Population Research and Policy Review (2023) 42:66
https://doi.org/10.1007/s11113-023-09813-y

RESEARCH BRIEFS

Persistence in UK Historical Data on Life Expectancy

Guglielmo Maria Caporale[1] · Juan Infante[2] · Marta del Rio[2] · Luis A. Gil-Alana[3,4]

Received: 14 February 2023 / Accepted: 30 June 2023 / Published online: 21 July 2023

Abstract
This paper provides estimates of persistence in historical UK data on life expectancy applying fractional integration methods to both an annual series from 1842 to 2019 and a 5-year average from 1543 to 2019. This method is the most appropriate for our purposes since it is more general and flexible than the classical methods based on integer differentiation. The results indicate that the former exhibits an upward trend and is persistent but mean reverting; the same holds for the latter, though its degree of persistence is higher. Similar results are obtained for the logged values. On the whole, this evidence suggests that the effects of shocks to the series are transitory though persistent, which is useful information for policy makers whose task is to take appropriate measures to increase life expectancy.

Keywords Life expectancy · Long memory · Fractional integration

JEL Classifications C22 · C40 · D60

Introduction

Life expectancy is a useful indicator of a population's health (Roser et al., 2013). Before the Age of Enlightenment, life expectancy was generally very low (around 30 years) across the world. However, during the early 1800s, it began to rise in countries where the Industrial Revolution took place, creating a significant gap in health conditions between wealthy and impoverished nations. Throughout the world, there has been a significant increase in the average lifespan since the early twentieth

✉ Luis A. Gil-Alana
alana@unav.es

1 Brunel University London, Uxbridge, London, UK

2 University Villanueva, Madrid, Spain

3 Faculty of Economics, DATAI and ICS, University of Navarra, Pamplona, Spain

4 Universidad Francisco de Vitoria, Madrid, Spain

century (World Bank Group, 2018). Oeppen and Waupel (2002) reported that this increase has been occurring at a rate of about a quarter of a year per year over a 160-year period. Among the factors contributing to this increase are improved living conditions and medical care (Katz et al., 1983). By the last decade of the twentieth century, the average lifespan had reached 50 years (World Bank Group, 2018). This marks the first time in history that entire populations have experienced long-term improvements in welfare (Riley, 2005). However, there are still considerable differences between countries in terms of life expectancy (Cutler et al., 2006): in 2019, the Central African Republic had the shortest one (53 years) and Japan the longest (three decades more). Differences also remain between individuals in the same country, although they have declined over time; these can be measured using the Gini coefficient in the same way as income inequality (Peltzman, 2009). Also, life expectancy is positively linked to GDP per person, a relationship known as the "Preston curve" (Preston, 1975), which can be estimated rather accurately using a Poisson common factor model (Li, 2013).

An interesting issue not much empirically investigated in the literature is the persistence of life expectancy. Various theories have been proposed to explain it, including medical advancements, economic development, lifestyle changes, environmental factors, and social support networks. Medical advancements have been crucial in reducing infectious diseases and improving health outcomes (Cutler et al., 2006). Economic growth has improved access to healthcare, living conditions, and education (Caporale & Gil-Alana, 2016). Lifestyle changes, such as increased physical activity, better diets, and reduced smoking rates, have also contributed to improved health outcomes (Katz et al., 1983). Improvements in environmental factors such as cleaner air and water have increased life expectancy (Riley, 2001). Strong social support networks and community engagement promote healthy behaviours and emotional and provide practical support (Peltzman, 2009). Overall, the complex interplay of these factors has contributed to the persistence of life expectancy (Roser et al., 2013).

