Università di Pisa

Sant'Anna Scuola Universitaria Superiore Pisa

Consiglio Nazionale delle Ricerche

Book of Short Papers SIS 2020

SIS 2020 Pisa

SIS Società Italiana di Statistica

Editors: Alessio Pollice, Nicola Salvati and Francesco Schirripa Spagnolo

Published by Pearson

www.pearson.com

ISBN 9788891910776

LI-CoD Model. From Lifespan Inequality to Causes of Death

Modello LI-CoD. Dalla Lifespan Inequality alle Cause di Morte

Andrea Nigri and Susanna Levantesi

Abstract The evolution of lifespan disparity is gaining a central role in mortality literature. From the beginning of the new millennium, its evolution has led scholars to give more emphasis to longevity and its relationship with causes-of-death evolution. Following this line of research, we propose a novel model aiming to provide the causes-of-death mortality surface, exploiting the relationship between mortality rates by cause-of-death and lifespan variability. Taking advantage of this relationship, and using data from Human Causes of Death Mortality Data Base, our model is able to reconstruct the mortality surface by causes, using a single value of lifespan inequality as input.

Abstract L'evoluzione della lifespan disparity sta acquisendo un ruolo centrale nella letteratura sulla mortalità. Dall'inizio del nuovo millennio, la sua evoluzione ha portato gli studiosi a dare maggiore enfasi alla longevità e al suo rapporto con l'evoluzione delle cause di morte. Seguendo questa linea di ricerca, proponiamo un nuovo modello che mira a fornire la superficie della mortalità per cause di morte, sfruttando la relazione tra i tassi di mortalità per causa di morte e la variabilità della durata della vita. Sfruttando questa relazione e utilizzando i dati dello Human Causes of Death Mortality Data Base, il nostro modello è in grado di ricostruire la superficie della mortalità per cause, usando come input un singolo valore della lifespan disparity.

Key words: Life expectancy, Forecasting, CoD.

Andrea Nigri
Sapienza, Viale Regina Elena 295-G, 00161 Rome, e-mail: andrea.nigri@uniroma1.it

Susanna Levantesi
Sapienza, Viale Regina Elena 295-G, 00161 Rome, e-mail: susanna.levantesi@uniroma1.it

1 Introduction

Over the last two centuries, longevity evolution has had a prominent impact on population dynamics, showing a rapid decline in mortality levels in developed countries. These improvements are driven by medical progress, better living conditions, leading to the infant mortality reduction. On the other hand, scientific achievements against chronic diseases ([8], [12]) lowered the adult age mortality after the second-world war, likewise the recent delaying of mortality at older ages ([14]). As a result, life expectancy at birth has increased over time, without any sign of an impending limit in human life boundaries ([5]). This latter behavior seems to be in lockstep with a significant decrease in lifespan variability, which could be due to different causes-of-death (CoD) composition. In light of that, the understanding of evolution of CoD composition is crucial in predicting the aging process and the population health. Concerning forecasting, cause-specific mortality is often based on predicting cause-specific death rates independently. Only a few methods incorporating dependence among causes have been suggested. An attractive alternative is to model and forecast cause-specific death distributions, rather than mortality rates, as dependence among the causes can be directly incorporated ([3]). Unfortunately, these processes are often not straight and show a lack of parsimony and interpretability. Furthermore, they are based on a rigid structure that does not taken into account the revolution of CoD pattern over time. Our paper contributes to the literature proposing a novel statistical tool able to provide the CoD mortality surface, using the relationship between CoD mortality rates and lifespan variability. In particular, given an age, our model is able to reconstruct the mortality surface by causes, using a single value of lifespan inequality as input. Thus, using data for several countries, we try to explain and forecast the remarkable changes in the transition phases that developed countries have exhibited in the period after the second world war, providing age-cause specific action.

2 Longevity indicators

The constant improvement of BPLE suggests that mortality reductions should not be viewed as a disconnected sequence of unrepeatable revolutions, but rather as a regular flow of continuous progress ([6]). Clearly, mortality is linked to social progress in terms of health, nutrition, education, hygiene, and medicine ([7]). Moreover, mortality improvements have been combined with an increase of lifespan equality in which the increase in the age-at-death corresponds to the increase of the compression of the distribution around its modal value ([4]). While life expectancy has been proven to hide heterogeneity in individual mortality courses, lifespan disparity measures

both uncertainty in the age-at-death distribution and heterogeneity[1] ([10]). In the following equations we provide the formal notation and definitions for both life expectancy and lifespan disparity.

- **Life expectancy**
  Let $S(x,t)$ and $\mu(x,t)$ be two continuous functions with respect to age x and time t, respectively representing the survival probability and the force of mortality of an individual aged x at time t in a given population. We denote $e_{x,t}$ the life expectancy at age x and time t, that is defined as follows:

$$e_{x,t} = \frac{\int_x^\infty S(y,t)dy}{S(x,t)} \tag{1}$$

  where $S(x,t) = \exp\left(-\int_0^\infty \mu(x+\xi,t+\xi)d\xi\right)$ are the survival probabilities.

