Corr}}(Y_{x\cdot}^{\text{LMI}}, Y_{x+h,\cdot}^{\text{LMI}}) \tag{13.4}$$

of the estimated correlation function for the four nearest ages. For Sweden and the UK, the correlation is around zero for low ages, but it starts to rise at the age of 60, so that a marked correlation of 0.5 can be seen at the age of 90. The pattern for US males is quite different, here we have a pronounced correlation of 0.7 in ages 20–40, then it dips towards zero at age 60 and finally it rises again to 0.7 at age 90. For US females, the correlation is not so high in ages 20–40 as for the males, but for ages 60 and higher, they are very close.

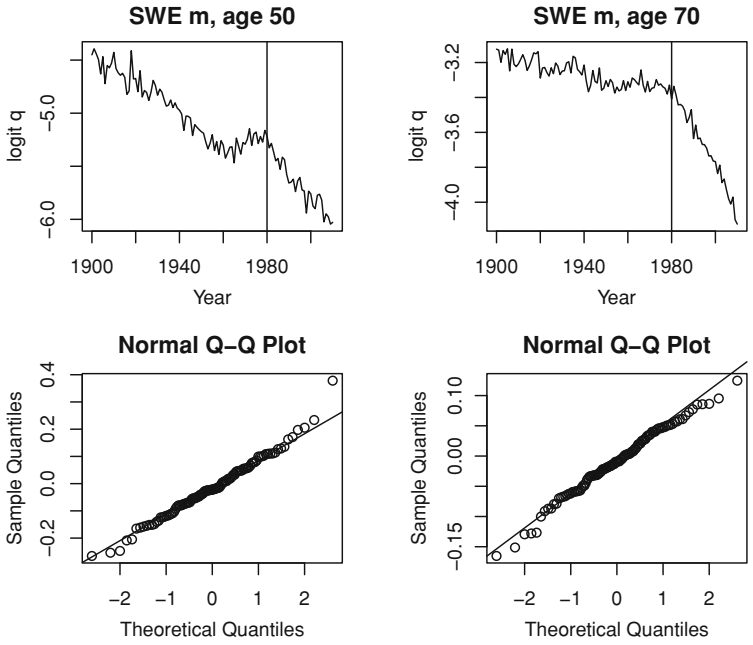


Fig. 13.1 Estimated logit mortality rates (13.2) and QQ-plots of increments (13.3) of estimated logit mortality rates for the Swedish data set, males of ages 50 and 70

We then look at the estimated lag 1 autocorrelation function

$$\widehat{\mathrm{ACF}}_x(1) = \widehat{\mathrm{Corr}}(Y^{\mathrm{LMI}}_{x\cdot}, Y^{\mathrm{LMI}}_{x,\cdot+1})$$

of the LMI data, and in Table 13.1 we have averaged these over all age classes for the Swedish, UK and US datasets. At this stage we define an MA(1)-process (without drift)

$$Y_t = \beta + w_t + \theta w_{t-1}, \tag{13.5}$$

with $w_t \sim N(0, \sigma^2)$ independent. Recall that it has an ACF(1) equal to $\theta/(1+\theta^2)$, see for instance [8], hence

$$\mathrm{ACF}(1) = -0.5 \Rightarrow \theta = -1.$$

When $\theta = -1$ we can rewrite (13.5) as

$$Y_t = (w_t + \alpha + \beta(t - \bar{t}) - (w_{t-1} + \alpha + \beta(t - 1 - \bar{t}))$$

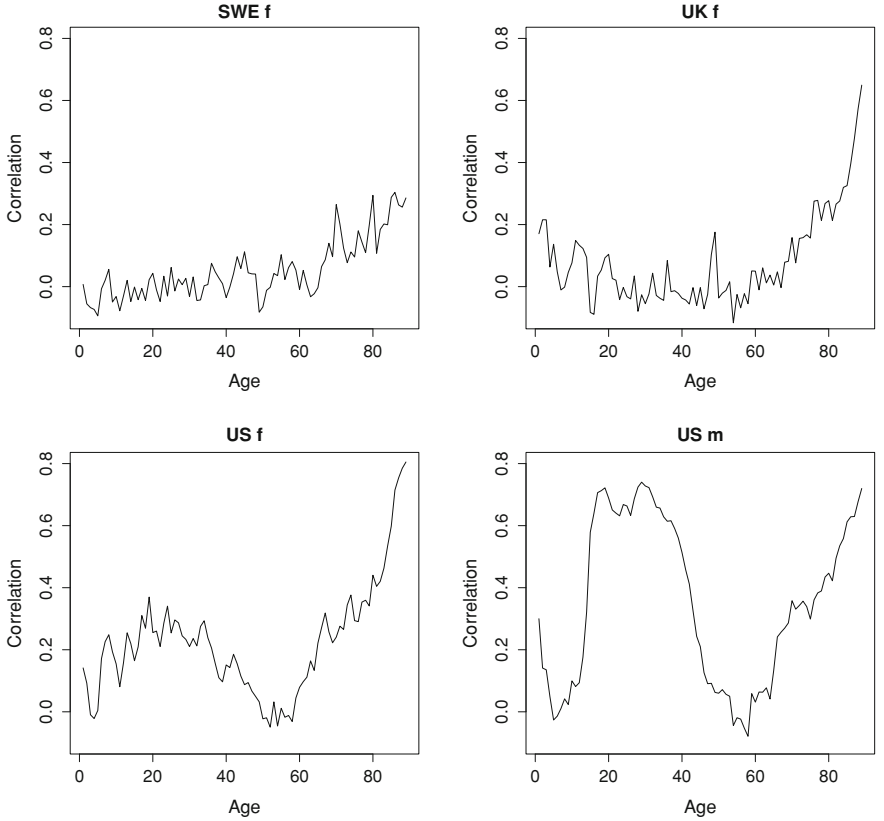


Fig. 13.2 Correlation of estimated increments (13.3) of logit mortality rates between each age group x and its four nearest age groups for the Swedish, UK and US datasets of males and females

and interpret it as LMI transformed data (13.3), in which $\alpha+\beta(t-\bar{t})$ and w_t represent a linear trend and independent noise of the estimated logit mortalities (13.2), with $\bar{t} = (t_1 + t_T)/2$ the midpoint of the observational interval of calendar years. An ACF(1) close to -0.5 thus indicates a high degree of independence of the logit mortalities between years, when a linear trend has been removed. The findings for Sweden and the UK, with low correlations between nearby ages and an ACF(1) close to -0.5, see Table 13.1, suggest a large amount of independence between years *and* over ages. We will find in Sect. 13.3 that this is caused by a binomial risk that is large in comparison to the unexplained systematic risk. The autocorrelations for US data suggests a structure with more dependencies, corresponding to a lower fraction of binomial risk.

