Coherent forecasting of multiple-decrement life tables: a test using Japanese cause of death data.

Jim Oeppen
Max Planck Institute for Demographic Research, Rostock, Germany;
oeppen@demogr.mpg.de

Abstract

Planners in public and private institutions would like coherent forecasts of the components of age-specific mortality, such as causes of death. This has been difficult to achieve because the relative values of the forecast components often fail to behave in a way that is coherent with historical experience. In addition, when the group forecasts are combined the result is often incompatible with an all-groups forecast. It has been shown that cause-specific mortality forecasts are pessimistic when compared with all-cause forecasts (Wilmoth, 1995). This paper abandons the conventional approach of using log mortality rates and forecasts the density of deaths in the life table. Since these values obey a unit sum constraint for both conventional single-decrement life tables (only one absorbing state) and multiple-decrement tables (more than one absorbing state), they are intrinsically relative rather than absolute values across decrements as well as ages. Using the methods of Compositional Data Analysis pioneered by Aitchison (1986), death densities are transformed into the real space so that the full range of multivariate statistics can be applied, then back-transformed to positive values so that the unit sum constraint is honoured. The structure of the best-known, single-decrement mortality-rate forecasting model, devised by Lee and Carter (1992), is expressed in compositional form and the results from the two models are compared. The compositional model is extended to a multiple-decrement form and used to forecast mortality by cause of death for Japan.

Key words: Compositional Data Analysis; mortality; cause of death; forecast; Japan; Lee-Carter; life tables; single-decrement; multiple-decrement.

1 Introduction

Disaggregated mortality forecasts offer two main advantages. Public and private financial institutions are exposed to longevity risk - the commitment to make future undefined payments in the form of pensions, annuities, and health expenditures for care and medical costs. The financial risks are known to vary by age, sex, time to death, cause of death, and other factors. The monetary sums involved are enormous and temporal shifts in these factors imply large changes for future costs. For example, it has been estimated that the health expenditure in the last year of life for a female who dies from cancer in the Netherlands is twice that of a female who dies from a heart attack (Polder et al., 2006).

The second reason for disaggregating mortality forecasts is that social planners have a duty to intervene so that morbidity and mortality are reduced, subject to budget constraints. Spending on research, capital investment, preventive measures, or palliative care could all be more efficient if accurate forecasts for sub-groups were available. While effective intervention would gradually invalidate a forecast, the statistical tools developed for disaggregated forecasting should also allow more rapid detection of changes in trends.

Thus there is both an attractive logic to, and a strong demand for, mortality forecasting of subgroups within populations. At the same time, it is desirable that the disaggregated forecast should be coherent with an overall forecast. This has been difficult to achieve in practice, so that disaggregated forecasts are not currently recommended by demographers as a general approach. For example, separate sex, or spatial, forecasts using the Lee-Carter model (Lee and Carter, 1992) have been forced to address these coherence problems (Lee and Nault, 1993; Li and Lee, 2005). The components frequently diverge in the long run in ways that are implausible when compared with the historical record. Similar problems apply in the case of multiple-decrement forecasts where there are several absorbing states, for example in cause of death forecasting. Wilmoth (1995) quantified the way in which forecasts based on proportional changes in mortality rates disaggregated by cause of death tend to be more pessimistic than single decrement forecasts. Over time, the causes with the lower rates of decline come to dominate the projection. It is clear that the dependencies, or relative balances, between the decrements have not been adequately modeled.

Multiple-decrement models maintain intra-age mathematical coherence for the following mortality measures: (Preston et al., 2001, p. 72)

$$\sum_i {}_nd_x^i = {}_nd_x; \qquad \sum_i {}_nm_x^i = {}_nm_x; \qquad \sum_i {}_nq_x^i = {}_nq_x; \tag{1}$$

where ${}_nd_x^i$ is the density of deaths from cause i in the life table in the age interval x to $x+n$, ${}_nm_x^i$ is the central death rate (deaths divided by person-years of exposure), and ${}_nq_x$ is the proportion who die before age $x+n$, conditional on survival to age x.

Causes of death are assumed to be independent when conditioned by age. This means that it is possible to make independent forecasts of the cause-specific components ${}_nm_x^i$ or ${}_nq_x^i$ and their combination should produce a plausible life table, although practical experience suggests that it does not. Independent cause-specific forecasts for ${}_nd_x^i$ cannot be used because they have to obey the additional constraint that

$$\sum_i \sum_x {}_nd_x^i = 1 \tag{2}$$

This constraint imposes an inter-age coherence on the distribution of lifetimes. Thus they are intrinsically relative rather than absolute values across decrements as well as ages. Changes in the density by age and cause have to be compensated by changes in other ages or causes. Rather than seeing this as a disadvantage, this paper argues that this is precisely what we would like to achieve in a multiple-decrement model, since disaggregated models usually fail because they do not maintain appropriately changing relative values across decrements. It is hoped that this feature can be exploited to obtain coherent forecasts. Modeling changes in the distribution of lifetimes d_x

was explored by Vaupel and Yashin (1986, 1987). It treats mortality improvement as saving lives and thus focuses on perturbation of the lifetime distribution.

Choosing to model $d^i_{x,t}$ is an unusual step, although there has been some recent interest in the analysis of this life table function in a non-forecasting context (Edwards and Tuljapurkar, 2005; Canudas-Romo, 2006). The established practice in mortality forecasting is to model the log transform of the central mortality rates $m_{x,t}$. This has two advantages. It exploits the approximately log-linear decline in mortality by age with respect to time. Secondly, it transforms the rates into the real space so that multivariate statistical methods for unbounded variables can be employed in the knowledge that the inverse transform ensures that all estimated rates exist in the positive real space.

The decision to model $d^i_{x,t}$ subject to the constraint in (2) requires a statistical methodology that is unfamiliar to demographers. The field is known as Compositional Data Analysis and was pioneered by Aitchison (1986).

The structure of the paper takes the following form: an introduction to Compositional Data Analysis (CoDa); construction of the CoDa equivalent of the Lee-Carter model; exploring the CoDa model in the single-decrement case; extension to the multiple-decrement form; comparing the single and multiple decrement models; and discussion.

2 Compositional Data Analysis

A composition is defined as a strictly positive vector of length D with a fixed sum. Compositional data arise when one has many observed compositions, all having the same sum. Such data are often expressed as percentages or densities. The example we consider in this paper is the d_x function of the life-table, which always sums to the life-table radix[1]. Another composition familiar to demographers is the components of the dependency ratio.

Aitchison (1986) argued that compositions provide only relative information, and that the appropriate space for analysis of sum-constrained data is the bounded space of the simplex, not the real space. For obvious reasons it is easier to understand compositional analysis when the simplex can be represented in two dimensions. Figure 1 shows a ternary plot that represents the simplex for the d_x function of Swedish females downloaded from the Human Mortality Database (2008) (HMD), divided into three large age groups: $0-14, 15-64, 65-110$. Where the dotted line meets the vertex represents unity for a proportion. Where it meets the base-line represents zero.

The dotted lines intersect at the uniform composition $(0.33, 0.33, 0.33)$ and we can see that the three components were effectively equal in the second half of the 18th century. By the modern era, deaths at older ages dominate, middle aged deaths comprise about 8% and the share of young deaths has collapsed.

Aitchison defined a remarkable set of operators that perform compositional addition, subtraction, multiplication and centreing on the simplex.[2] These operators obey the usual algebraic rules for

[1]For demographic modeling, it can be noted that L_x/e_0, and some measures of fertility and parity such as f_x/TFR, are also compositions. Compositional Data Analysis could give natural models for deriving age-specific tables from forecasts of the denominators.

[2]The R software for compositional analysis used in this paper was based on the library R:compositions written by K. G. van den Boogaart and available at ttp://www.stat.boogaart.de/compositions/

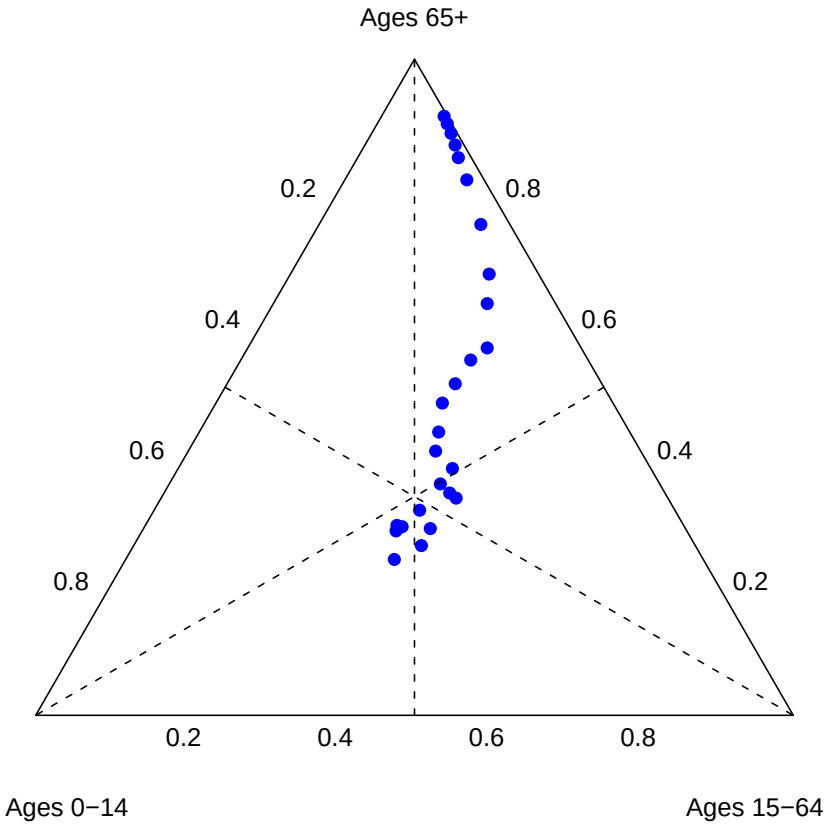

Figure 1: Sweden, female, 1751-(10)-2001: proportionate life-table deaths by age-groups

these processes. Compositional perturbations for addition and subtraction are defined by

$$x = z \oplus \hat{\xi} = \left[z_1\hat{\xi}_1, \ldots, z_D\hat{\xi}_D\right] / \left[z_1\hat{\xi}_1 + \ldots + z_D\hat{\xi}_D\right] \tag{3}$$

$$= C\left[z_1\hat{\xi}_1, \ldots, z_D\hat{\xi}_D\right] \tag{4}$$

$$z = x \ominus \hat{\xi} = C\left[\frac{x_1}{\hat{\xi}_1}, \ldots, \frac{x_D}{\hat{\xi}_D}\right] \tag{5}$$

where the closure operator $C[\ldots]$ ensures a unit sum:

$$C\left[x_i, \ldots, x_D\right] = \frac{(w_i, \ldots, w_D)}{(w_i + \ldots + w_D)} \tag{6}$$