- **Lifespan disparity**
  We denote $e^\dagger_{x,t}$ the lifespan disparity that is an indicator of the lifespan variation representing the life expectancy lost due to death by an individual aged x at time t ([13]). Formally, its functional form is defined as follow:

$$e^\dagger_{x,t} = -\int_x^\infty S(y,t)\ln S(y,t)dy \tag{2}$$

  The lifespan disparity at birth is: $e^\dagger_{0,t} = -\int_0^\infty S(y,t)\ln S(y,t)dy$. Lifespan disparity varies among populations and over time. It measures the dispersion in the age-at-death: when mortality is highly variable, some individuals will die at a much lower age than the expected age-at-death, contributing many lost years to life disparities; conversely, when mortality is highly concentrated around older ages or the modal age, life disparity decreases.

These two variables show latent behaviors that should be represented by incorporating in forecasting the relationship between lifespan inequality and CoD composition. Our approach aims to introduce a new perspective in the forecasting of CoD mortality surface, thus providing more accurate predictions.

3 Data and model

We use data from Human Causes of Death Mortality Data Base that provides death rates specific for ages and causes of death. This database is coded by using the international classification of diseases (ICD), providing different aggregation levels: full list, intermediate list, and shortlist. Each classification has been developed using the same criteria for all countries, ensuring homogeneity and comparability. For our aims,

[1] In addition to life disparity, other inequality measures have been proposed in literature, e.g. the Gini coefficient and the Keyfitz's entropy ([9], [11]) that appear to be linearly related and negatively correlated to life expectancy at birth ([1]).

we use the shortlist that provides the highest level of ICD aggregation, underlining 10 different macro CoD. Afterward, we perform an additional aggregation as follows:

- **Circolatory:** Heart diseases (I00-I52); Cerebrovascular diseases (G45, I60-I69); Other and unspecified disorders of the circulatory system (I70-I99)
- **Neoplasy:** Neoplasms (C00-D48)
- **Diab:** Endocrine, nutritional and metabolic diseases (E00-E90)
- **External:** External causes (V01-Y98)
- **Perinatal:** Diseases of the genitourinary system and complications of pregnancy, childbirth and puerperium (N00-O99) Certain conditions originating in the perinatal period and congenital malformations/anomalies (P00-Q99, R95)
- **Respiratory:** Acute respiratory diseases (J00-J22, U04); Other respiratory diseases (J30-J98)
- **Infectius:** Certain infectious diseases (A00-B99)
- **Digestive:** Diseases of the digestive system (K00-K93)
- **Other:** Diseases of the skin and subcutaneous tissue, musculoskeletal system and connective tissue (L00-M99); Diseases of the blood and blood-forming organs (D50-D89); Mental and behavioral disorders (F00-F99); Diseases of the nervous system and the sense organs (G00-G44, G47-H95)

Let x be the age, t the year and c a specific CoD. Given a certain age x, the aim of the model is to convert a value of lifespan inequality $e^{\dagger}_{0,t}$ into a list of CoD specific rates $m_{c,t}$ as follows:

$$log(\hat{m}_{c,t}) = \beta_c log(e^{\dagger}_{0,t}) + \delta_c \gamma + \varepsilon_c \tag{1}$$

Where: β_c is CoD-specific pattern of human mortality, δ_c the correction of mortality improvement over CoD, γ the parameter to be optimize and ε_c are the errors such that $\varepsilon_c \sim \mathcal{N}(\mu, \sigma^2)$. In the numerical experiment we use the following input data: $x \in \{0, ..., 100\}$, causes: $c \in \{Circ, Diab, ..., Resp\}$, year: $t \in \{1999, ..., 2013\}$.

The projected values of lifespan inequality and total mortality rate(m_t), which constitute the model's input, can be given by a certain extrapolation method or be the target values resulted by official forecasting (e.g. WHO).

In order to obtain the desired CoD mortality rates we define the following steps, starting from $m_{c,t}$ for a given age x, $e^{\dagger}_{0,t}$ and m_t as input:

- We estimate the slope, β_c, of the linear relation between the logarithmic transformation of lifespan inequality and the CoD specific rates over the observation time t. This is done by using the method of the least-squares, by minimizing the sum of squared residuals: $min \sum_c [log(m_{c,t}) - \beta_c log(e^{\dagger}_{0,t})]^2$;
- We estimate the parameter δ_c by computing the singular value decomposition (SVD) of the matrix of regression residuals, R, obtained in the previous step: $SVD[R] = PDQ^T$. Where D and Q are matrices of left and right singular vectors, and P is a diagonal matrix with singular values along the diagonal. The first term of the SVD, is used for obtaining the estimates of δ_c that can be interpreted as the adjustment of mortality improvement over CoD;
- We compute the mortality rates: $\hat{m}_{c,t} = exp\{\beta_c log(e^{\dagger}_{0,t}) + \delta_c \gamma\}$, where $\gamma = 0$

- Give the total level of mortality, m_t, we optimize the CoD mortality rates finding the optimal value of γ where $\sum_c |\hat{m}_{c,t} - m_t| = 0$.

