Forecasting vital rates from demographic summary measures

Giancarlo Camarda[a] and José Manuel Aburto[b]

July 1, 2019

Abstract

[a] Institut national d'études démographiques. E-Mail: carlo-giovanni.camarda@ined.fr
[b] Interdisciplinary Centre on Population Dynamis University of Southern Denmark & Max Planck Institute for Demographic Research E-Mail: jmaburto@sdu.dk

1 Introduction

Future levels of mortality and fertility can be predicted by modelling and extrapolating rates over age and time, or by forecasting summary measures of each phenomena and then convert to age-specific rates. For example, in developed countries, the linear increase in life expectancy over some periods has made it easier to fit trends over time of this summary indicator than fitting more complex models based on age-specific dynamics of mortality (White 2002). Therefore, several methods have been proposed to forecast life expectancy. For instance, Torri and Vaupel (2012) forecast life expectancy for a given country assuming a tendency towards a predicted best practice life expectancy. Pascariu et al. (2018) proposed incorporating the analysis of the gap between female and male life expectancy to more accurately predict the overall level of mortality in a country. Similarly, Raftery et al. (2013) forecast life expectancy for several countries using a Bayesian hierarchical model for females, and then model the sex gap to estimate male life expectancy (Raftery et al. 2014). Subsequently, the overall level of mortality given by life expectancy is converted to a age-specific profile (Ševčíková et al. 2016). This latter method has been adopted by the United Nations. However, life expectancy, as an average, conceals the variation in the age-at-death distribution (van Raalte et al. 2018). While the sources of variance in lifespans, or the health inequalities it reflects, are not fully understood, they however are a key problem for policy as well for modelling and forecasting mortality (Tuljapurkar and Edwards 2011). Recently, Bohk-Ewald et al. (2017) proposed to incorporate the variation in ages at death as an additional indicator to evaluate mortality forecast. This variation is often called lifespan variation or lifespan inequality and refers to how similar ages at death are in a population. They found that some methods struggle to account for trends in lifespan variation, which results in a mismatch between life expectancy and lifespan variation. In most countries, life expectancy and lifespan variation are often negatively correlated (Smits and Monden 2009; Vaupel et al. 2011; Colchero et al. 2016; Alvarez et al. 2019; Gonzaga et al. 2018), however in some countries this association is less strong when looking at first differences over time. For example, in Central and Eastern European countries and in some Latin American countries life expectancy and lifespan variation moved independently from each other in periods where life expectancy

improvements slowed-down (Aburto and van Raalte 2018; Aburto and Beltrán-Sánchez 2019; García and Aburto 2019). This pattern is also more frequent in recent decades in low mortality countries (Aburto et al. 2019). Therefore, incorporating the dynamics of both life expectancy and lifespan variation to obtain an age-specific mortality profile that matches both is a step forward on more accurately predicting future longevity and the mechanisms underpinning new patterns in mortality.

In fertility forecasting, the challenge of accurately predicting levels and age-specific fertility dynamics rises from the complex association between structural changes (e.g. trajectory of total fertility) and changing age patterns (tempo) (Booth 2006). In general, forecasts of completed fertility aim to predict the number of children by women in reproductive age. There are two big strands in fertility forecasting. (1) Cohort fertility, which is informative on what a cohort of women experience. Regarding (1), Bohk-Ewald et al. (2018) compared the performance of 20 major methods, including parametric curve fitting methods, extrapolation, Bayesian approaches and context-specific methods, aimed at completing lifetime fertility of women that have not yet reached their last reproductive age. The authors found that more complex methods do not necessarily outperform simpler methods. The second strand is period fertility, which summarizes fertility within a period and central for our paper(Bohk-Ewald et al. 2018). As in mortality forecasting, there is the case that total fertility or mean age at childbearing are forecasted (Miller 1986), but then there is the challenge of getting age-specific fertility rates consistent with those forecasts. Lee (1993) modelled age-specific fertility rates over time imposing lower and upper bounds and an ultimate level of fertility to address the issues rising from structural change. Later on, Lee and Tuljapurkar (1994) used this method with a different ultimate level of fertility and without bounds. Similarly, another approach to avoid the problem of structural change is setting a target total fertility (e.g. the average expectation of a group of experts) and then again face the problem of deriving age-specific dynamics (Lutz et al. 1996). To overcome this challenge, Thompson et al. (1989) forecasted the total fertility rate, the mean age at childbearing, and the standard deviation of the age at childbearing using the gamma distribution and then estimated age-specific fertility rates from these parameters. Others, rely on probabilistic projections of the total fertility rate (TFR). For example, Alkema et al. (2011) developed a methodology to forecast TFR for all countries

using a Bayesian projection model. From these estimates, the age-specific profile can be derived (Ševčíková et al. 2016). However, important information in other summary measures such as variance of childbearing age is often ignored (Hruschka and Burger 2016). Our method pertains to this last method to derive age-specific fertility rates in the case that an aggregated summary measure such as TFR or mean age at childbearing is forecasted along with the standard deviation of the age at childbearing or other measure of variation.