Figure 2 shows compositional addition. The perturbation p, which is itself a composition shown by the black dot, is applied to x_0, x_1, etc.. The share of component y dominates x_0, but p is dominated by x, so successive perturbations drive the composition towards domination by X.

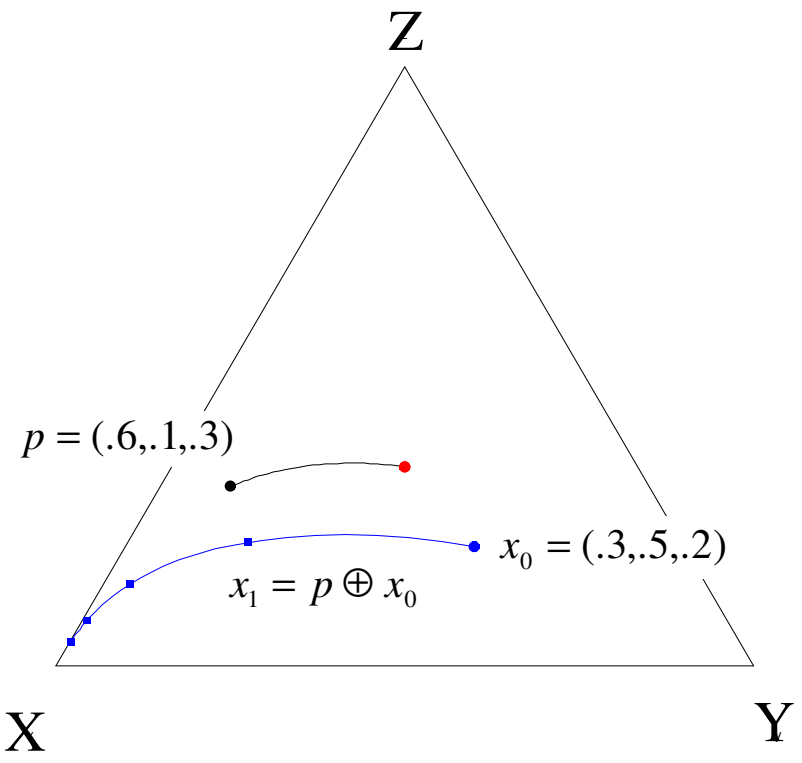

Figure 2: Compositional addition

The effect of the perturbation p can be assessed by its position relative to the uniform composition, marked by a red dot.

Compositional multiplication is defined by

$$t \otimes x = C\left[x_1^t, \ldots, x_D^t\right] \tag{7}$$

The effect is shown in Figure 3, where the dots mark the effect of multiplying the composition x_0 by successive values of the scalar t. The filled blue dot represents the result when $t = 1$ and the red dot is the uniform composition defined when $t = 0$. Positive values of t drive the composition towards its dominant component, while negative values drive it towards its minimum component.

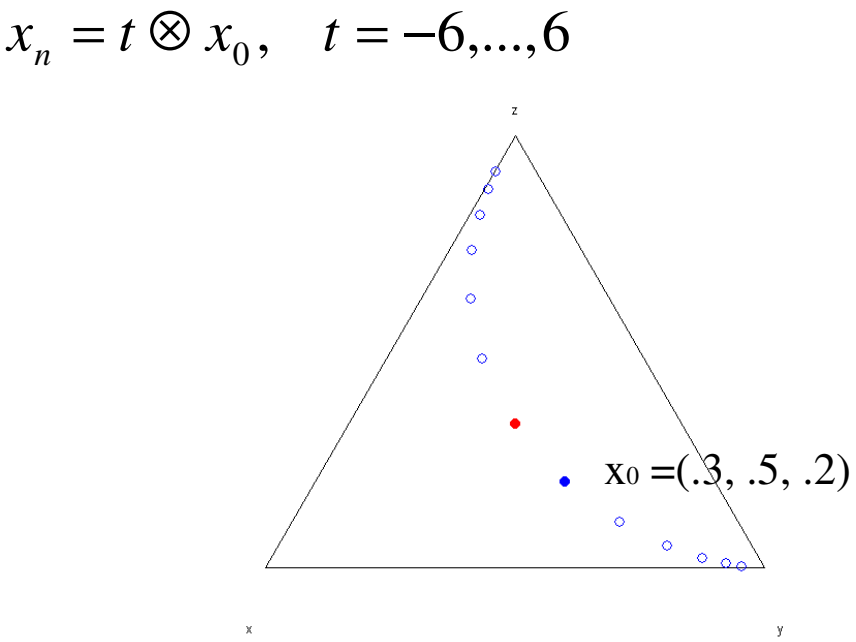

Figure 3: Compositional multiplication

If we assume a compositional data set with D columns and N rows summing to unity, then centreing is defined by calculating the geometric mean composition $\hat{\xi}$ and subtracting it from each row.

$$\hat{\xi} = cen(x) = C[\exp\{E(\log x)\}] \tag{8}$$
$$= C[g_1, \ldots, g_D] \tag{9}$$

where g_i is the geometric mean of the i^{th} column.

The statistical pitfalls of modeling compositional data have been known for more than a century. They arise because of the dependency between the columns - specifically the covariance matrix of compositional data always has at least one negative term in each row (Aitchison, 1986). Aitchison's innovation was to open the sum-constrained data of the simplex to the full range of linear models

in the real space by defining the centred log-ratio and its inverse

$$CLR(z) = \log\left[\frac{z_1}{g(z)}, \ldots, \frac{z_D}{g(z)}\right] \tag{10}$$

$$CLR^{-1}(z) = C[\exp(z_1), \ldots, \exp(z_D)] \tag{11}$$

Any strictly positive denominator is permissible in (10), but it is conventional to use the geometric mean of the row composition. It is clear from (10) that a model of $CLR(z)$ is a model of relative values.

To illustrate the effect of centreing compositions and calculating a linear model after transformation by $CLR()$, Figure 4 shows the Swedish data, the centred form, and its first principal component.

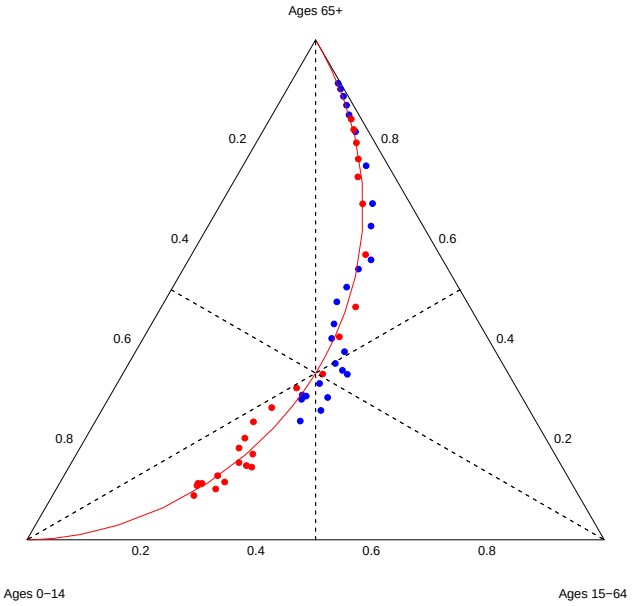

Figure 4: Compositional centreing and the first Principal Component: original data (blue), centred data (red)

It can be seen that one principal component is sufficient to capture most of the variation and that the apparent non-linearity of the data on the simplex is well represented by a straight line in the real space.

Compositional regression is also possible, but this paper will make use of the singular value decomposition (SVD) of $CLR(z)$ which allows one to select r singular vectors as the best rank-r least-squares approximation to $CLR(z)$(Aitchison, 2003). Using the compositional operators defined above, the full model for the rank-r approximation of a compositional data set is then

$$x_n^{(r)} = \hat{\xi} \oplus (u_{n1} s_1 \otimes \beta_1) \oplus \ldots \oplus (u_{nr} s_r \otimes \beta_r)\,, \qquad r < D \tag{12}$$

where the u are the left singular vectors, s are the singular values, and the β are the inverse of the centred log ratio transform of the transposed right singular vectors.