Table 13.1 Autocorrelations $\widehat{\text{ACF}}(1)$ = average $\widehat{\text{ACF}}_x(1)$ of lag 1, for increments of logit transformed data, averaged over all age classes, for the Swedish, UK and US datasets

Category	$\widehat{\text{ACF}}(1)$		
SWE males	−0.45	SWE females	−0.46
UK males	−0.42	UK females	−0.44
US males	−0.15	US females	−0.29

13.3 Mixed Regression Model for Transformed Data

We will formalise the procedure of the previous section, and notice that (13.2) and (13.3) are both transformations

$$Y_{xt} = f_{xt}(\hat{\mathbf{q}}), \quad (x,t) \in \Omega \tag{13.6}$$

of the estimated mortalities $\hat{\mathbf{q}} = \{\hat{q}_{xt};\ (x,t) \in \Omega\}$, computed for a collection

$$\Omega \subset \{(x,t);\ x_l \le x \le x_u, t_1 \le t \le t_T\}$$

of ages and calendar years. The analogous transformations of the true but unknown mortalities $\mathbf{q} = \{q_{xt};\ (x,t) \in \Omega\}$ are denoted as

$$Y_{xt}^{\infty} = f_{xt}(\mathbf{q}), \quad (x,t) \in \Omega,$$

where the superscript ∞ signifies a hypothetical population of infinite size with no binomial risk.

Imagine that we have a regression model with response variables $\{Y_{xt};\ (x,t) \in \Omega\}$, covariates (x,t) and parameters θ. In order to assess how much of the variation in Y_{xt} is a function of changes in the underlying $\mathbf{q}$, not explained by our model (systematic variation), and how much is due to random noise (binomial risk), we write

$$\begin{aligned} Y_{xt} &= m_{xt} + \varepsilon_{xt} \\ &= m_{xt} + \varepsilon_{xt}^{s} + \varepsilon_{xt}^{b}, \end{aligned} \tag{13.7}$$

as a sum of one part

$$m_{xt} = m_{xt}(\theta) = E_{\theta}(Y_{xt}^{\infty}) \tag{13.8}$$

explained by the regression model, and another part ε_{xt} not explained by the regression model. The explained part depends on a number of regression parameters $\theta = (\theta_1, \ldots, \theta_p)^T$, the unexplained part can further be decomposed into a sum of $\varepsilon_{xt}^{s} = Y_{xt}^{\infty} - m_{xt}$, the unexplained systematic variation, which by definition satisfies $E(\varepsilon_{xt}^{s}) = 0$, and $\varepsilon_{xt}^{b} = Y_{xt} - Y_{xt}^{\infty}$, the unexplained random noise, i.e. binomial risk. We assume that

$$E(\varepsilon_{xt}^b) = 0, \tag{13.9}$$

which is accurate to order N_{xt}^{-1} for smooth transformations f_{xt}.

Based on (13.7), we decompose the variance

$$\begin{aligned} v_{xt} &= \mathrm{Var}(Y_{xt}) \\ &= \mathrm{Var}(\varepsilon_{xt}^s) + \mathrm{Var}(\varepsilon_{xt}^b) \\ &=: v_{xt}^s + v_{xt}^b \end{aligned} \tag{13.10}$$

of Y_{xt} into two parts, of which v^s represents unexplained systematic variation and v^b binomial risk.

In particular we will study linear mixed regression models

$$\mathbf{Y} = \mathbf{X}\theta + \varepsilon, \tag{13.11}$$

where $\mathbf{Y} = (Y_{xt};\ (x,t) \in \Omega)^T$ and $\varepsilon = (\varepsilon_{xt};\ (x,t) \in \Omega)^T$ are $n \times 1$ column vectors, $\mathbf{X}$ is an $n \times p$ design matrix and n is the number of elements in Ω. The least squares estimator

$$\hat{\theta} = (\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}^T\mathbf{Y} \tag{13.12}$$

will be used in a model selection step below for computing estimates $\hat{m}_{xt} = m_{xt}(\hat{\theta})$ of the regression function.

13.4 Basic Model

We assume a simple two-factor model

$$\mathrm{logit} q_{xt} = \alpha_x + \beta_x(t - \bar{t}) + \delta_{xt} \tag{13.13}$$

for the logit transformed mortalities, with age and calendar years as covariates, whereas cohort effects $t - x$ are not included. The deterministic period effect $(t - \bar{t})$ is linear, with $\bar{t} = (t_1 + t_T)/2$ the mid-point of the chosen time interval. This often provides a good approximation, see for instance Sect. 9 of [17]. The intercepts α_x and slopes β_x represent deterministic age effects, for which we assume a parametrisation

$$\begin{aligned} \alpha_x &= \sum_{j=0}^{p_1} a_j \phi_j(x), \\ \beta_x &= \sum_{j=0}^{p_2} b_j \phi_j(x) \end{aligned}$$

in terms of basis functions ϕ_j that are either polynomials,

$$\phi_j(x) = x^j, \tag{13.14}$$

or single age class indicators

$$\phi_j(x) = 1_{\{x=x_{j+l}\}}, \tag{13.15}$$

so that each age class is assigned a separate intercept and slope parameter, corresponding to $p_1 = p_2 = x_u - x_l$, $a_j = \alpha_{x_l+j}$ and $b_j = \beta_{x_l+j}$.

While (13.14) has the advantage of smoothing the logit transformed mortalities age-wise, (13.15) is better at capturing age specific effects. It is also possible to choose the basis functions as B-splines [19, 26].

The δ_{xt} terms are random variables with $E(\delta_{xt}) = 0$. If these are all independent, we get a generalised (or hierarchical) linear mixed model, for which various approximate estimation algorithms are available, see for instance [7, 28].

In the following two subsections, we analyse two transformations (13.2) and (13.3) of raw data in more detail for the model in (13.13).

13.4.1 Logit Mortality

Assume that $Y_{xt} = Y_{xt}^{\mathrm{LM}}$ in (13.2) is defined for all (x, t) in $\Omega = \{(x, t);\ x_l \le x \le x_u, t_1 \le t \le t_T\}$. The three terms in (13.7), and the parameter vector, then have the form

$$\begin{aligned} m_{xt} &= \sum_{j=0}^{p_1} a_j \phi_j(x) + (t - \bar{t}) \sum_{j=0}^{p_2} b_j \phi_j(x), \\ \varepsilon_{xt}^s &= \delta_{xt}, \\ \varepsilon_{xt}^b &= \mathrm{logit}\hat{q}_{xt} - \mathrm{logit}q_{xt}, \\ p &= p_1 + p_2 + 2, \\ \theta &= (a_0, \ldots, a_{p_1}, b_0, \ldots, b_{p_2})^T. \end{aligned} \tag{13.16}$$

In the Lee–Carter model and many of its extensions, age and period parameters enter bi-linearly into the regression function. However, since time enters as a fixed known covariate in terms of a linear time trend in (13.13), Eq. (13.7) can be rewritten as a multiple linear regression model (13.11), where the design matrix $\mathbf{X}$ has rows

$$(\phi_0(x), \phi_1(x), \ldots, \phi_{p_1}(x), (t - \bar{t})\phi_0(x), (t - \bar{t})\phi_1(x), \ldots, (t - \bar{t})\phi_{p_2}(x))$$

for all $(x, t) \in \Omega$.

It follows from (13.9) that the binomial risk variance function satisfies

$$v_{xt}^b = E\left(\mathrm{Var}(\mathrm{logit}\hat{q}_{xt}|q_{x,t})\right) \approx E\left(\frac{1}{N_{xt}q_{xt}(1 - q_{xt})}\right), \tag{13.17}$$

where the variance of a transformed binomial variable is computed by means of a Gauss approximation in the last step. Hence we can estimate v_{xt}^b from the data as

$$\hat{v}_{xt}^b = \frac{1}{N_{xt}\bar{q}_{xt}(1-\bar{q}_{xt})}, \tag{13.18}$$

where either

$$\bar{q}_{xt} = \frac{e^{m_{xt}(\hat{\theta})}}{1+e^{m_{xt}(\hat{\theta})}}, \tag{13.19}$$

or more simply $\bar{q}_{xt} = \hat{q}_{xt}$.