In this article, we propose a model to obtain future mortality and fertility age-patterns that comply with the projected summary measures. Unlike comparable approaches, we assume only smoothness of future vital rates which is achieved by a two-dimensional P-spline approach as in Currie et al. (2004). Since summary measures are commonly nonlinear functions of the estimated penalized coefficients, Lagrangian multipliers cannot be directly implemented. We hence opted for a Sequential Quadratic Programming (SQP) procedure (Nocedal and Wright 2006) to perform the associated constrained nonlinear optimization. We illustrate our approach with two data sets: mortality of Japanese females, based on future life expectancy predicted by United Nations World Population Prospects (United Nations 2017) and Spanish fertility constrained to total fertility rates, mean and variance of age at childbearing derived by classic time-series analysis.

2 Model on Italian mortality data

For ease of presentation, we formulate the model on mortality data. Suppose that we have deaths, and exposures to risk, arranged in two matrices, $\boldsymbol{Y} = (y_{ij})$ and $\boldsymbol{E} = (e_{ij})$, both with dimension $m \times n_1$. Rows and columns are classified by age at death, $\boldsymbol{a}$, $m \times 1$, and year of death, $\boldsymbol{t}_1$, $n_1 \times 1$, respectively. We assume that the number of deaths y_{ij} at age i in year j is Poisson distributed with mean $\mu_{ij}\, e_{ij}$, where μ_{ij} is the force of mortality. The aim of our model is to reconstruct trends in μ_{ij} for n_2 future years, $\boldsymbol{y}_2, n_2 \times 1$ [*This sentence is not completely clear to me*].

2.1 Life expectancy

Demographers and actuaries often summarize mortality age-patterns with life expectancy. Life expectancy at birth is the average years a newborn is expected to live given the current mortality rates. In lifetable notation, life expectancy at birth is defined as

$$e(0) = \int_0^\infty \ell(x)\, dx,$$

where $\ell(x)$ is survival function. Time-trends of this summary measure are often regular and well-understood. Forecasting a single time-series is therefore relatively easy. Figure 1 (left panel) presents observed life expectancy at age 1 for Japanese females from 1960 to 2016 along with the medium variant up to 2050 as computed by the UN. This variant is calculated considering available data for all countries in the world. The aim of our model is to reconstruct age-specific future mortality patterns that are consistent with this predicted trend.

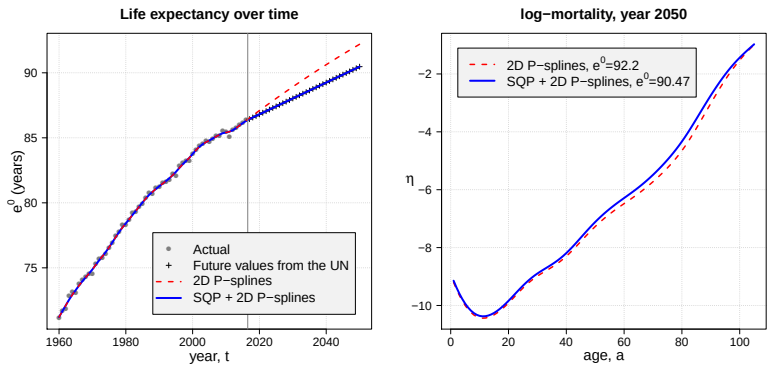

Figure 1: Left panel: Actual, estimated and forecast life expectancy at age 1 by United Nations, 2D P-splines and the SQP+2D P-splines. Right panel: Mortality age-pattern in 2050 by 2D P-splines and the SQP+2D P-splines. Japanese females, ages 1-105, years 1960-2016, forecast up to 2050.