3 A CoDa Model for Mortality Forecasting

In this section we take as a benchmark the best known mortality forecasting procedure, the Lee-Carter model (Lee and Carter, 1992), and attempt to define its CoDa equivalent. The mortality component of the Lee-Carter model is defined by

$$\ln m_{x,t} = a_x + b_x k_t + \epsilon_{x,t} \tag{13}$$

where $m_{x,t}$ is the central death rate at age x and time t, and a_x is the average pattern of mortality by age. To match the life-table in each period t, this average pattern is distorted by an age-period interaction term $b_x k_t$. The temporal intensity of the distortion is controlled by k_t and its age-specific pattern by b_x. The a_x are calculated as age-specific averages of $\ln m_{x,t}$ across time. The $\ln m_{x,t}$ are then centred on a_x, and k_x and b_t are estimated by SVD as the first left and right singular vectors. To define a unique solution the b_x are required to sum to unity and the k_t to zero.

For advanced economies in the 20th century this rank-1 approximation is usually adequate and the k_t are approximately linear which allows them to be forecast with a simple ARIMA time series model. Lee and Carter used a random walk with drift.

$$k_t = k_{t-1} + d + e_t \tag{14}$$

They observed that fitting the log mortality rates ensured positivity, but emphasised the weight given to low rates. As a result they used an iterative procedure to adjust the k_t with the objective of fitting the observed annual totals of deaths D_t. In most recent applications, this has been replaced with a procedure that fits $e(0)$, life expectancy at birth (Lee and Miller, 2001).

To express the Lee-Carter structure in CoDa form the data structure and model notation follows the practice of Compositional Data Analysis. A transposed Lexis surface of $d_{x,t}$ values for one country and sex is organized with $t = 1 \dots N$ time periods as rows and $x = 1 \dots D$ age groups as columns. Exposure and deaths tables are used. Since the $d_{x,t}$ values used in the final analysis have to be strictly positive for log transformation, the oldest age column in which every cell has positive exposure is regarded as ω and older exposure is summed into the open interval starting at ω. Very light smoothing is applied to each row of the deaths table to ensure that any zero counts are replaced by a positive value. The smoothing is carried out by a histogram smoother with a second-order penalized likelihood that leaves the moments and sums unchanged.[3] As with exposure, deaths are then accumulated in the open interval. The d_x values are calculated from the M_x using the usual method employing a_x. Since the life-table d_x function is a density, each row of the table to be analyzed sums to the life table radix, assumed here to be unity, and is therefore a "composition" containing only relative information.

The construction of the CoDa equivalent to the Lee-Carter model can be summarised in the following steps:

1. Construct an $N \times D$ matrix of the $d_{x,t}$ values where $t = 1 \dots N$ and $x = 1 \dots D$ so that each row is a density of life-table deaths.
2. Centre the matrix by calculating the vector of age-specific geometric means and subtracting it from each row of the matrix using the CoDa subtraction operator.
3. Calculate the centred log-ratios of each row of the resulting matrix to transform it into the real space.
4. Calculate the SVD of the result.
5. Construct the selected low rank approximation to the centred log-ratio matrix.

[3] I should like to thank Paul Eilers and Giancarlo Camarda for their help with the R code.

6. Use the inverse of the centred log-ratio transform to convert the low rank approximation to compositional data.

7. Using the compositional addition operator, add the column-specific geometric means to the low rank compositional matrix to obtain the fitted $\hat{d}_{x,t}$ matrix.

Since we are modeling the $d_{x,t}$, or distribution of lifetimes, directly there seems to be no reason to adjust the left singular vector to match average life-expectancy or deaths in total or by age. We do not scale the singular vectors as they automatically sum to zero.

4 Single-Decrement Examples

[Note: for this summary, only the first left and right singular vectors are shown since they dominate the dynamics. Some second and third vectors are significant and can be interpreted as information about the AIDS epidemic, Gorbachev's vodka controls, data errors, the collapse of Communism in Eastern Europe, etc..]

To illustrate the model fit, Figure 5 shows the centred log ratio data and rank-2 model estimates for three years spanning the experience of French females from 1955 to 2005.

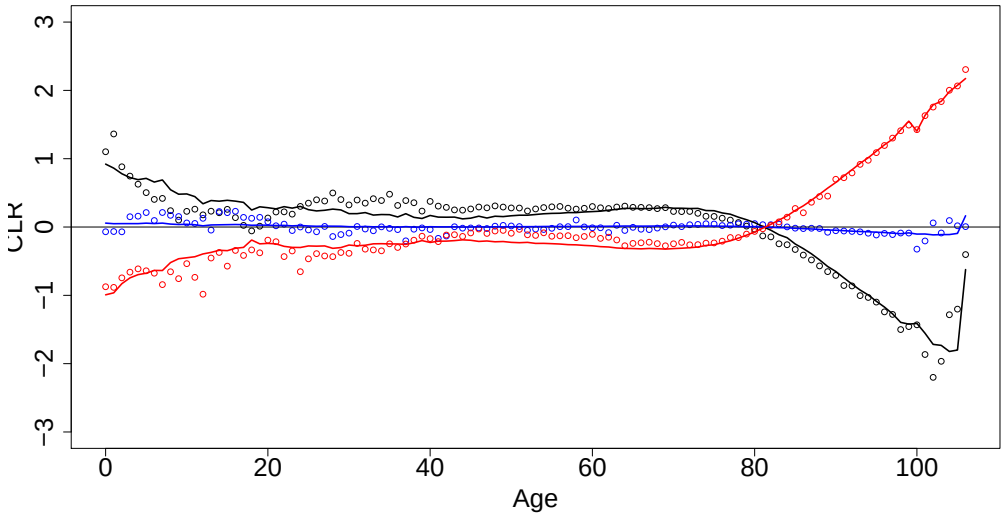

Figure 5: France, female, 1955-2005: CLR(centred d_x) (points) and rank-2 model estimates (lines) for 1955 (black), 1979 (blue), 2005 (red).

Figure 6 shows the Period and Age perturbation profiles for the same data, before the CLR^{-1} transform is applied to convert them to a composition. The Period factor shows the same linearity observed in Lee-Carter estimates and has the same interpretation. The Age perturbation has a different interpretation from that of the Lee-Carter model. When the Period factor is positive, d_x density for ages where the Age Factor is negative is transferred to ages where it is positive. For French females in the years after 1980, relative to the central life-table, lives are being saved below age 80 and transferred to higher ages.

France is a high-performance country in survival terms, but the Period Factor trend in Figure 6 is typical of the patterns for either sex in advanced economies in the HMD during the second half

of the Twentieth Century, but with some exceptions. For example, the Period factors for Japan are more curvilinear, although the age pattern is similar to France. The Age Factor profiles for all countries in the HMD are basically similar to the French female pattern with the dominant positive value at age ω, with the exception of most of the former Communist countries, the USA, and Canadian males.

To illustrate that these results are not a function of rigidity in the models, Figure 7 and Figure 8 show patterns for males in Hungary.

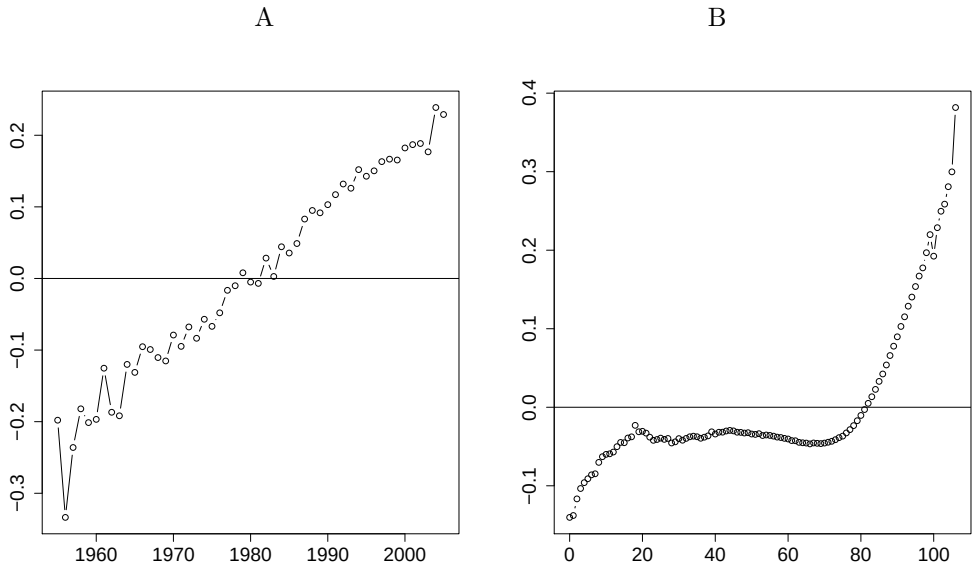

Figure 6: France, female, 1955-2005: (A) Period Factor; (B) Age Factor.

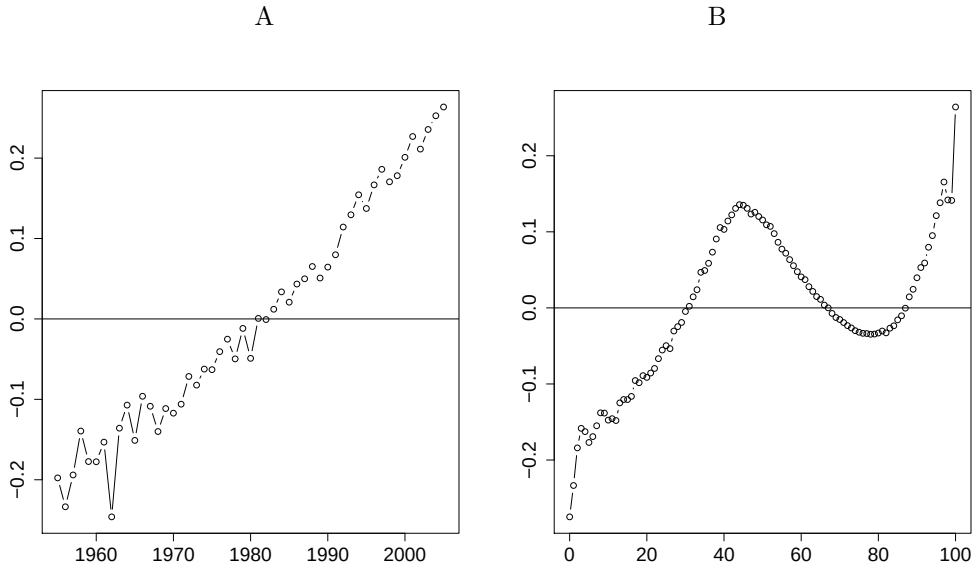

Figure 7: Hungary, male, 1955-2005: (A) Period Factor; (B) Age Factor.