A logistic regression model is obtained if the unexplained systematic errors δ_{xt} in (13.13) equal zero. This is a generalised linear model (GLM) with a logit link. Then the death counts D_{xt} will have an (unconditional) binomial distribution

$$D_{xt} \sim \text{Bin}(N_{xt}, q_{xt}) = \text{Bin}\left(N_{xt}, \frac{e^{m_{xt}(\theta)}}{1+e^{m_{xt}(\theta)}}\right), \tag{13.20}$$

with $m_{xt}(\theta)$ as in (13.16). The model parameters θ can be estimated directly from untransformed raw data D_{xt} by means of maximum likelihood

$$\tilde{\theta} = \arg\max_{\theta} \prod_{(x,t)\in\Omega} P_\theta(D_{xt}|N_{xt}) \tag{13.21}$$

and by plugging these into the regression function, the mortality rate estimates (13.19) can be refined as

$$\tilde{q}_{xt} = \frac{e^{m_{xt}(\tilde{\theta})}}{1+e^{m_{xt}(\tilde{\theta})}}. \tag{13.22}$$

Renshaw and Haberman [32] also use a GLM approach with an over-dispersed Poisson distribution. When their over-dispersion parameter is set to unity, so that the data is Poisson distributed, the resulting model is very similar to (13.20).

13.4.2 Logit Mortality Increments

If the time trend in (13.13) is of central interest, we use instead $Y_{xt} = Y_{xt}^{\text{LMI}}$ in (13.3) for all (x,t) in $\Omega = \{(x,t);\ x_l \le x \le x_u, t_2 \le t \le t_T\}$. Then the three terms in (13.7), and the parameter vector, have the form

$$\begin{aligned} m_{xt} &= \sum_{j=0}^{p_2} b_j \phi_j(x), \\ \varepsilon^s_{xt} &= \delta_{xt} - \delta_{x,t-1}, \\ \varepsilon^b_{xt} &= \left(\text{logit}\hat{q}_{xt} - \text{logit}q_{xt}\right) - \left(\text{logit}\hat{q}_{x,t-1} - \text{logit}q_{x,t-1}\right), \\ p &= p_2 + 1, \\ \theta &= (b_0, \ldots, b_{p_2})^T. \end{aligned} \tag{13.23}$$

We can write this as a multiple linear regression model (13.11) with a design matrix $\mathbf{X}$ of dimension $n \times p$ whose row corresponding to (x, t) is $(\phi_0(x), \phi_1(x), \ldots, \phi_{p_2}(x))$. It follows from (13.9) that the binomial risk variance function satisfies

$$\begin{aligned} v^b_{xt} &= E\left(\text{Var}(\text{logit}\hat{q}_{x,t-1}|q_{x,t-1})\right) + E\left(\text{Var}(\text{logit}\hat{q}_{xt}|q_{xt})\right) \\ &\approx E\left(\frac{1}{N_{x,t-1}q_{x,t-1}(1-q_{x,t-1})}\right) + E\left(\frac{1}{N_{xt}q_{xt}(1-q_{xt})}\right), \end{aligned} \tag{13.24}$$

which we estimate as

$$\hat{v}^b_{xt} = \frac{1}{N_{x,t-1}\bar{q}_{x,t-1}(1-\bar{q}_{x,t-1})} + \frac{1}{N_{xt}\bar{q}_{xt}(1-\bar{q}_{xt})}, \tag{13.25}$$

with $\bar{q}_{x,t-1}$ and $\bar{q}_{xt}$ as in (13.19), using the LS estimate $\hat{\theta}$ of LM (not LMI) transformed data, or we put $\bar{q}_{x,t-1} = \hat{q}_{x,t-1}$ and $\bar{q}_{xt} = \hat{q}_{xt}$.

The LMI transformation will only be used for goodness of fit testing, not for refining mortality estimates, as in (13.22).

13.5 Variance Decomposition and Overdispersion Test

We can use (13.7–13.10) in order to define a variance decomposition of the transformed mortalities Y_{xt} as follows: Let w_{xt} be weights assigned to all elements of Ω and assume that ω is randomly chosen from Ω with probabilities proportional to w_{xt}. Then

$$E(Y_\omega) = m = m(\theta) = \sum_{(x,t)\in\Omega} w_{xt} m_{xt} / \sum_{(x,t)\in\Omega} w_{xt}.$$

Following [24, 25], we will decompose the variance of Y_ω into three parts;

$$\begin{aligned}\mathrm{Var}(Y_\omega) &= \sum_{(x,t)\in\Omega} w_{xt} E\left((Y_{xt}-m)^2\right) / \sum_{(x,t)\in\Omega} w_{xt} \\ &= \left(\sum_{(x,t)\in\Omega} w_{xt}(m_{xt}-m)^2 + \sum_{(x,t)\in\Omega} w_{xt} v^s_{xt} + \sum_{(x,t)\in\Omega} w_{xt} v^b_{xt}\right) / \sum_{(x,t)\in\Omega} w_{xt} \\ &= \sigma^2_{\mathrm{exp}} + \sigma^2_s + \sigma^2_b\end{aligned}$$

corresponding to explained, systematic unexplained and binomial variance. The weights can be chosen in many different ways, see [24]. One possibility is to use

$$w_{xt} = \lambda^{t_T - t}, \tag{13.26}$$

where $0 < \lambda \leq 1$ is a forgetting factor that quantifies to which extent older calendar years should be down-weighted. Whereas $\lambda < 1$ may be preferable when the ultimate purpose is prediction of future mortality risks, uniform weights $\lambda = 1$, i.e.

$$w_{xt} = 1 \tag{13.27}$$

are more appropriate for parameter estimation of historical data. Yet another possibility is to downweight observations with a high binomial variance. This yields a scheme

$$w_{x,t} = \left(v^b_{xt}\right)^{-1} \tag{13.28}$$

referred to in [24] as inverse non-dispersed variance weighting.