We arrange data as a column vector, that is, $\boldsymbol{y} = \mathtt{vec}(\boldsymbol{Y})$ and $\boldsymbol{e} = \mathtt{vec}(\boldsymbol{E})$ and we model our Poisson death counts as follows: $\boldsymbol{\eta} = \ln(E(\boldsymbol{y})) = \ln(\boldsymbol{e}) + \boldsymbol{B}\,\boldsymbol{\alpha}$, where $\boldsymbol{B}$ is the regression matrix over the two dimensions: $\boldsymbol{B} = \boldsymbol{I}_{n_1} \otimes \boldsymbol{B}_a$, with $\boldsymbol{B}_a \in \mathbb{R}^{m \times k_a}$. Over time, we employ an identity matrix of dimension n_1 because we will incorporate a constraint

for each year. In order to forecast, data and bases are augmented as follows:

$$\breve{\boldsymbol{E}} = [\boldsymbol{E} : \boldsymbol{E}_2]\,, \qquad \breve{\boldsymbol{Y}} = [\boldsymbol{Y} : \boldsymbol{Y}_2]\,, \qquad \breve{\boldsymbol{B}} = \boldsymbol{I}_{n_1+n_2} \otimes \boldsymbol{B}_a\,, \tag{1}$$

where $\boldsymbol{E}_2$ and $\boldsymbol{Y}_2$ are filled with arbitrary future values. If we define a weight matrix $\boldsymbol{V} = \text{diag}(\texttt{vec}(\boldsymbol{1}_{m\times n_1} : \boldsymbol{0}_{m\times n_2}))$, the coefficients vector $\boldsymbol{\alpha}$ can be estimated by a penalised version of the iteratively reweighted least squares algorithm:

$$(\breve{\boldsymbol{B}}^T\boldsymbol{V}\tilde{\boldsymbol{W}}\breve{\boldsymbol{B}} + \boldsymbol{P})\tilde{\boldsymbol{\alpha}} = \breve{\boldsymbol{B}}^T\boldsymbol{V}\tilde{\boldsymbol{W}}\tilde{\boldsymbol{z}}\,, \tag{2}$$

where a difference penalty $\boldsymbol{P}$ enforces smoothness behaviour of mortality both over age and time. Outcomes from this approach in terms of life expectancy is depicted with a dashed red line in Figure 1 (left panel), and a departure from the UN projected values is evident.

Note that life expectancy is a nonlinear function of the coefficients vector $\boldsymbol{\alpha}$:

$$\boldsymbol{e}^0(\boldsymbol{\alpha}) = (\boldsymbol{1}_{1\times m} \otimes \boldsymbol{I}_n)\, \exp[(\boldsymbol{I}_n \otimes \boldsymbol{C})\texttt{vec}(\exp(\boldsymbol{B}\boldsymbol{\alpha}))] + 0.5 \tag{3}$$

where $\boldsymbol{C}$ is a $(m \times m)$ lower triangular matrix filled only with -1.

Constrained nonlinear optimization is therefore necessary and a SQP [*we need to explain more on SQP, perhaps a new subsection*] approach is implemented. Let denote with $\boldsymbol{e}^0_{\text{T}}$ the n_2-vector of target life expectancy for future years and with $\boldsymbol{N}$ the $(k_a n_2 \times n_2)$ matrix with derivatives of (3) with respect to $\boldsymbol{\alpha}$ for each future year. The solution of the associated system of equations at the step $\nu + 1$ is given by

$$\begin{bmatrix} \boldsymbol{\alpha}_{\nu+1} \\ \boldsymbol{\omega}_{\nu+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{L}_\nu & \boldsymbol{H}_\nu \\ \boldsymbol{H}_\nu^T & \boldsymbol{0}_{n_2\times n_2} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{r}_\nu - \boldsymbol{L}_\nu\boldsymbol{\alpha}_\nu \\ \boldsymbol{e}^0_{\text{T}} - \boldsymbol{e}^0(\boldsymbol{\alpha}_\nu) \end{bmatrix}, \tag{4}$$

where $\boldsymbol{L}$ and $\boldsymbol{r}$ are left- and right-hand-side of the system in (2), and matrix $\boldsymbol{H}^T = \left[\boldsymbol{0}_{n_2\times k_a n_1} : \boldsymbol{N}^T\right]$. Vector of $\boldsymbol{\omega}$ denotes the current solution of the associated Lagrangian multipliers.

Forecast e^0 by the proposed method is exactly equal to the UN values (Figure 1, left panel). The right panel of Figure 1 shows the forecast mortality age-pattern in 2050:

Shape obtained by the suggest approach is not a simple linear function of the plain P-splines outcome.