The Age perturbation in Figure 7 reveals that the d_x function is tending towards bimodality in the short and medium term and thus the model is not limited to unimodal perturbations. It should also be noted that the temporal trend in Hungary is very simple when compared with the USA (see Figure 15 later), and the temporal dynamics of the survival probability distribution (l_x) in Figure 8 come from the "fixed" life-saving pattern by age, not from non-linear changes in the intensity of perturbation over time. This is a surprising result and is repeated for males in the Czech Republic, Slovakia and Slovenia, although the other former Communist states in the HMD show non-linear changes in the Period factor.

5 Exploring the Single Decrement Model

In this section, single-decrement CoDa models are compared with variants of the Lee-Carter model. The objective is to determine if the CoDa model is a plausible technique for forecasting mortality. For this purpose it is convenient to follow the structure and model abbreviations adopted by Booth et al. (2006). They used the acronym LC to indicate the original Lee-Carter model as the SVD of the log-mortality rates, with k_t adjusted to match the total observed deaths D_t. The adjusted k_t are extrapolated by a random walk with drift, and the forecast "jumps off" the fitted rates in the final year. The Lee-Miller variant (LM) (Lee and Miller, 2001) adjusts the k_t to match the observed average life expectancy at birth $e_{0,t}$ and the jump-off rates are the observed values in the final year of the fitting period. In the Booth-Maindonald-Smith variant (BMS) (Booth et al., 2002), the jump-off is from the fitted rates as in LC, but the adjustment of k_t involves fitting the observed age-specific deaths $D_{x,t}$.

For all the models, the fitting period ended in 1985 and the forecast was made for 1986-2000. The starting years for fitting vary by model. In the spirit of the original publication of the LC model, long series were used in its estimation. Eight countries start in 1900 but data for Australia and Canada are only available from 1921. For LM the starting date is 1950. The novel feature of the BMS model is that starting dates are country-specific and chosen to maximise goodness of fit to a linear k_t. The years for each country are given in Booth et al. (2006, Table 1). The lengths of the

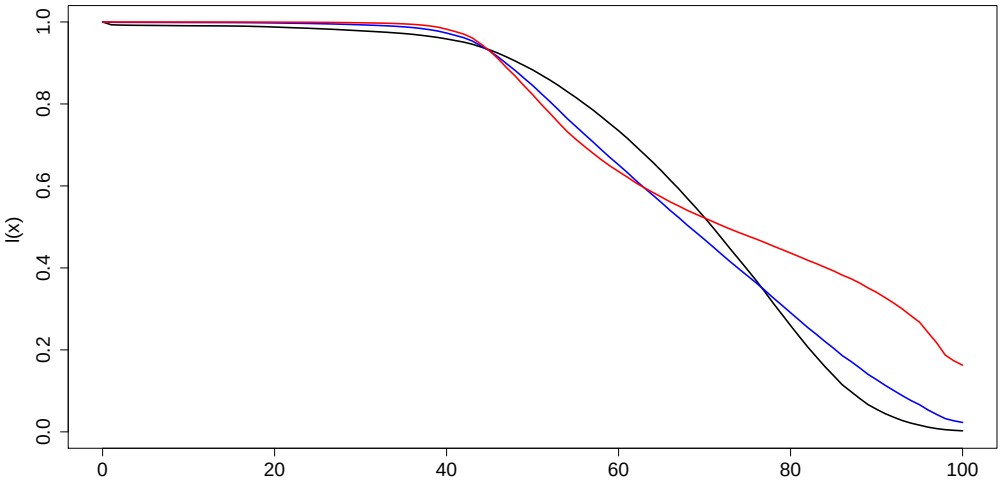

Figure 8: Hungary, male: survival probability ($\hat{l}_x$) in 2005 (black), 2050 (blue), 2100 (red).

selected fitting periods vary from 24 years to as low as 10, for a forecast horizon of 15 years.[4] An open interval was defined starting at age 95.[5]

As with the original article, the data are taken from the HMD, but the data have since been revised. In the tables that follow, the parameters are estimated from the revised data.[6] A comparison with the published tables shows only very minor differences. It should be noted that the tables presented here adopt the definition of error as (observed - forecast) which was reversed in the published tables. The interpretation of the success of the LC model over the longer time-period is complicated by the inclusion of periods of age-specific mortality that were unlikely to be relevant to the forecast period. The involvement of the countries in the two World Wars varied, which also has a sex-specific effect, and the data for all but Australia and Canada include the 1918 influenza pandemic.

In the tables that follow, each Lee-Carter variant (LC, LM and BMS) is followed by the CoDa results estimated from the same data. There are two additional columns ":unadj" and ":actual". The first shows the effect on the LC model of not adjusting the k_t to match D_t. The second uses the actual final life-table as the jump-off point for the forecast rather than the fitted values.

Table 1: **Mean error in *male* log death rates by model.**

Country	**LC**	**:unadj.**	**:actual**	**CoDa**	**LM**	**CoDa**	**BMS**	**CoDa**
Australia	0.24	-0.16	-0.06	-0.10	-0.10	-0.09	-0.04	-0.01
Canada	0.13	-0.09	-0.02	-0.04	-0.04	-0.04	0.06	0.07
Denmark	-0.04	-0.03	-0.06	-0.03	-0.12	-0.13	-0.11	-0.12
Eng & Wales	0.28	-0.06	0.02	0.00	-0.04	-0.03	-0.03	-0.01
Finland	0.24	-0.09	-0.02	-0.03	-0.01	0.00	0.04	0.04
France	0.19	-0.11	-0.03	-0.03	-0.08	-0.07	-0.07	-0.04
Italy	0.06	-0.02	0.00	0.01	0.00	-0.01	0.03	0.03
Norway	-0.17	-0.05	-0.07	-0.06	-0.10	-0.12	-0.11	-0.09
Sweden	0.09	-0.03	-0.02	-0.02	-0.06	-0.08	0.01	0.05
Switzerland	0.12	-0.01	0.01	0.05	-0.02	-0.01	-0.02	0.01
Average	0.11	-0.06	-0.02	-0.03	-0.06	-0.06	-0.02	-0.01

Table 2: **Mean error in *female* log death rates by model.**

Country	**LC**	**:unadj.**	**:actual**	**CoDa**	**LM**	**CoDa**	**BMS**	**CoDa**
Australia	0.16	-0.09	-0.02	-0.02	-0.05	-0.02	-0.01	0.05
Canada	0.24	0.02	0.08	0.09	0.03	0.09	0.07	0.07
Denmark	0.35	-0.03	0.03	0.04	-0.05	-0.01	-0.04	-0.05
Eng & Wales	0.20	0.00	0.04	0.06	0.00	0.03	-0.02	0.03
Finland	0.68	0.04	0.17	0.15	0.16	0.22	0.17	0.23
France	0.26	-0.07	0.01	0.05	-0.02	0.02	-0.03	0.02
Italy	0.24	0.00	0.05	0.10	0.06	0.09	0.07	0.10
Norway	0.56	-0.05	0.06	0.02	-0.01	-0.01	0.03	-0.03
Sweden	0.61	-0.05	0.07	0.05	0.00	0.06	0.04	0.01
Switzerland	0.43	-0.04	0.07	0.12	0.02	0.07	0.02	0.06
Average	0.37	-0.03	0.06	0.06	0.01	0.05	0.03	0.05

For the long time-series in the LC data-set, the CoDa results are remarkably similar to "LC:actual", the variant with the actual jump-off life-table. For the shorter periods, there is no clear winner but

[4]We presume that the search took place over the interval 1950 to 1985.

[5]It is not clear how cells with zero mortality rates at young ages were handled in the error analysis. In this study, zero death totals were replaced by 0.5 to create a rate more extreme than the data.

[6]The Lee-Carter variants were estimated using the package R:demography by Hyndman, R. J. and his colleagues, 2006.

Table 3: Mean absolute error in *male* log death rates by model.

Country	LC	:unadj.	:actual	CoDa	LM	CoDa	BMS	CoDa
Australia	0.46	0.28	0.19	0.20	0.18	0.17	0.13	0.13
Canada	0.30	0.19	0.12	0.12	0.11	0.11	0.12	0.12
Denmark	0.19	0.19	0.18	0.18	0.21	0.21	0.18	0.21
Eng & Wales	0.44	0.28	0.19	0.18	0.14	0.14	0.12	0.12
Finland	0.44	0.29	0.24	0.23	0.20	0.20	0.19	0.21
France	0.35	0.30	0.14	0.14	0.13	0.12	0.12	0.11
Italy	0.25	0.23	0.18	0.18	0.17	0.17	0.18	0.16
Norway	0.23	0.22	0.19	0.18	0.20	0.20	0.18	0.18
Sweden	0.24	0.19	0.18	0.18	0.20	0.20	0.16	0.19
Switzerland	0.25	0.23	0.18	0.18	0.18	0.18	0.16	0.17
Average	0.31	0.24	0.18	0.18	0.17	0.17	0.15	0.16

Table 4: Mean absolute error in *female* log death rates by model.