The variance components can be estimated as

$$\begin{aligned}\hat\sigma^2_{\mathrm{exp}} &= \sum_{(x,t)\in\Omega} \hat w_{xt}(\hat m_{xt}-\hat m)^2 / \sum_{(x,t)\in\Omega} \hat w_{xt} \\ \hat\sigma^2_b &= \sum_{(x,t)\in\Omega} \hat w_{xt} \hat v^b_{xt} / \sum_{(x,t)\in\Omega} \hat w_{xt} \\ \hat\sigma^2_{\mathrm{unexp}} &= \sum_{(x,t)\in\Omega} \hat w_{xt}(Y_{xt}-\hat m_{xt})^2 / \sum_{(x,t)\in\Omega} \hat w_{xt},\end{aligned} \tag{13.29}$$

where $\hat\sigma^2_{\mathrm{unexp}}$ is an estimate of the total unexplained variance $\sigma^2_{\mathrm{unexp}} = \sigma^2_s + \sigma^2_b$, $\hat v^b_{xt}$ and $\hat m_{xt} = m_{xt}(\hat\theta)$ are estimates of the binomial risk variance and regression function respectively, $\hat w_{xt} = w_{xt}$ if deterministic weights (13.26–13.27) are used, $\hat w_{xt} = (\hat v^b_{xt})^{-1}$ for inverse variance weights (13.28), and

$$\hat m = \sum_{(x,t)\in\Omega} \hat w_{xt} \hat m_{xt} / \sum_{(x,t)\in\Omega} \hat w_{xt}.$$

The coefficient of determination

$$R^2 = \frac{\hat{\sigma}^2_{\text{exp}}}{\hat{\sigma}^2_{\text{exp}} + \hat{\sigma}^2_{\text{unexp}}}$$

quantifies how large a fraction of the total variance is explained by the model. However, in this paper we will focus on the fraction

$$\rho = 1 - \frac{\sigma_b^2}{\sigma^2_{\text{unexp}}} \tag{13.30}$$

of the unexplained variance that originates from systematic risk. It represents the part of the unexplained variation which potentially could be explained. We can interpret ρ as the correlation coefficient between two random variables Y_ω and Y'_ω, computed from two hypothetical populations with the same mortalities $\mathbf{q}$, and with estimated mortalities that both satisfy (13.1) but the two populations are conditionally independent, given $\mathbf{q}$. An estimate of ρ is

$$\hat{\rho} = \left(1 - \frac{\hat{\sigma}_b^2}{\hat{\sigma}^2_{\text{unexp}}}\right)_+,$$

where a truncation is applied in order to avoid a negative estimate of a non-negative parameter. This may happen, either due to the randomness of the estimated mortalities, or if the model is over-parametrised, then a simpler model should be considered. We will use $\hat{\rho}$ as a model selection tool as follows: Let $0 < \rho_{\text{crit}} < 1$ be a pre-defined threshold value of the correlation coefficient and $\varepsilon^s = \{\varepsilon^s_{xt};\ (x,t) \in \Omega\}$ the unexplained systematic risk. Then, if

$$\begin{aligned} \hat{\rho} \le \rho_{\text{crit}} &\Longrightarrow \text{discard } \varepsilon^s, \\ \hat{\rho} > \rho_{\text{crit}} &\Longrightarrow \text{include } \varepsilon^s, \end{aligned} \tag{13.31}$$

with the rationale of choosing a simpler model when the binomial risk dominates the unexplained systematic risk. The outcome of the test (13.31) can thus serve as a tool for model selection. With $\hat{\rho}$ sufficiently close to 0, we disregard unexplained systematic variation, so that the mortality rates q_{xt} are deterministic. The death counts will then follow the logistic regression model (13.20), knowing that it will mostly explain what there is to explain. On the other hand, a high value of $\hat{\rho}$ indicates that the logistic regression model fails to explain a significant amount of variation in data, and then ε^s should be included in the model. Various ways of modelling unexplained systematic risk are discussed in Sect. 13.7.

We can regard (13.31) as a test of over-dispersion for the number of deaths D_{xt}, with null and alternative hypotheses

$$\begin{aligned} H_0 &: \rho = 0 \Leftrightarrow \varepsilon^s = \mathbf{0}, \\ H_1 &: \rho > 0 \Leftrightarrow \varepsilon^s \neq \mathbf{0}, \end{aligned}$$

respectively. Under the alternative hypothesis, the unconditional distribution of D_{xt} will be a mixed binomial, with a mixture distribution caused by the unexplained systematic risk. This gives an over-dispersion

$$\begin{aligned}\text{Var}(D_{xt}) &= \text{Var}\left(E(D_{xt}|q_{xt})\right) + E\left(\text{Var}(D_{xt}|q_{xt})\right)\\ &= N_{xt}^2\text{Var}(q_{xt}) + N_{xt}E\left(q_{xt}(1-q_{xt})\right)\\ &= N_{xt}E(q_{xt})\left(1-E(q_{xt})\right) + (N_{xt}^2 - N_{xt})\text{Var}(q_{xt})\\ &\overset{H_1}{>} N_{xt}E(q_{xt})\left(1-E(q_{xt})\right)\end{aligned}$$

of untransformed data for all $(x, t) \in \Omega$. For large populations, the transformations (13.2) and (13.3) are approximately linear functions of $\{D_{xt};\ x_l \le x \le x_u, t_1 \le t \le t_T\}$. Therefore, transformed data will be over-dispersed ($\text{Var}(Y_{xt}) > v_{xt}^b$ for all $(x, t) \in \Omega$), precisely under the alternative hypothesis, as shown in (13.10).

The threshold ρ_{crit} in (13.31) can either be defined as a fixed value, for instance in the range 0.1–0.3, depending on how much a simpler model, without unexplained systematic variation, is preferred. It can also be derived as a quantile of the null distribution of $\hat{\rho}$, which can either be approximated by parametric bootstrap, when new data is generated from the null model (13.20), but with θ replaced by an estimate $\hat{\theta}$, or from an asymptotic approximation of the null distribution of $\hat{\rho}$. It is motivated in the appendix that

$$\hat{\rho} \sim \sqrt{\frac{C}{n}}U_+ \text{ under } H_0 \tag{13.32}$$

when the number of elements n of Ω is large, where $U \sim N(0, 1)$ and C is a constant that depends on the weighting scheme and the type of transformation used. For LM transformed data, all Y_{xt} are independent under the null hypothesis, and therefore

$$C = \frac{2n\sum_{(x,t)}\left(w_{xt}v_{xt}^b\right)^2}{\left(\sum_{(x,t)}w_{xt}v_{xt}^b\right)^2},$$

with a minimum value of $C = 2$ for inverse variance weighting (13.28). For LMI transformed data, C will be slightly smaller, as motivated in the appendix.

We see from (13.32) that the null distribution of $\hat{\rho}$ is approximately a 0.5:0.5 mixture of a point mass at zero and a continuous one-dimensional distribution. More generally, statistics for testing parameters at the boundary of a parameter space often have null distributions that are mixtures of distributions of different dimensions [33, 34].

13.6 Data Analysis

In this section, we proceed with a data analysis in order to investigate whether the simple model (13.13) could be used for Swedish, UK and US data sets.