Lifespan variation

Lifespan variation can be measured with multiple indicators (van Raalte and Caswell 2013). These indicators are highly correlated when they measure variation over the full age range, i.e. starting from age 0 (Colchero et al. 2016). Here we measure lifespan variation with two indicators: 1) Years of life lost and 2) the Gini coefficient of the life table.

Years of life lost, denoted with $e^\dagger$, is defined as the average remaining life expectancy when death occurs (Vaupel and Canudas-Romo 2003; Vaupel et al. 2011). It is an indicator of absolute variation in lifespans. For example, when death is highly variable, some people will die well before their expected age at death, contributing many lost years to life disparity. When survival is highly concentrated around older ages, the difference between the age at death and the expected remaining years decreases, and life disparity decreases. In life table notation $e^\dagger$ is given by (Goldman and Lord 1986; Vaupel and Canudas-Romo 2003)

$$e^\dagger = -\int_0^\infty \ell(x)\, \ln \ell(x)\, dx \;=\; \int_0^\infty e(x)\, d(x)\, dx,$$

where $e(x)$ is life expectancy at age x, and $d(x)$ is the age-at-death distribution.

The Gini coefficient is an indicator of relative variation. It was originally proposed in Economics to measure income or wealth inequality and has been adopted in demography and survival analysis to measure lifespan variation (Hanada 1983; Shkolnikov et al. 2003; Bonetti et al. 2009; Gigliarano et al. 2017) There exist several alternative and equivalent ways to define the Gini coefficient (Yitzhaki and Schechtman 2013). For our purposes and the remainder of this article, we will use the following formulation in terms of life table notation (Michetti and Dall'Aglio 1957; Hanada 1983):

$$G = 1 - \frac{\int_0^\infty \ell(x)^2\, dx}{\int_0^\infty \ell(x)\, dx}, \tag{5}$$

where $\ell(x,t)$ is the survival function and $\int_0^\infty \ell(x)^2\, dx$ is the resulting life expectancy at

birth of doubling the hazard at all ages. The Gini coefficient takes values between 0 and 1. A coefficient equal to 0 corresponds to the case of perfect equality in ages at death. The Gini index increases as lifespans become more spread and unequal in the population, reaching a value of 1 in the case of perfect inequality.

3 Spanish Fertility Data

We forecast Spanish fertility using three commonly-used summary measures: Total Fertility Rate, mean and variance of childbearing age, forecast by conventional time-series analysis. We then smooth and constrain future fertility age-patterns to comply these forecast values. Summary measures as well as fertility rates in 2050 are presented in Figure 2.