Country	LC	:unadj.	:actual	CoDa	LM	CoDa	BMS	CoDa
Australia	0.30	0.24	0.17	0.16	0.15	0.14	0.12	0.15
Canada	0.26	0.16	0.12	0.13	0.10	0.13	0.12	0.13
Denmark	0.42	0.21	0.22	0.22	0.24	0.24	0.21	0.24
Eng & Wales	0.26	0.14	0.13	0.13	0.12	0.12	0.10	0.12
Finland	0.76	0.33	0.31	0.29	0.27	0.30	0.26	0.32
France	0.36	0.22	0.12	0.14	0.11	0.11	0.10	0.11
Italy	0.34	0.23	0.17	0.18	0.15	0.16	0.15	0.15
Norway	0.64	0.26	0.23	0.22	0.20	0.20	0.18	0.20
Sweden	0.67	0.25	0.20	0.20	0.18	0.20	0.18	0.19
Switzerland	0.50	0.29	0.22	0.24	0.20	0.22	0.18	0.22
Average	0.45	0.23	0.19	0.19	0.17	0.18	0.16	0.18

the CoDa models can be regarded as viable alternatives. This applies equally to the error measures using log mortality-rates, which are not subject to minimisation in the CoDa SVD. There seems to be no penalty for the lack of an adjustment procedure in the CoDa models.

Although these errors measured over a short forecasting horizon are reassuring for the CoDa models, fundamental differences do arise in longer-term forecasts. Figure 9 shows the Lee-Miller and CoDa forecasts for female life-expectancy in England and Wales.[7] The extrapolated "Best Practice" line estimated by Oeppen and Vaupel (2002) is provided as a benchmark. A similar pattern can be seen for France in Figure 10.

The pattern of "optimism" shown in the CoDa forecasts seems to be a general feature for countries with advanced economies and high life-expectancy. Across the female data for the ten countries in the tables, only Norway has approximately the same forecast life-expectancy from both models. The differences between the Lee-Miller and CoDa forecasts are less pronounced for former communist countries. This implicates high levels of survival and rapid improvement, but Lee (2000) pointed out that the $b_x.(k_t - k_{t-1})$ term in his model is constant if k_t is linear. This leads to a fixed rate of decline in log mortality i.e. each age-specific mortality rate changes at a constant exponential rate (Wilmoth, 1993). If one examines the log mortality rates derived from CoDa models, their first-differences are linear but not necessarily constant.[8] As a result, CoDa forecasts of life expectancy are generally linear for most advanced economies and do not show the asymptotic

[7] In this figure and the following one, the terminal year for fitting was the maximum available in the HMD, and the open interval started at the maximum age for which all death counts were strictly positive.

[8] This result was derived by inspection and not analytically.

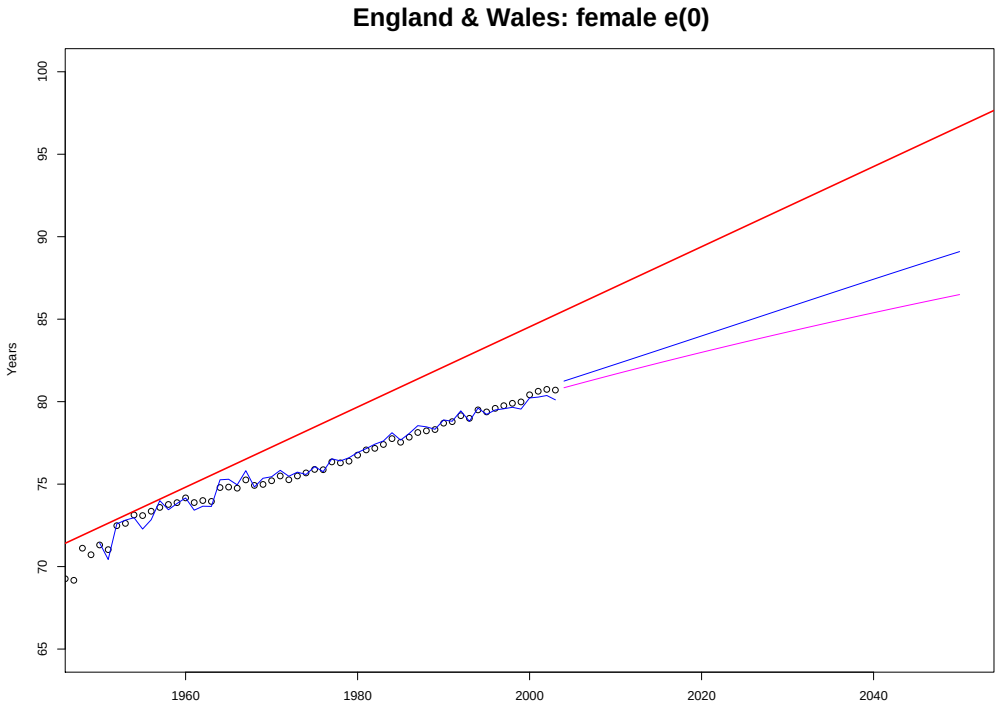

Figure 9: England and Wales, female: life expectancy at birth (circles), CoDa fitted and forecast (blue), Lee-Miller forecast (magenta), Best Practice (red).

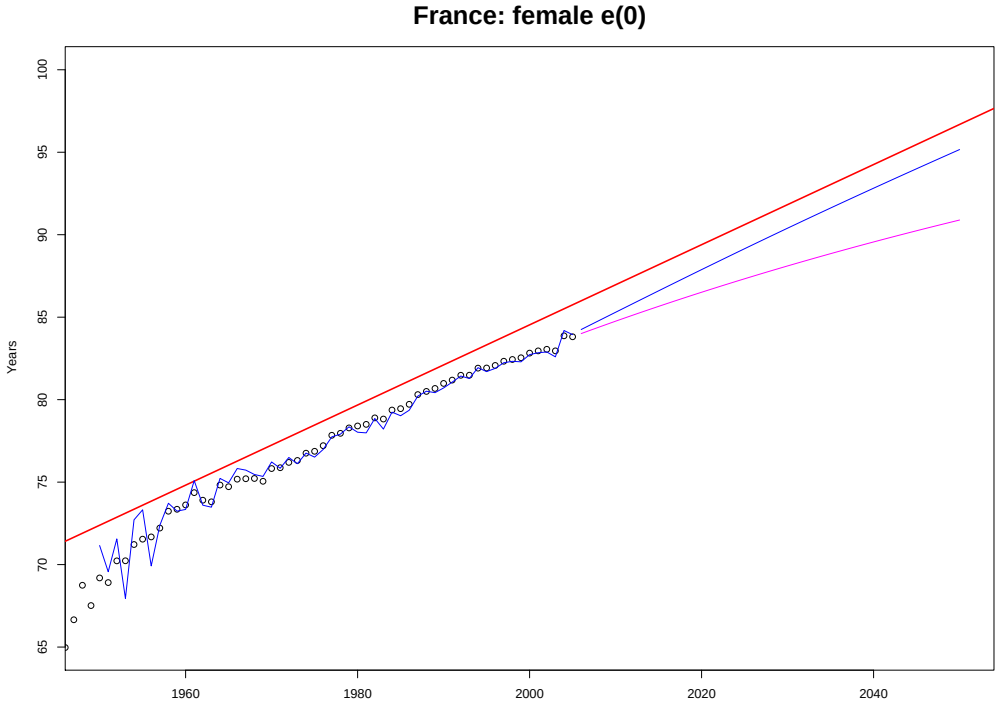

Figure 10: France, female: life expectancy at birth (circles), CoDa fitted and forecast (blue), Lee-Miller forecast (magenta), Best Practice (red).

Table 5: Mean error in *male* life expectancy by model.

Country	LC	:unadj.	:actual	CoDa	LM	CoDa	BMS	CoDa
Australia	1.07	3.41	1.62	1.85	1.55	1.41	0.63	0.31
Canada	0.75	2.08	0.87	0.95	0.73	0.70	-0.17	-0.25
Denmark	0.53	0.45	0.51	0.29	1.10	1.09	1.18	1.12
Eng & Wales	0.57	2.42	1.02	1.12	1.07	0.99	0.84	0.64
Finland	0.66	2.44	1.01	1.04	0.60	0.43	0.11	-0.02
France	0.56	2.82	0.96	0.94	1.01	0.90	0.85	0.58
Italy	1.33	1.86	1.15	0.97	1.12	1.13	0.79	0.64
Norway	1.59	0.81	1.08	0.99	1.50	1.55	1.12	1.06
Sweden	0.63	1.26	1.04	1.01	1.24	1.30	0.59	0.37
Switzerland	-0.05	1.02	0.31	0.01	0.39	0.27	0.28	0.00
Average	0.77	1.86	0.96	0.92	1.03	0.98	0.62	0.44

Table 6: Mean error in *female* life expectancy by model.

Country	LC	:unadj.	:actual	CoDa	LM	CoDa	BMS	CoDa
Australia	0.79	2.28	1.00	1.01	0.86	0.61	0.21	-0.16
Canada	-0.43	0.93	-0.24	-0.32	-0.42	-0.85	-0.40	-0.30
Denmark	-1.45	0.75	-0.50	-0.60	-0.48	-0.89	-0.40	-0.26
Eng & Wales	-0.03	1.17	0.43	0.31	0.44	0.10	0.43	0.13
Finland	-0.52	2.03	0.11	0.12	-0.47	-0.93	-0.81	-1.21
France	0.36	2.47	0.70	0.46	0.41	0.12	0.23	-0.05
Italy	0.65	2.05	0.80	0.48	0.50	0.21	0.24	0.02
Norway	-0.73	1.69	0.04	0.18	-0.02	-0.08	-0.34	-0.06
Sweden	-0.65	2.04	0.19	0.27	-0.10	-0.51	-0.13	0.06
Switzerland	-0.76	1.99	-0.11	-0.43	-0.28	-0.67	-0.51	-0.84
Average	-0.28	1.74	0.24	0.15	0.04	-0.29	-0.15	-0.27

behaviour of the Lee-Carter variants.