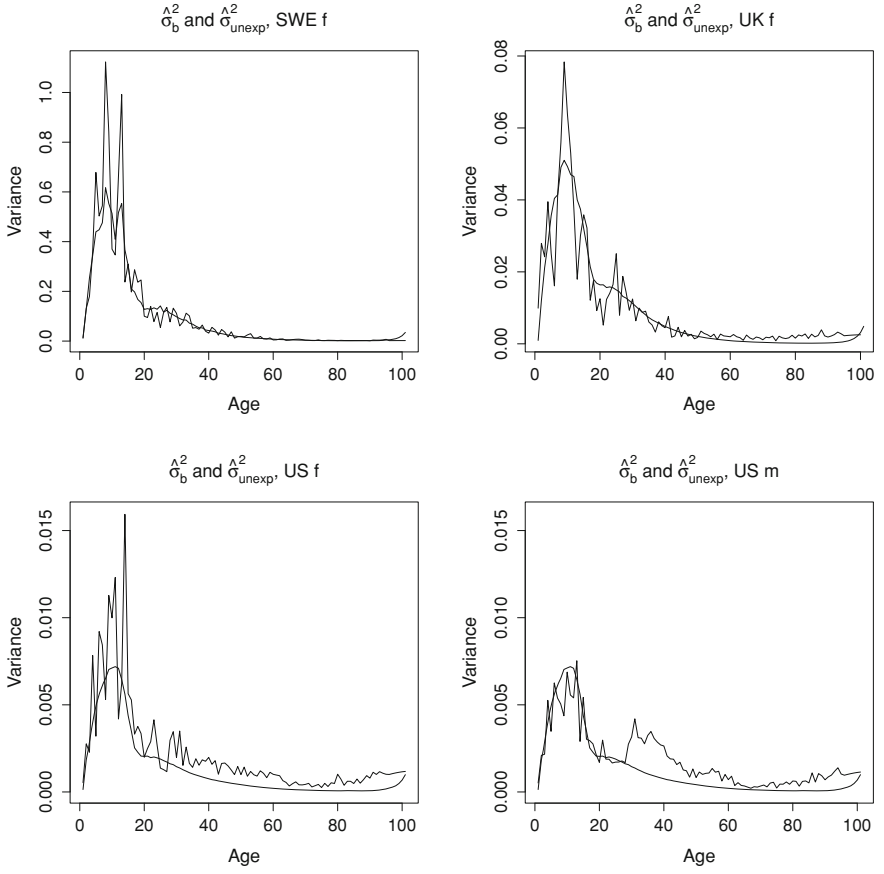


Fig. 13.3 Estimated age-specific variance components $\hat{\sigma}^2_{bx}$ and $\hat{\sigma}^2_{\text{unexp},x}$, for the Swedish female, UK female and US datasets as a function of age x. We use uniform weights (13.27) and for the $\hat{v}_{b,x}$ estimate in (13.25), we put $\bar{q}_{xt} = \hat{q}_{xt}$. In all four subplots, the more smoothed curves represent the estimated binomial variances

13.6.1 Variance Decomposition

We start by fitting a multiple linear regression model (13.23) to LMI data, with single age class indicators as defined in (13.15). We then compute estimated variance components $\hat{\sigma}^2_{bx}$ and $\hat{\sigma}^2_{\text{unexp},x}$ given in (13.29), when Ω consists of one single age, class, i.e. $x_l = x_u = x$, using uniform weights (13.27), see Fig. 13.3.

Some general features can be seen. Since the binomial variance $\hat{\sigma}^2_{bx} \propto q_x^{-1}$ for younger ages, it has a maximum around the age of ten, then it declines, but starts to grow again at age 90 due to the rapidly declining population size. For US data, the unexplained variance is above the binomial variance for almost all ages, and for all over 25. For the Swedish data set the two variances are very close to each other.

Table 13.2 Estimated correlation coefficient $\hat{\rho}$ for different age bands and populations based on LM and LMI data

Transformation	Age	0–100		1–45		46–60		61–90	
	Quantity	$\hat{\rho}$	ρ_{crit}	$\hat{\rho}$	ρ_{crit}	$\hat{\rho}$	ρ_{crit}	$\hat{\rho}$	ρ_{crit}
LM	SWE f	0.15	0.0488	0.10	0.0730	0.14	0.1265	0.29	0.0895
	UK f	0.80	0.0504	0.02	0.0754	0.50	0.1307	0.92	0.0924
	US f	0.90	0.0521	0.72	0.0781	0.84	0.1353	0.94	0.0956
LMI	SWE f	0.12	0.0495	0	0.0742	0.10	0.1285	0.34	0.0909
	UK f	0.68	0.0512	0	0.0767	0.37	0.1329	0.86	0.0940
	US f	0.72	0.0531	0.36	0.0795	0.68	0.1377	0.83	0.0974

British data is in between, $\hat{\sigma}^2_{bx}$ and $\hat{\sigma}^2_{\text{unexp},x}$ are very close up to age 45–50, then the unexplained variance rises above the binomial variance. Since the estimated variance components vary a great deal over ages, we use the inverse non-dispersed variance weighting (13.28) when computing $\hat{\rho}$. In order not to confound finer nuances at higher ages with larger variances from lower ages, we calculate $\hat{\rho}$ for several age bands, as it is apparent that the ratio $\hat{\sigma}^2_{bx}/\hat{\sigma}^2_{\text{unexp},x}$ changes with age.

For Swedish data, $\hat{\rho}$ is significantly different from 0, but yet so low that we can settle for a simpler regression model with systematic unexplained risk excluded. The same holds for the UK, up to an age of 45. For the US data, the systematic unexplained variation $\hat{\sigma}^2_s$ is the main source of unexplained variation. Hence one should try a model with systematic unexplained risk included.

Table 13.2 presents the estimated correlation coefficient $\hat{\rho}$ for different age bands and populations based on LM and LMI data, no age-wise smoothing (13.15) and inverse non-dispersed variance weighting (13.28). Also displayed is the approximation (13.32) of the 0.975 quantile $\rho_{\text{crit}} = 1.96\sqrt{2/n}$ of the null distribution, using $C = 2$ for the asymptotic variance, which is exact for LM and a conservative upper bound for LMI. The number of age-year cells n equals $(x_u - x_l + 1)T$ and $(x_u - x_l + 1)(T - 1)$ respectively for LM and LMI data, with $T = 32$, 30 and 28 for the female SWE, UK and US populations.

13.6.2 Residual Plots

Even with the above finding, it is instructive to study the residual plots for the simple regression model (13.23) based on LMI transformed data, with systematic unexplained risk disregarded. In Fig. 13.4 we have plotted the residuals

$$\hat{\varepsilon}_{xt} = Y_{xt} - \hat{m}_{xt}$$

from an ordinary least squares fit (13.12).

When one observes clear patterns in a residual plot it is a sign that there are systematic effects, not captured by the model. One can then ask if one should modify or extend the model in order to explain them by the covariates, or if such an extension