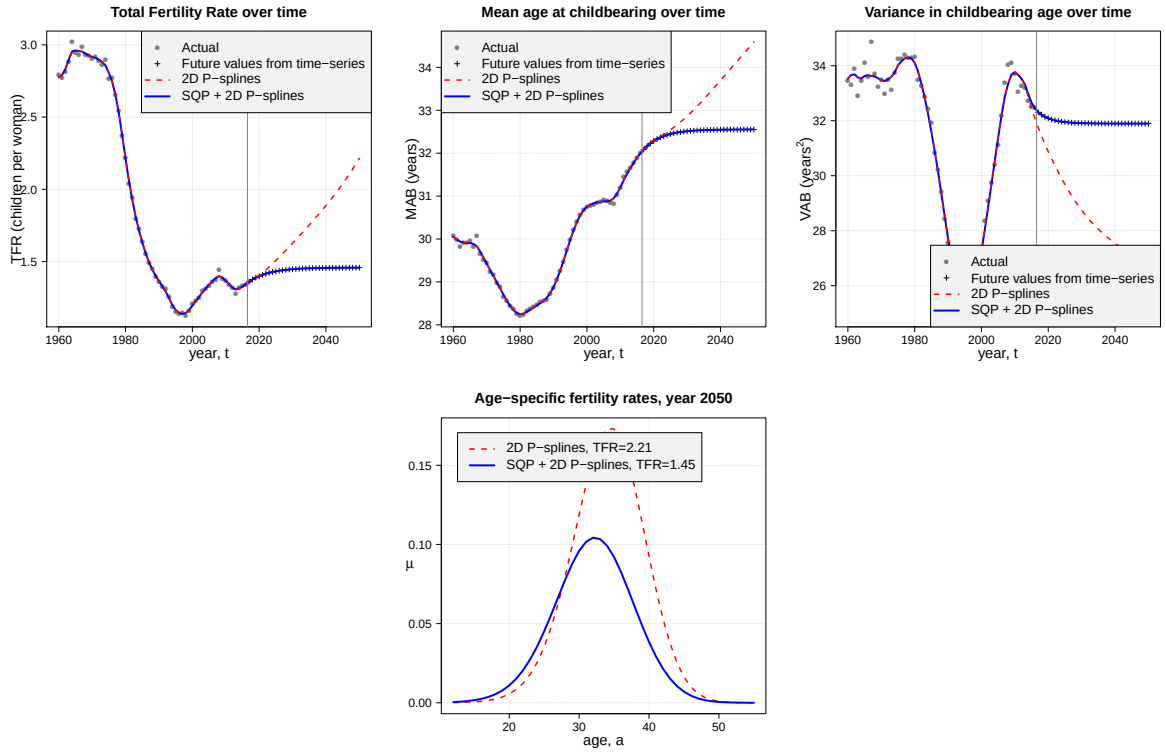

Figure 2: Top and left-bottom panels: Actual, estimated and forecast Total Fertility Rate, Mean and Variance in childbearing age by time-series analysis, 2D P-splines and the SQP+2D P-splines. Right-bottom panel: Age-specific fertility rate in 2050 by 2D P-splines and the SQP+2D P-splines. Spain, ages 12-55, years 1960-2016, forecast up to 2050.

4 Discussion

Appendix A

R code

References

Aburto, J. M. and H. Beltrán-Sánchez (2019). Upsurge of homicides and its impact on life expectancy and life span inequality in mexico, 2005–2015. *American journal of public health 109*(3), 483–489.

Aburto, J. M. and A. A. van Raalte (2018). Lifespan dispersion in times of life expectancy fluctuation: The case of Central and Eastern Europe. *Demography 55*(6), 2071–2096.

Aburto, J. M., F. Villavicencio, U. Basellini, S. Kjærgaard, and J. W. Vaupel (2019). Dynamics of life expectancy and lifespan equality. Working paper presented at PAA 2019 (available from authors).

Alkema, L., A. E. Raftery, P. Gerland, S. J. Clark, F. Pelletier, T. Buettner, and G. K. Heilig (2011). Probabilistic projections of the total fertility rate for all countries. *Demography 48*(3), 815–839.

Alvarez, J.-A., J. M. Aburto, and V. Canudas-Romo (2019). Latin american convergence and divergence towards the mortality profiles of developed countries. *Population studies*, 1–18.

Bohk-Ewald, C., M. Ebeling, and R. Rau (2017). Lifespan disparity as an additional indicator for evaluating mortality forecasts. *Demography 54*(4), 1559–1577.

Bohk-Ewald, C., P. Li, and M. Myrskylä (2018). Forecast accuracy hardly improves with method complexity when completing cohort fertility. *Proceedings of the National Academy of Sciences 115*(37), 9187–9192.

Bonetti, M., C. Gigliarano, and P. Muliere (2009). The gini concentration test for survival data. *Lifetime Data Analysis 15*(4), 493–518.

Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting 22*(3), 547–581.

Chandola, T., D. A. Coleman, and R. W. Hiorns (1999). Recent european fertility patterns: Fitting curves to distorteddistributions. *Population Studies 53*(3), 317–329.

Coale, A. J. and T. J. Trussell (1974). Model fertility schedules: variations in the age structure of childbearing in human populations. *Population index 1*(1), 185–258.

Colchero, F., R. Rau, O. R. Jones, J. A. Barthold, D. A. Conde, A. Lenart, L. Németh, A. Scheuerlein, J. Schoeley, C. Torres, ..., S. C. Alberts, and J. W. Vaupel (2016). The emergence of longevous populations. *Proceedings of the National Academy of Sciences 113*(48), E7681–E7690.

Currie, I. D., M. Durban, and P. H. Eilers (2004). Smoothing and forecasting mortality rates. *Statistical modelling 4*(4), 279–298.

García, J. and J. M. Aburto (2019). The impact of violence on venezuelan life expectancy and lifespan inequality. *International journal of epidemiology*.

Gigliarano, C., U. Basellini, and M. Bonetti (2017). Longevity and concentration in survival times: The log-scale-location family of failure time models. *Lifetime Data Analysis 23*(2), 254–274.

Goldman, N. and G. Lord (1986). A new look at entropy and the life table. *Demography 23*(2), 275–282.