6 Exploring the Multiple-Decrement Model

The intention of specifying a CoDa model was to allow a very simple extension to the multiple-decrement case. If we have K causes of death, then the $N \times D \times K$ array of $d^i_{x,t}$ values is "unfolded" to an $N \times DK$ matrix. The rows are still compositions and sum to unity and the SVD proceeds as before, after centring on the $D \times K$ vector of the geometric means of the columns. The left singular vectors control the temporal dynamics and their interpretation is unchanged. The elements of the right singular vectors are now defined by pairs of ages and causes. As in the single-decrement case they indicate which columns are losing and gaining deaths, but the gains and losses need not be balanced within causes or within ages. In the 20th century, we would imagine that infectious causes would lose deaths at most ages, whereas cancer might gain deaths at most adult ages.

The death counts for Japan were downloaded from the Berkeley Mortality Database (2008) (BMD) and cover the years $1951 \ldots (1) \ldots 1990$ and ages $0, 1-4, 5-9, \ldots, 95-99, 100+$. Totals for unknown ages at death were distributed proportionately. Causes of death are disaggregated into 40 categories, but it was felt that this was too many for initial exploration. A rather arbitrary categorisation of the 40 into 6 groups was used: Infectious (2:5, 8:9), Malignant Neoplasm (11:21), Heart Disease (23:25), Cerebrovascular (27:29), Respiratory (6:7, 30:31), and Miscellaneous (1, 10, 22, 32:40). The HMD $d_{x,t}$ data for Japan were aggregated to match the BMD age groups and

Table 7: Mean absolute error in *male* life expectancy by model.

Country	LC	:unadj.	:actual	CoDa	LM	CoDa	BMS	CoDa
Australia	1.18	3.41	1.62	1.85	1.55	1.41	0.63	0.32
Canada	0.79	2.08	0.87	0.95	0.73	0.70	0.19	0.25
Denmark	0.53	0.45	0.51	0.32	1.10	1.09	1.18	1.12
Eng & Wales	0.70	2.42	1.02	1.12	1.07	0.99	0.84	0.64
Finland	0.84	2.44	1.01	1.04	0.62	0.47	0.27	0.20
France	0.63	2.82	0.96	0.94	1.01	0.90	0.85	0.58
Italy	1.33	1.86	1.15	0.97	1.12	1.13	0.79	0.64
Norway	1.59	0.96	1.12	1.04	1.51	1.56	1.15	1.10
Sweden	0.79	1.26	1.04	1.01	1.24	1.30	0.61	0.40
Switzerland	0.49	1.02	0.44	0.31	0.49	0.40	0.40	0.29
Average	0.89	1.87	0.97	0.96	1.04	1.00	0.69	0.55

Table 8: Mean absolute error in *female* life expectancy by model.

Country	LC	:unadj.	:actual	CoDa	LM	CoDa	BMS	CoDa
Australia	0.79	2.28	1.00	1.01	0.86	0.61	0.23	0.29
Canada	0.43	0.93	0.25	0.32	0.42	0.85	0.40	0.31
Denmark	1.45	0.75	0.50	0.60	0.49	0.89	0.40	0.27
Eng & Wales	0.19	1.17	0.43	0.31	0.44	0.13	0.43	0.14
Finland	0.55	2.03	0.30	0.29	0.48	0.93	0.82	1.21
France	0.41	2.47	0.70	0.46	0.41	0.20	0.23	0.24
Italy	0.66	2.05	0.80	0.48	0.50	0.21	0.24	0.10
Norway	0.73	1.69	0.25	0.32	0.21	0.18	0.34	0.18
Sweden	0.65	2.04	0.24	0.29	0.16	0.52	0.17	0.14
Switzerland	0.76	1.99	0.14	0.44	0.30	0.68	0.51	0.84
Average	0.66	1.74	0.46	0.45	0.43	0.52	0.38	0.37

the three-dimensional array was constructed by using Equation 10.31 from Elandt-Johnson and Johnson (1980).

$$ {}_n d^i_{x,t} = {}_n d_{x,t} \frac{{}_n D^i_{x,t}}{{}_n D_{x,t}} \tag{15} $$

Conditioning on a fixed age x removes the usual non-stationarity problems that would be caused by using $D_{x,t}$.

To illustrate the fit of the model it is convenient to calculate $l^i_{x^*}$, the number of persons aged x^* who will eventually leave the table from cause i: details are given on p. 76 of Preston et al. (2001).

$$ l^i_{x^*} = \sum_{x=x^*}^{\infty} {}_n d^i_x \tag{16} $$

Figure 11 shows the period life-table probability at birth that a Japanese female will die of a specific cause. The open circles mark the data used to fit the model and the lines show that a rank-2 approximation is very good. The solid points from 1999 to 2005 are outside the fitting period and represent the three leading causes of death in Japan. They are taken from calculations by the Japanese Ministry of Health, Labour and Welfare.[9] The fit between the forecast and the post 1999 data for malignant neoplasms and cerebrovascular disease is reasonable, but the forecast probability for heart disease is so much higher that one might hope that the definitions are different. Unfortunately, the source does not specify the causes under each broader heading.

[9] http://www.mhlw.go.jp/english/database/db-hw/index.html

The cerebrovascular category has two dominant groups with very different dynamics, so combining them probably represents a sub-optimal model.

The life-saving interpretation gives us new insights into the change in mortality by cause. To illustrate this, Figure 12 shows the components of the first right singular vector that represent infectious diseases and malignant neoplasms by age. The latter cause gains deaths at all ages when the left-singular vector is positive, whereas infectious diseases have become less important over most of the age-range.

7 Comparing the Single and Multiple Decrement Models

In both cases, identification of the ARIMA model suggested that a random walk with drift is inadequate. Using the AICc criterion (Shumway and Stoffer, 2000), ARIMA(0,2,2) was selected. The first two singular vectors of the single-decrement model for females in Japan, 1951-1990, capture 95.3 and 3.6% of the variance in the centred matrix. For the multiple-decrement model the figures are 91.2 and 4.1%. Although the contribution of the second vector seems to be small, it is necessary to recreate the curvature in the cerebrovascular group (see Figure 11). Figure 13 shows the first and second left-singular vectors. This indicates that disaggregating by cause does not introduce extra dynamics to the model. If this indicates that there are effectively just two fundamental temporal processes driving all causes then it is an important result.

The life expectancy forecasts for both models are shown in Figure 14. It is clear that we cannot regard the multiple-decrement model as pessimistic with respect to the single-decrement version. Wilmoth (1995) reported differences of 5 and 10 years by 2020 and 2050, respectively. Preliminary

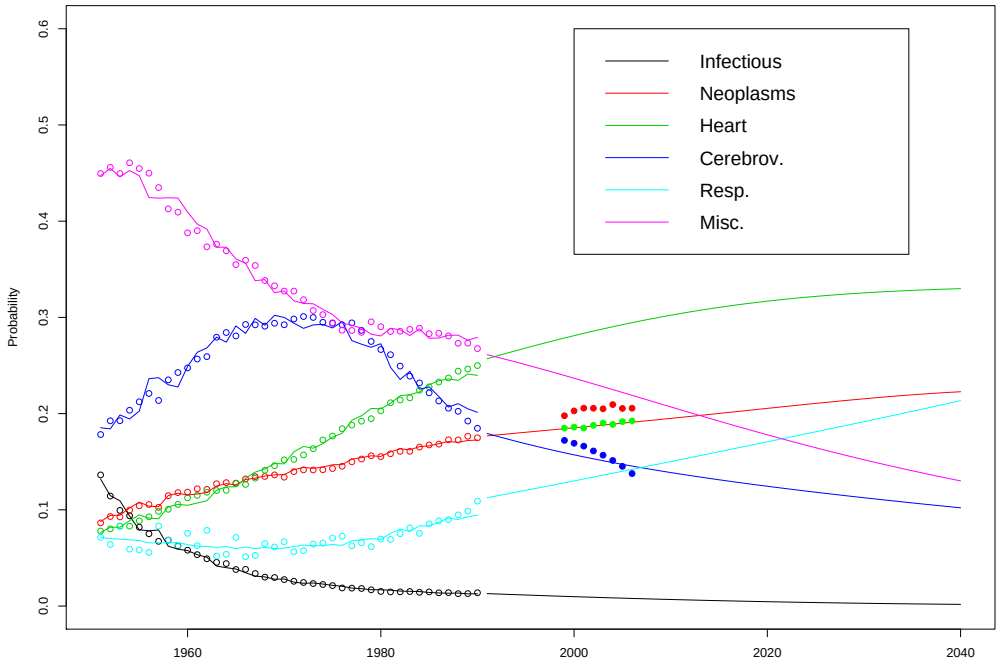

Figure 11: Japan, female period life-table probability at birth of dying from a specific cause: points = data, lines = estimates.