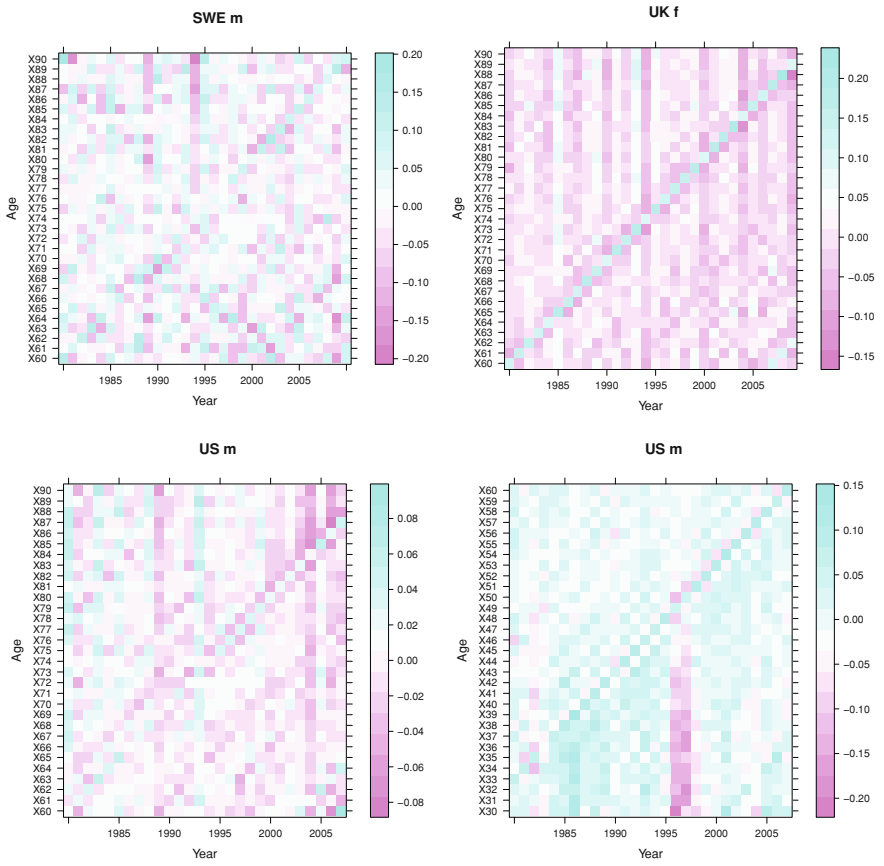


Fig. 13.4 Residuals of a least squares fit (13.12) to one year increments (13.3) of estimated logit mortality rates for Swedish male, UK female and US data

adds more complexity than is motivated by these effects. The patterns can in some cases provide additional insight into the underlying processes.

13.6.2.1 Calendar Year Effects

Calendar year effects can be seen as vertical lines in the residual plots. They can be spotted mostly in higher ages, above 60, and a probable cause are phenomena such as a seasonal influenza, heat waves and cold spells that are known to vary in severity from year to year. There is a notable exception from the old age only effects. An increase in mortality for US males in their 30s appears during 1985–89, when the AIDS epidemic started and a steep drop is observed in 1996–97, the same years as the HIV inhibitor medicines reached the markets. This effect is also evident in the

inter age correlation graph, see Fig. 13.2, where a high degree of correlation is seen among US males in this ageband.

The calendar year effects, while recurring, seem to be random in nature. They could be incorporated in a random effect model but not in an ordinary regression model with error terms that are independent between ages *and* calendar years.

13.6.2.2 Cohort Effects

There are some evident cohort effects in the residual plots. One, emanating from the generation around 1920, is more or less evident for all studied populations. Another stems from the 1946–47 generations, although this is not visible for Swedish data.

What these periods have in common is that they are post war years with a baby boom. Birth rates in the UK went up almost 40 % in 1920 and 22 % in 1946, whereas in Sweden they went up by 20 % in 1920, but no particular increase occurred in 1946, since births started to increase already in 1942 and did so in a steadier fashion.

A sudden increase (or drop) in birth rate skews the distribution of births over the year, and this might lead to a systematic error in estimating N_{xt} around that cohort ([36], p. 11). So these single year cohort effects might be due to statistical errors rather than real effects.

13.6.3 Estimated and Predicted Mortality Rates

For Swedish data we disregard systematic unexplained risk and perform the regression analysis (13.20) with an age-specific parametrisation based on (13.15). The results are very similar to the least squares estimates (13.12) (not shown here) obtained from LM transformed data.

In Fig.13.5 we plot both estimated and empirical logit mortalities for 2011 as well as the estimated trend, for Swedish females.

The mortality improvement varies in a wavelike pattern over ages. It is fastest for infants with almost -0.04 per year, from age 85 the improvements decrease in a linear manner to age 100 were very small improvements are observed.

Extrapolation of the trend gives a prediction of future mortality. However, more plausible results may be obtained by first smoothing the trend using for example the polynomial parametrisation based on (13.14).

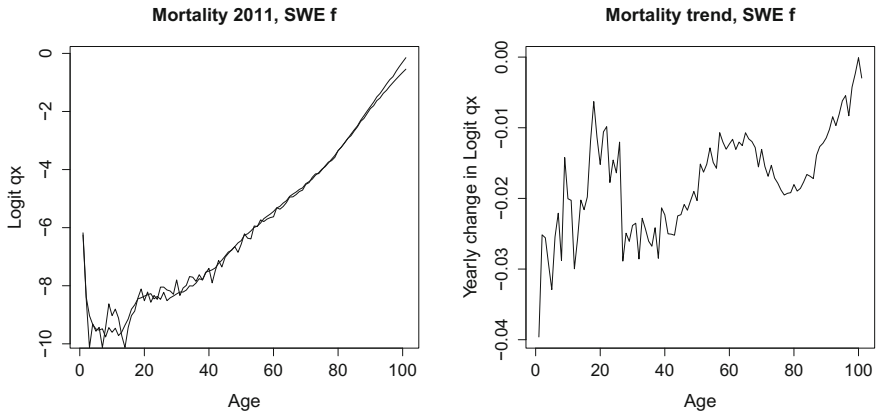


Fig. 13.5 *Left* Estimates of logit morality rates ($\text{logit}\hat{q}_{xt}$) for Swedish females of different ages x in calendar year $t = 2011$, together with a logistic regression analysis (13.20) with fitted logit mortality rates ($\text{logit}\tilde{q}_{xt}$) from (13.22). *Right* Corresponding estimates $\tilde{\beta}_x$ of the one year increments of the logit mortality rates. An age-specific parametrisation is used, based on (13.15)

How long into the future should the present trend be extrapolated?

Looking further back in mortality data it is clear that there have been shifts in the speed of improvements over different age spans and time periods.

We can think of different scenarios that will change the present trend, but predicting if and when is not possible within the framework of the model.

13.7 Discussion

In this paper we have focused on the stochastics of mortality rates, starting with an explorative data analysis. Using data from Sweden, the UK and the USA, we found clear signs of randomness in the logit mortalities for Swedish data, after a linear trend had been removed, whereas for US data, there was more underlying structure.

In order to quantify these effects and separate random noise from over-dispersion in terms of systematic unexplained variation, we fitted a parametric regression function, where logit mortalities have a deterministic linear period effect, with a separate intercept and slope parameter for each age class. Then we performed a variance decomposition of the residual variance, which enabled us to quantify the amount of unexplained variation in terms of systematic and diversifiable (binomial) risk. Based on formulas for estimates of these two variance components we were able to calculate an estimate of the fraction ρ of the unexplained variance that originates from systematic unexplained variation.

We found that the estimates of ρ were low for Swedish data, around 0.15, depending on the age span. The somewhat surprising conclusion is that a naive regression model captures the essentials, leaving very little further variance to a more elaborate model to explain. However, for US data ρ was estimated to values around 0.9,

indicating a lot of over-dispersion or systematics effects not captured by the naive regression model. Looking at residual plots we see the existence of calendar year effects, indicating that this is something that should be included in a model with a better fit. UK data falls somewhere in between. For ages 1–45, the simple model without systematic unexplained variation explains almost everything in regards of variance, but for higher ages there is substantial unexplained variation.

Population size is the key here, even with almost 10 million inhabitants in Sweden, almost all underlying changes in the estimated mortality rates $\hat{q}_{xt}$, except for the deterministic trend, is drowned by random binomial noise. This would certainly be the case even for smaller populations. For the practitioner working with mortality in a life or pension company the message is clear, keep your models simple!