Gonzaga, M. R., B. L. Queiroz, and E. E. C. De Lima (2018). Compression of mortality: the evolution in the variability in the age of death in latin america. *Revista Latinoamericana de Población* (23), 9–35.

Hanada, K. (1983). A formula of gini's concentration ratio and its application to life tables. *Journal of Japanese Statistical Society 13*(2), 95–98.

Hruschka, D. J. and O. Burger (2016). How does variance in fertility change over the demographic transition? *Philosophical Transactions of the Royal Society B: Biological Sciences 371*(1692), 20150155.

Hyndman, R. J. and H. Booth (2008). Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting 24*(3), 323–342.

Lee, R. D. (1993). Modeling and forecasting the time series of us fertility: Age distribution, range, and ultimate level. *International Journal of Forecasting 9*(2), 187–202.

Lee, R. D. and S. Tuljapurkar (1994). Stochastic population forecasts for the united states: Beyond high, medium, and low. *Journal of the American Statistical Association 89*(428), 1175–1189.

Lutz, W., W. Sanderson, S. Scherbov, and A. Goujon (1996). World population scenarios for the 21st century. *The future population of the world. What can we assume today*, 361–396.

Michetti, B. and G. Dall'Aglio (1957). La differenza semplice media. *Statistica 7*(2), 159–255.

Miller, R. B. (1986). A bivariate model for total fertility rate and mean age of childbearing. *Insurance: Mathematics and Economics 5*(2), 133–140.

Nocedal, J. and S. J. Wright (2006). Sequential quadratic programming. In *Numerical optimization*, pp. 529–562. Springer.

Pascariu, M. D., V. Canudas-Romo, and J. W. Vaupel (2018). The double-gap life expectancy forecasting model. *Insurance: Mathematics and Economics 78*, 339–350.

Peristera, P. and A. Kostaki (2007). Modeling fertility in modern populations. *Demographic Research 16*, 141–194.

Raftery, A. E., J. L. Chunn, P. Gerland, and H. Ševčíková (2013, Jun). Bayesian probabilistic projections of life expectancy for all countries. *Demography 50*(3), 777–801.

Raftery, A. E., N. Lalic, and P. Gerland (2014). Joint probabilistic projection of female and male life expectancy. *Demographic research 30*, 795.

Schmertmann, C. (2003). A system of model fertility schedules with graphically intuitive parameters. *Demographic research 9*, 81–110.

Ševčíková, H., N. Li, V. Kantorová, P. Gerland, and A. E. Raftery (2016). Age-specific mortality and fertility rates for probabilistic population projections. In *Dynamic demographic analysis*, pp. 285–310. Springer.

Shkolnikov, V. M., E. E. Andreev, and A. Z. Begun (2003). Gini coefficient as a life table function: Computation from discrete data, decomposition of differences and empirical examples. *Demographic Research 8*(11), 305–358.

Smits, J. and C. Monden (2009). Length of life inequality around the globe. *Social Science & Medicine 68*(6), 1114–1123.

Thompson, P. A., W. R. Bell, J. F. Long, and R. B. Miller (1989). Multivariate time series projections of parameterized age-specific fertility rates. *Journal of the American Statistical Association 84*(407), 689–699.

Torri, T. and J. W. Vaupel (2012). Forecasting life expectancy in an international context. *International Journal of Forecasting 28*(2), 519–531.

Tuljapurkar, S. and R. D. Edwards (2011). Variance in death and its implications for modeling and forecasting mortality. *Demographic Research 24*(21), 497–526.

United Nations (2017). *World population prospects: the 2017 revision*. United Nations.

van Raalte, A. A. and H. Caswell (2013). Perturbation analysis of indices of lifespan variability. *Demography 50*(5), 1615–1640.

van Raalte, A. A., I. Sasson, and P. Martikainen (2018). The case for monitoring life-span inequality. *Science 362*(6418), 1002–1004.

Vaupel, J. W. and V. Canudas-Romo (2003). Decomposing change in life expectancy: A bouquet of formulas in honor of Nathan Keyfitz's 90th birthday. *Demography 40*(2), 201–216.

Vaupel, J. W., Z. Zhang, and A. A. van Raalte (2011). Life expectancy and disparity: An international comparison of life table data. *BMJ Open 1*(1), bmjopen–2011–000128.

White, K. (2002). Longevity advances in high-income countries, 1955-96. *Population and Development Review 28*(1), 59–76. cited By 75.

Yitzhaki, S. and E. Schechtman (2013). *The Gini Methodology*. Springer New York.