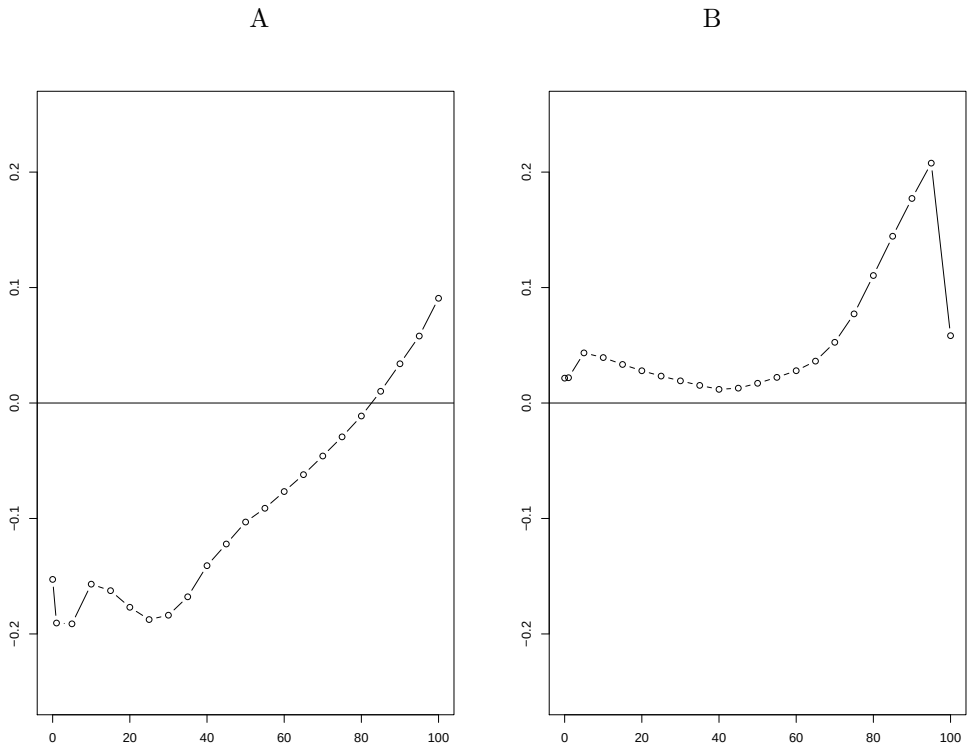

Figure 12: Japan, female, 1951-1990, right-singular vector components for: (A) Infectious Diseases; (B) Malignant Neoplasms.

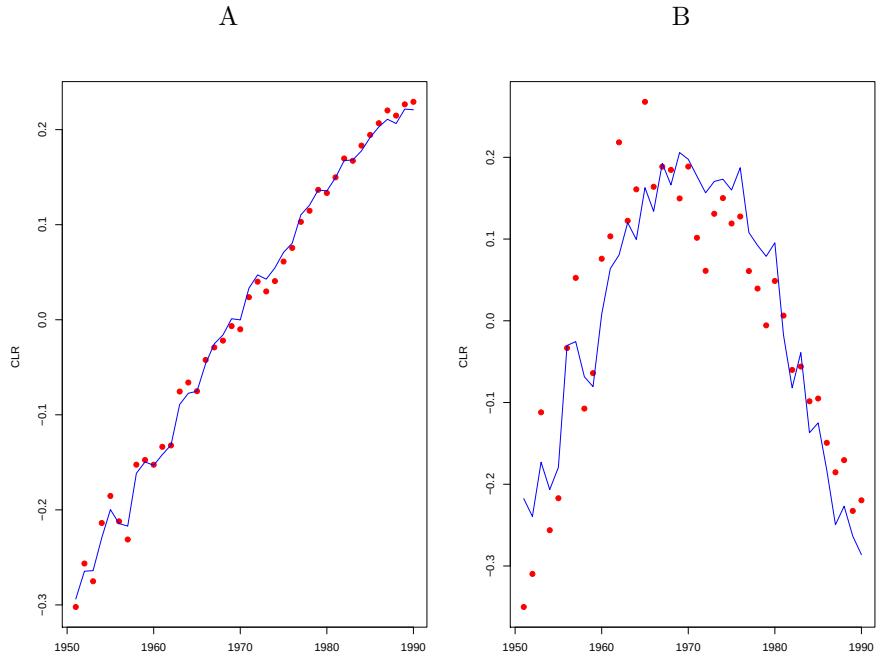

Figure 13: Japan, female, 1951-1990: (A) first left-singular vector; (B) second left-singular vector. Symbols: single decrement (red); multiple decrement (blue).

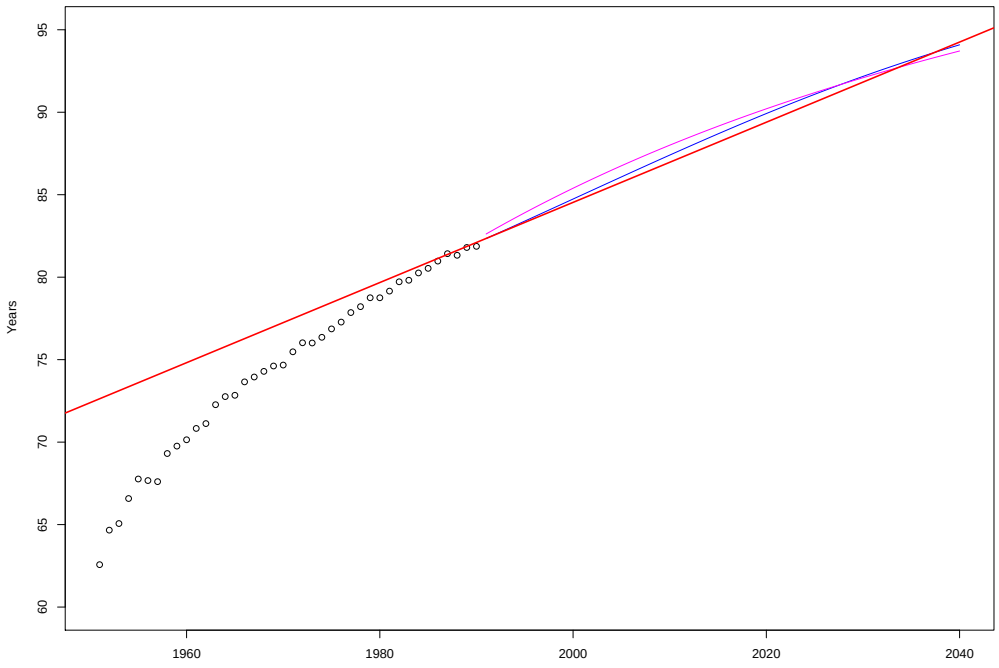

Figure 14: Japan, female, 1951-1990: life expectancy at birth (circles), single decrement forecast (magenta), multiple decrement forecast (blue), "best practice" (red).

results, not presented here, suggest that this coherence between the two models is also true for France and the Netherlands.

8 Discussion

8.1 Data quality

Mortality by cause of death is thought to be particularly inaccurate at high ages, partly because of misdiagnosis and lack of autopsy confirmation, and partly because multiple causes of death are often ascribed. Secondly, deaths are often tabulated by 5 year age groups and the open age interval into which the deaths of the oldest-old are aggregated is often defined at a relatively young age such as 85. Unfortunately, it is at these high ages where most of the temporal dynamics are occurring. This makes it unlikely that a single-decrement forecast aggregated up from a multiple-decrement forecast should be considered as being better. It is more realistic to regard the multiple-decrement forecast as being special-purpose and subject to additional sources of uncertainty, although it may be revealing as to why higher-rank approximations may be required in the single-decrement forecast. In the Japanese case, the rank 2 component seems to be driven by non-linear changes in the components of cerebrovascular mortality.

8.2 Model complexity

The separation of temporal and fixed age-cause-specific effects in the Lee-Carter structure means that forecasting the multiple-decrement CoDa version is not penalised by having a large number of causes and ages. However, it is not clear whether one should use all available causes and aggregate after modeling, or select groups on the basis of disease similarity or common dynamics. One constraint is imposed by the requirement that compositions should be strictly positive.

8.3 Combining forecasts

The CoDa model presented above aims to make a unified forecast of $\hat{d}^i_x$ whose aggregation across causes can be used as a single-decrement forecast. It is also possible to combine single- and multiple-decrement models so that the the single-decrement forecast is maintained. Assume that we have forecasts of ${}_n\hat{m}_{x,t}$, perhaps from a Lee-Carter model, and ${}_n\hat{d}^i_x$, from a CoDa model. They can be combined using (Elandt-Johnson and Johnson, 1980, Eqn. 10.9) as

$$ {}_n\hat{m}^i_{x,t} = {}_n\hat{m}_{x,t} \frac{{}_n\hat{d}^i_{x,t}}{{}_n\hat{d}_{x,t}} \tag{17} $$

This guarantees that the multiple-decrement forecast ${}_n\hat{m}^i_{x,t}$ is consistent with the single-decrement model ${}_n\hat{m}_{x,t}$, but does not of course resolve the issue of why the forecast of $\hat{d}^i_x$ was not sufficient on its own.

8.4 Rectangularisation of the survival curve

The CoDa model offers a very simple test of long-run rectangularisation - the growing tendency in advanced economies for the survival probability to remain close to one and then collapse quickly to zero at high ages. If rectangularisation exists and mortality becomes compressed as Fries (1980) suggested, then the first right-singular vector should show a peak at an age "interior" to the age range. For France in Figure 6, single positive and negative regions are divided at circa age 80 and

the scale of this process in the positive limb of the Age Factor increases with age, showing that if the trend in the Period Factor continued indefinitely, all deaths would occur at $\omega = 106$. In 2005 only 0.19% occurred at this age. A random-walk with drift projection of the Period Factor would lead to 15% dying at the maximum age in 2050, but by 2100 this would have risen to 86%. Thus there is no evidence from the model and these data for the combination of Period and Age Factor patterns that might indicate a long-term rectangularisation process below the age-limit of the life-table.