If the estimated fraction of unexplained systematic variance ρ is small, we suggested to predict mortality rates from a logistic regression fit of raw data. On the other hand, if the estimated ρ large, this signifies a non-negligible amount of unexplained systematic variation. Then there are several possible ways to proceed. Firstly, a logistic regression analysis could be employed, but with an enlarged parametric model. Secondly, a low-dimensional parametric model could be retained, but with overdispersion explicitly modelled, using for instance negative binomial distributions [30] or generalised linear models with over-dispersed Poisson data [18, 32] for which parameters can be estimated by extended quasi-likelihood methods, or some generalised linear mixed model. In [20], we propose modelling the unexplained systematic variation as a time series that includes a random white noise component, a random walk component, and a third seasonal effects component that incorporates correlation between age classes. Thirdly, nonparametric smoothing methods can be employed, such as two-dimensional penalised splines [16], Generalised Additive Models [22] or Kalman filter techniques for time series [21].

We have argued that a simple logistic regression model often works well for fitting mortality data in a small country. However, when prediction of future mortalities is of concern, it is often more flexible to have a random component of systematic variation. This facilitates calculation of more realistic predictive distributions and simulation of various scenarios of future mortalities. See [9, 12, 14, 21] and references therein for more details on mortality prediction.

Acknowledgments Ola Hössjer's research was financially supported by the Swedish Research Council, contract nr. 621-2008-4946, and the Gustafsson Foundation for Research in Natural Sciences and Medicine.

Appendix

Motivation of (13.32). We first rewrite and approximate (13.30) as

$$\begin{aligned}\hat{\rho} &= \left(\frac{\sum_{(x,t)} \hat{w}_{xt}\left((Y_{xt}-\hat{m}_{xt})^2-\hat{v}^b_{xt}\right)}{\sum_{(x,t)} \hat{w}_{xt}(Y_{xt}-\hat{m}_{xt})^2}\right)_+ \\ &\overset{H_0}{\approx} \left(\frac{\sum_{(x,t)} w_{xt}\left((Y_{xt}-m_{xt})^2-v^b_{xt}\right)}{\sum_{(x,t)} w_{xt}(Y_{xt}-m_{xt})^2}\right)_+ \\ &\approx \left(\frac{\sum_{(x,t)} w_{xt}v^b_{xt}\left(U^2_{xt}-1\right)}{\sum_{(x,t)} w_{xt}v^b_{xt}U^2_{xt}}\right)_+,\end{aligned} \tag{13.33}$$

where U_{xt} are standard normal variables that approximate the null hypothesis Pearson residuals $(Y_{xt} - \hat{m}_{xt})/\sqrt{\hat{v}^b_{xt}}$.

In order to motivate the approximations in (13.33), we first notice that the last step follows from a Multivariate Central Limit Theorem

$$\left((Y_{xt} - m_{xt})/\sqrt{v^b_{xt}};\ (x,t) \in \Omega\right) \xrightarrow{\mathscr{L}} \mathbf{U} = (U_{xt};\ (x,t) \in \Omega)\,,$$

under H_0 as the population size tends to infinity, and $\mathbf{U} \sim N(\mathbf{0}, \Sigma)$ has a multivariate normal distribution, with a covariance matrix $\Sigma = (\Sigma_{xt,x't'})$ that equals the covariance matrix of $\left(\varepsilon^b_{xt}/\sqrt{v^b_{xt}};\ (x,t) \in \Omega\right)$.

For LM transformed data, it follows from (13.16) that all ε^b_{xt} are independent, and by definition in (13.10), $v^b_{xt} = \mathrm{Var}(\varepsilon^b_{xt})$. Therefore Σ equals the identity matrix of order n. For LMI transformed data, it follows analogously from (13.23) that ε^b_{xt} are no longer independent. Therefore the elements of Σ are slightly more complicated; $\Sigma_{xt,xt} = 1$, $\Sigma_{xt,x't} = 0$ if $x \neq x'$, $\Sigma_{xt,x't'} = 0$ if $|t - t'| \geq 2$ and

$$\Sigma_{xt,x,t+1} = -\frac{1}{N_x(q_{xt}(1-q_{xt}))\sqrt{v^b_{xt}v^b_{x,t+1}}},$$

where

$$v^b_{xt} = \frac{1}{N_{x,t-1}q_{x,t-1}(1-q_{x,t-1})} + \frac{1}{N_{xt}q_{xt}(1-q_{xt})}$$

is the expression for the binomial variance (13.24) when there is no overdispersion.

For the second step of (13.33) we assume for simplicity that weights are uniform, introduce

$$\bar{v}^b = \max_{(x,t)\in\Omega} v^b_{xt},$$

and notice that under the null hypothesis the numerator and denominator of the second line of (13.33) satisfy

$$\begin{aligned}\sum_{x,t}\left((Y_{xt}-m_{xt})^2-v_{xt}^b\right)&=O_p\left(n^{1/2}\bar{v}^b\right)\\ \sum_{x,t}(Y_{xt}-m_{xt})^2&=O_p\left(n\bar{v}^b\right),\end{aligned}\tag{13.34}$$

where $X_n=O_p(A_n)$ is a sequence of random variables such that X_n/A_n is bounded in probability as n grows.

Under mild regularity conditions, the least squares estimator $\hat{\theta}$ is consistent as n grows, at a rate $|\hat{\theta}-\theta|=O_p\left(n^{-1/2}(\bar{v}^b)^{1/2}\right)$, both in (13.16) and (13.23), see for instance [23] for asymptotics of linear regression estimators. It can be seen that this leads to approximation errors between the numerators and denominators of the first and second lines of (13.33) that equal

$$\begin{aligned}\sum_{x,t}\left((Y_{xt}-\hat{m}_{xt})^2-(Y_{xt}-m_{xt})^2\right)&=O_p\left(n|\hat{\theta}-\theta|^2\right)=O_p\left(\bar{v}^b\right),\\ \sum_{x,t}\left(\hat{v}_{xt}^b-v_{xt}^b\right)&=O_p\left(n^{1/2}(\bar{v}^b)^{3/2}\right),\end{aligned}\tag{13.35}$$

using Taylor expansions of $m_{xt}=m_{xt}(\theta)$ with respect to $\theta=\theta^{\mathrm{LM}}$ or $\theta=\theta^{\mathrm{LMI}}$ in the upper equation, and another Taylor expansion of $v_{xt}^b=v_{xt}^b(\theta^{\mathrm{LM}})$ with respect to θ^{LM} in the lower equation. Under the null hypothesis we have for LM transformed data that (13.17) and (13.20) simplify so that

$$v_{xt}^b=\frac{1}{N_{xt}q_{xt}(1-q_{xt})}=\frac{\left(1+e^{m_{xt}(\theta^{\mathrm{LM}})}\right)^2}{N_{xt}e^{m_{xt}(\theta^{\mathrm{LM}})}},$$

and analogously (13.24) can be simplified for LMI transformed data. In either case we find that

$$M=\max_{(x,t)\in\Omega}\left|\frac{\partial v_{xt}^b}{\partial\theta^{\mathrm{LM}}}\right|=O(\bar{v}^b),$$

which was used on the right-hand side of the second equation of (13.35), since the left-hand side can be bounded above by $O_p(Mn|\hat{\theta}-\theta|)$.

We conclude that the approximation errors in (13.35) are of smaller order than the relevant main terms in (13.34), and this justifies the second step of (13.33).