4 Results

The model is applied to the USA data from 1999 to 2013. Figure 1 shows the model fitting with 95% confidence interval in the year 1999. In order to verify the model's

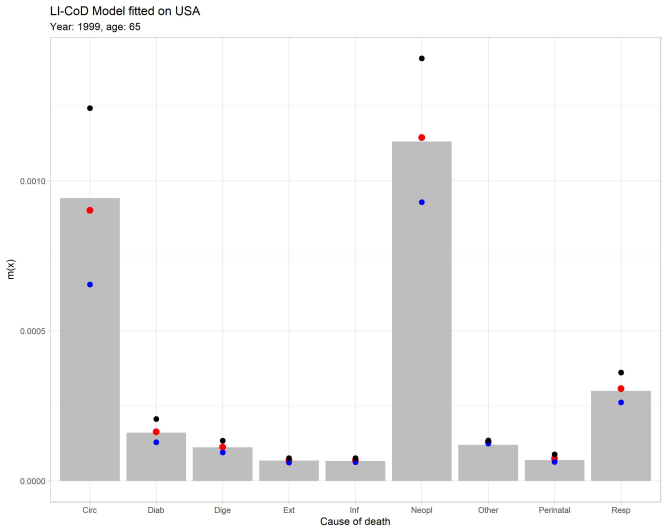

Fig. 1: Model fitting. Predicted values (black dots) and 95% confidence intervals (red and blue dots). Country USA. Year 1999.

accuracy, we perform a backtesting setting the fitting period to 1999-2010 and the forecasting period to 2011-2013. The forecasting results given by the backtesting are compared by the root mean squared error (RMSE) and the mean average error (MAE). Both the values of RMSE and MAE are very low, highlighting the accuracy of the model in predicting CoD mortality rates.

5 Conclusion

In this paper, we provide an innovative model to forecast CoD mortality, able to catch the hidden pattern and the relationship between two summary demographic measures, lifespan inequality and age-specific CoD mortality rates. Our model can have an important application in the context of incomplete data, when the official

Table 1: RMSE and MAE by CoD. Country USA. Years 2011-2013.

CoD	RMSE	MAE
Circolatory	0.000038511	0.000036751
Neoplasy	0.000026127	0.000024954
Diab	0.000007973	0.000007960
External	0.000001244	0.000001218
Perinatal	0.000006181	0.000006170
Respiratory	0.000005397	0.000004857
Infectius	0.000004339	0.000004266
Digestive	0.000002565	0.000002045
Other	0.000007631	0.000006738

registries just provide summary measures and incomplete data. We also suggest further development, modeling the CoD evolution using a Markov chain framework.

References

1. Aburto, J. M., Villavicencio, F., Basellini, U., Kjærgaard, S., and Vaupel J. W. (2020). Dynamics of life expectancy and life span equality. PNAS: 117(10): 5250–5259
2. Human Mortality Database 2018. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Data downloaded on 01/01/2019. `https://www.mortality.org`.
3. Kjeargaard, S., Ergemen, Y., Kallestrup Lamb, M., Oeppen, J., Lindahl Jacobsen, R.: Forecasting causes of death by using compositional data analysis: the case of cancer deaths. Journal of the Royal Statistical Society, Series C **68**,5, 1351-1370 (2019)
4. Kannisto, V.: Measuring the Compression of Mortality. Demographic Research **3**, article 6 (2000)
5. Oeppen, J., Vaupel, J.W.: Broken Limits to Life Expectancy. Science **296** (5570): 1029–1031 (2002)
6. Oeppen, J., Vaupel, J. W.: The Linear Rise in the Number of Our Days. Social Insurance Studies, 3. The Linear Rise in Life Expectancy: History and Prospects (2006)
7. Riley J. Rising Life Expectancy: A Global History. Cambridge. Cambridge University Press (2001)
8. Rau, R., Soroko, E., Jasilionis, D., Vaupel, J. W.: Continued Reductions in Mortality at Advanced Ages. Population and Development Review, **34**: 747-768 (2008)
9. Shkolnikov, V.M., Andreev, E.M., Begun, A.Z.: Gini coefficient as a life table function: Computation from discrete data, decomposition of differences and empirical examples. Demographic Research **8** (article 11): 305–358 (2003)
10. van Raalte, A., Sasson, I., Martikainen, P.: The case for monitoring life-span inequality. Science, 1002-1004 (2018)
11. van Raalte, A.A., Caswell, H.: Perturbation analysis of indices of lifespan variability. Demography, **50**: 1615–1640 (2013)
12. Vaupel J. W.: The remarkable improvements in survival at older ages. Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences **352**: 1799–1804 (1997)
13. Vaupel J. W.: How Change in Age-Specific Mortality Affects Life Expectancy. Population Studies **40**: 147–157 (1986)
14. Zuo, W., Jiang, S., Guo, Z., Feldman, M.W., and Tuljapurkar, S.: Advancing front of old-age human survival. PNAS **115** (44) 11209-11214 (2018)