Some countries do suggest rectangularisation. Figure 15 shows the patterns for females in the USA. As expected, there is a temporal trend in $d_{x,t}$ perturbation, but the process is non-linear in time

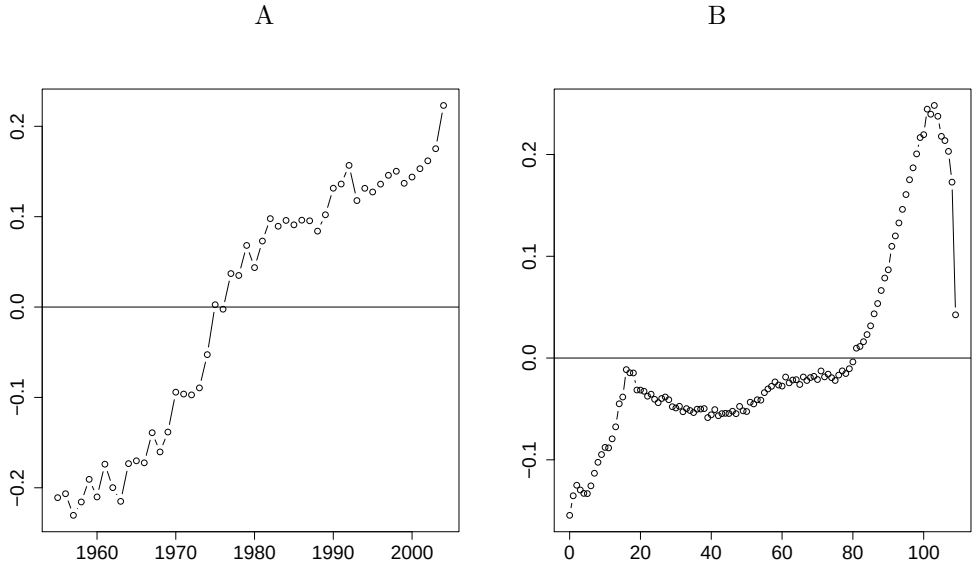

Figure 15: USA, female, 1955-2004: (A) Period Factor; (B) Age Factor.

and has short periods of stagnation centred on 1960 and 1985. The Age Factor does show evidence of ultimate rectangularisation, but at a modal age of 103, well above the age of 85 suggested by Fries (1980). For males in both the USA and in Canada, the limit modal age is 101 years with ω equal to 109 and 108 respectively. Figure 16 shows the projected dynamics of the USA female d_x function over the next century.

As the USA and Canada are the only advanced economies in the HMD showing rectangularisation, it is important to recall that the four-fold classification by Kannisto (1994) of the quality of old-age mortality in 32 sets of life-tables placed these two countries together with Chile and New Zealand Maori in the lowest category, entitled "Weak Quality".[10] The weight of the HMD evidence is that rectangularisation is a short-term effect in advanced economies and is not indicative of a widespread limiting process. More remarkably, an analysis of the Swedish data by 50-year periods from 1800 shows no evidence of rectangularisation as a limiting process (Oeppen, 2008).

8.5 Changing hazards versus life-saving as models of mortality

The CoDa model shifts deaths from one age to another, rather than altering mortality rates. Thus it is essentially a life-saving model in the style of Vaupel and Yashin (1986, 1987). The advantages or disadvantages of life-saving models are unknown in a forecasting context.

[10] I should like to thank Domantas Jasilionis for bringing this to my attention.

9 Conclusion

This paper suggests that the problem of coherent forecasting of multiple-decrement life tables can be approached via Compositional Data Analysis: a methodology explicitly developed for constrained data. As a test of the method, the Lee-Carter model was expressed in compositional form and proved to be as good as the Lee-Carter variants in short-horizon forecasts for ten countries. For a compositional model, the disaggregation of deaths by age into specific causes presents no additional structural problems. A brief examination of multiple-decrement fits and forecasts for Japan suggests that the results look promising, although it is clear that a great deal more testing would be required before the methodology could be regarded as established. It does not seem to be the case that the multiple-decrement compositional forecast is always more pessimistic than the single-decrement one, as has been observed in proportional change models of mortality rates.

Acknowledgments
I should like to thank Giancarlo Camarda, Iain Currie, Paul Eilers, Jutta Gampe and the members of her Statistical Laboratory at the MPIDR, Domantas Jasilionis, Nico Keilman, Trifon Missov, Mike Murphy, Roland Rau, Tiziana Torri and John Wilmoth for their help and suggestions. I am grateful to K. G. van den Boogaart who contributed the R:compositions library and to Rob Hyndman, Heather Booth, Leonie Tickle and John Maindonald who contributed the R:demography library.

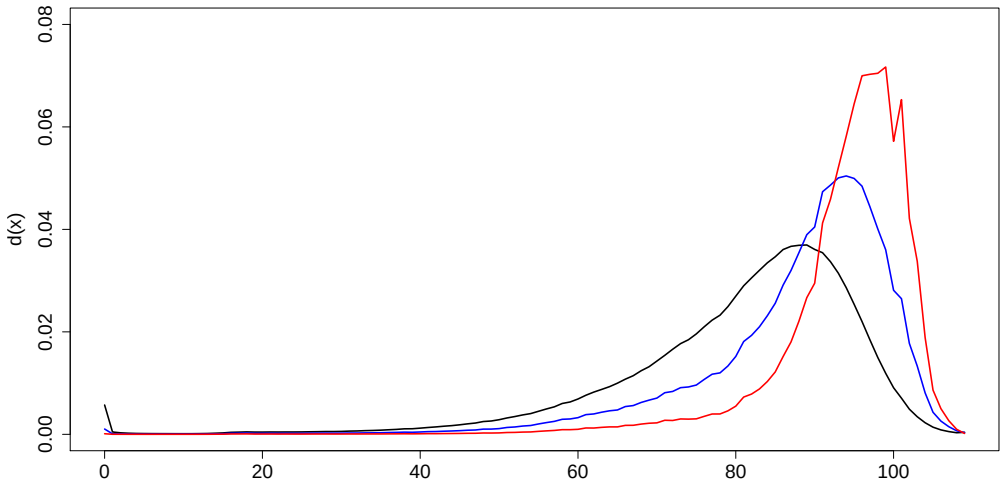

Figure 16: USA, female: $\hat{d}_x$ in 2005 (black), 2050 (blue), 2100 (red).

References

Aitchison, J. (1986). *The statistical analysis of compositional data*. London: Chapman and Hall.

Aitchison, J. (2003). A concise guide to Compositional Data Analysis.

Berkeley Mortality Database (2008). University of California, Berkeley (USA). www.demog.berkeley.edu/ bmd/.

Booth, H., R. J. Hyndman, L. Tickle, and P. de Jong (2006). Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions. *Demographic Research 15*, 289–310.

Booth, H., J. Maindonald, and L. Smith (2002). Applying Lee-Carter under conditions of variable mortality decline. *Population studies 56*(3), 325–36.

Canudas-Romo, V. (2006). The modal age at death and the shifting mortality hypothesis.

Edwards, R. D. and S. Tuljapurkar (2005). Inequality in life spans and a new perspective on mortality convergence across industrialized countries. *Population and Development Review 31*(4), 645–+.

Elandt-Johnson, R. C. and N. L. Johnson (1980). *Survival models and data analysis*. New York: Wiley.

Fries, J. F. (1980). Aging, natural death, and the compression of morbidity. *New England Journal of Medicine 303*(3), 130–135.

Human Mortality Database (2008). University of California, Berkeley (USA). Max Planck Institute for Demographic Research (Germany). www.humanmortality.de.

Kannisto, V. (1994). *Development of oldest-old mortality, 1950-1990 : evidence from 28 developed countries*. Monographs on Population Aging. Odense: Odense University Press.

Lee, R. and T. Miller (2001). Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography 38*, 537–550.

Lee, R. D. (2000). The Lee-Carter method for forecasting mortality, with various extensions and applications. *North American Actuarial Journal 4*(1), 80–91.

Lee, R. D. and L. R. Carter (1992). Modeling and forecasting United-States mortality. *Journal of the American Statistical Association 87*(419), 659–671.

Lee, R. D. and F. Nault (1993). Modeling and forecasting provincial mortality in Canada.

Li, N. and R. Lee (2005). Coherent mortality forecasts for a group of populations: An extension of the lee-carter method. *Demography 42*(3), 575–594.

Oeppen, J. (2008). Model evidence against rectangularisation as a limit to survival improvement.

Oeppen, J. and J. W. Vaupel (2002). Demography - broken limits to life expectancy. *Science 296*(5570), 1029–+.

Polder, J. J., J. J. Barendregt, and H. van Oers (2006). Health care costs in the last year of life - the Dutch experience. *Social Science and Medicine 63*(7), 1720–1731.

Preston, S. H., P. Heuveline, and M. Guillot (2001). *Demography: measuring and modeling population processes*. Malden, MA: Blackwell Publishers.

Shumway, R. H. and D. S. Stoffer (2000). *Time series analysis and its applications*. Springer texts in statistics. New York: Springer.

Vaupel, J. W. and A. I. Yashin (1986). Targeting lifesaving - demographic linkages between population-structure and life expectancy. *European Journal of Population-Revue Europeenne De Demographie 2*(3-4), 335–360.

Vaupel, J. W. and A. I. Yashin (1987). Repeated resuscitation - how lifesaving alters life-tables. *Demography 24*(1), 123–135.

Wilmoth, J. (1993). Computational methods for fitting and extrapolating the Lee-Carter model of mortality change. Technical report, University of California, Berkeley. U.S.A.

Wilmoth, J. R. (1995). Are mortality projections always more pessimistic when disaggregated by cause of death? *Mathematical Population Studies 5*(4), 293–319.