In order to motivate (13.32), we use the Central Limit Theorem for the numerator and Law of Large Numbers for the denominator of the ratio within the $(\cdot)_+$ sign on the third line of (13.33). From this we deduce that the ratio has an asymptotic $N(0,2C/n)$ distribution, with

$$C=\frac{n\sum_{(x,t),(x',t')}w_{xt}v_{xt}^b\cdot w_{x't'}v_{x't'}^b\cdot\mathrm{Cov}\left(U_{xt}^2,U_{x't'}^2\right)}{\left(\sum_{(x,t)}w_{xt}v_{xt}^b\right)^2},$$

which for LM transformed data reduces to (13.32), since Σ is then the identity matrix of order n and $\mathrm{Var}(U^2_{xt}) = 2$ for all $(x, t) \in \Omega$. For LMI transformed data, the negative correlations between $U^2_{x,t}$ and $U^2_{x,t+1}$ make C slightly smaller, and in particular $C < 2$ for inverse variance weighting.

References

1. Alho, J.M.: Discussion of Lee. North Am. Actuar. J. **4**, 91–93 (2000)
2. Barrieu, P., et al.: Understanding, modelling and managing longevity risk: key issues and main challenges. Scand. Actuar. J. **3**, 203–231 (2012)
3. Bell, W.R.: Comparing and assessing time series methods for forecasting age-specific fertility and mortality rates. J. Official Stat. **13**(3), 279–303 (1997)
4. Bell, W.R., Monsell, B.C.: Using principal components in time series modelling and forecasting of age specific mortality rates. In: Proceedings of the American Statistical Association, Social Statistics Session, pp. 154–159 (1991)
5. Booth, H., Tickle, L.: Mortality modelling and forecasting: a review of methods. Ann. Actuar. Sci. **3**(I/II), 3–43 (2008)
6. Bozik, J.E., Bell, W.R.: Forecasting age specific mortality using principal components. In: Proceedings of the American Statistical Association, Social Statistics Session, pp. 396–401 (1987)
7. Breslow, N.E., Clayton, D.G.: Approximate inference in generalized linear mixed models. J. Am. Stat. Assoc. **88**(421), 9–25 (1993)
8. Brockwell, P.J., Davis, R.A.: Time Series: Theory and Methods, 2nd edn. Springer, New York (1991)
9. Brouhns, N., Denuit, M., van Keilegom, I.: Bootstrapping the log-bilinear model for mortality forecasting. Scand. Actuar. J. **2005**(3), 212–224 (2005)
10. Brouhns, N., Denuit, M., Vermunt, J.K.: A poisson log-bilinear regression approach to the construction of projected lifetables. Insur. Math. Econ. **31**, 373–393 (2002)
11. Cairns, A.J.G., Blake, D., Dowd, K.: Modeling and management of mortality risk: a review. Scand. Actuar. J. **2008**(2–3), 79–113 (2008)
12. Cairns, A.J.G., Blake, D., Dowd, D., Couglan, G.D., Epstein, D., Ong, A., Balevich, I.: A quantitative comparison of stochastic mortality models using data from england and wales and the united states. North Am. Actuar. J. **13**, 1–35 (2009)
13. Cairns, A.J.G., Blake, D., Dowd, D., Couglan, G.D., Khalaf-Allah, M.: Bayesian stochastic mortality modelling for two populations. ASTIN Bull. **41**, 29–59 (2011b)
14. Cairns, A.J.G.: Modelling and management of longevity risk. Manuscript (2013)
15. Cairns, A.J.G., Dowd, K., Blake, D., Guy, D.: Longevity hedge effectiveness: a decomposition. Quant. Financ. **14**(2), 217–235 (2014)
16. Currie, I.D., Durban, M., Eilers, P.H.C.: Smoothing and forecasting mortality rates. Stat. Model. **4**, 279–298 (2004)
17. Denton, F.T., Feaver, C.H., Spencer, B.G.: Time series analysis and stochastic forecasting: an econometric study of mortality and life expectancy. J. Popul. Econ. **18**, 223–227 (2004)
18. Djeundje, V.A.B., Currie, I.D.: Smoothing dispersed counts with applications to mortality data. Ann. Actuar. Sci. **5**(1), 33–52 (2010)
19. Eilers, P.H.C., Marx, B.D.: Flexible smoothing with *b*-splines and penalties (with discussion). Stat. Sci. **11**, 89–121 (1996)
20. Ekheden, E., Hössjer, O.: Multivariate time series modeling, estimation and prediction of mortalities. Submitted (2014)
21. Guerrero, V.M., Silva, E.: Non-parametric and structured graduation of mortality rates. Popul. Rev. **49**(2), 13–26 (2010)

22. Hall, M., Friel, N.: Mortality projections using generalized additive models with applications to annuity values for the irish population. Ann. Actuar. Sci. **5**(1), 19–32 (2010)
23. Huber, P.J.: Robust regression: asymptotics, conjectures and monte carlo. Ann. Stat. **1**(5), 799–821 (1973)
24. Hössjer, O.: On the coefficient of determination for mixed regression models. J. Stat. Plan. Infer. **138**, 3022–3038 (2008)
25. Hössjer, O., Eriksson, B., Järnmalm, K., Ohlsson, E.: Assessing individual unexplained variation in non-life insurance. ASTIN Bull. **39**(1), 249–273 (2009)
26. Imoto, S., Konishi, S.: Selection of smoothing parameters in *b*-spline nonparametric regression models using information criteria. Ann. Inst. Statist. Math. **55**(4), 671–687 (2003)
27. Lee, R.D., Carter, L.R.: Modelling and forecasting U.S. mortality. J. Am. Stat. Assoc. **87**(419), 659–671 (1992)
28. Lee, Y., Nelder, J.A., Pawitan, Y.: Generalized linear models with random effects. Unifiied Analysis via *H*-likelihood. Chapman and Hall/CRC, Boca Raton (2006)
29. Li, N., Lee, R.: Coherent mortality forecasts for a group of populations: an extension of the lee-carter model. Demography **42**(3), 575–594 (2005)
30. Li, J.S.H., Hardy, M.R., Tan, K.S.: Uncertainty in mortality forecasting: an extension to the classic lee-carter approach. ASTIN Bull. **39**, 137–164 (2009)
31. Osmond, C.: Using age, period and cohort models to estimate future mortality rates. Int. J. Epidemiol. **14**, 124–129 (1985)
32. Renshaw, A., Haberman, S.: Lee-carter mortality forecasting: a parallel generalized linear modelling approach for england and wales mortality projections. Appl. Stat. **52**(1), 119–137 (2003)
33. Self, S.G., Liang, K.Y.: Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J. Am. Stat. Assoc. **82**, 605–610 (1987)
34. Silvapulle, M.J., Sen, P.K.: Constrained Statistical Inference: Order, Inequality and Shape Constraints. Wiley, Hoboken (2005)
35. Wilmoth, J.R.: Computational methods for fitting and extrapolating the Lee-Carter model of mortality change. Technical Report, Department of Demography, University of California, Berkeley (1993)
36. Wilmoth, J.R., Andreev, K., Jdanov, D., Glei, D.A.: Methods protocol for the Human Mortality Database (2007). http://www.mortality.org/Public/Docs/MethodsProtocol.pdf