Various authors have proposed different approaches to examine life expectancy. Caporale and Gil-Alana (2016) carried out univariate analysis to investigated time trends in life expectancy in Sub-Saharan Africa, while other authors have explored the relationship between life expectancy and other variables. For instance, Cutler et al. (2006) examined the determinants of mortality, Peltzman (2009) studied mortality inequality, and Preston (1975) investigated the changing relationship between mortality and the level of economic development. Moreover, Riley (2001, 2005) examined global and regional estimates of life expectancy from 1800 to 2001. Some authors have proposed specific models for projecting mortality and life expectancy; for instance, Li (2013) developed a Poisson common factor model for jointly projecting mortality and life expectancy for females and males, and Oeppen and Waupel (2002) studied the limitations of life expectancy and the potential for extending the human lifespan.

Examining life expectancy in the UK is particularly informative owing to its well-developed system for collecting and reporting health data. The UK experienced significant changes in life expectancy over the past century, and thus it represents an important case study shedding light on the factors driving such changes.

In particular, the UK government focussed on improving health outcomes through policy interventions, and thus comparing its experience with that other countries can provide insights into the factors accounting for differences in health outcomes across populations. This type of analysis can inform policy decisions concerning health and social care and provide a better understanding of factors that affect health outcomes over time.

The purpose of analysing life expectancy in the UK is to detect time trends, investigate the causes of the recent decline in life expectancy improvements, and evaluate the impact of policy interventions such as public health campaigns, healthcare policy changes, or social welfare programmes, on life expectancy and related health outcomes, in order to identify the most effective policies. In particular, our investigation shed light on the degree of persistence of the series and on whether exogenous shocks have transitory or permanent effects; this is essential information for policy makers to decide whether or not policy actions should be taken.

The rate of progress in life expectancy in the UK has slowed down in recent years and come to a halt between 2011 and 2019 after significant improvement over the past century (Riley, 2001, 2005; Our World in Data 2023). The Covid-19 pandemic has further affected it and exposed existing health disparities. Factors contributing to the slowdown in life expectancy improvements include rising inequality, changes in lifestyle and diet, and reductions in public health funding (Peltzman, 2009). Various types of models, such as the Chebyshev polynomials-based approach, Bloomfield's exponential model, and Granger and Joyeux's long memory time series, can be used to model the persistence and non-linearities of life expectancy changes over time (Bloomfield, 1973; Cuestas & Gil-Alana, 2016; Granger & Joyeux, 1980; Hosking, 1981). Other statistical models, such as nonstationary hypotheses testing and Poisson common factor models, can be used to identify the factors that affect life expectancy (Gil-Alana & Robinson, 1997; Li, 2013; Robinson, 1994). Ongoing data collection and analysis are crucial to understanding and improving health outcomes in the UK, informing policy decisions and interventions aimed at reducing health disparities and improving overall health (Cutler et al., 2006; Oeppen & Waupel, 2002; Preston, 1975).

Applying fractional integration methods fills a gap in the literature by providing a more general and flexible method for modelling long-range dependence in time series data on life expectancy. Previous studies have examined various aspects of life expectancy, such as time trends and the factors driving it. However, they have not used the long-memory approach followed in the present study, which sheds new light on the stochastic properties of the series of interest and can also yield more accurate forecasts. The layout of the paper is the following: "Data and Methodology" section describes the data and the methodology; "Empirical Results" section discusses the empirical findings; Final section offers some concluding remarks.

Data and Methodology

The series used for the analysis are UK life expectancy from 1842 to 2019 at an annual frequency and its 5-year average from 1543 to 2018. They have been constructed by 'OurWorldinData', which is a project of the Global Change Data Lab, a non-profit organisation based in the UK (Registered Charity Number 1186433), and available from Roser et al. (2013) at https://ourworldindata.org/life-expectancy (online resource). The annual figures are the average number of years a newborn would live if age-specific mortality rates in the c rrent year were to stay the same throughout its life, whilst the 5-year average is calculated in each case as the arithmetic mean over 5 years. The 5-year average is used to filter out short-run fluctuations.

The chosen, univariate, fractional integration approach is ideal for our purposes given its ability to capture long-term dependencies and persistence in the data; it has previously been used to estimate time trends in life expectancy in sub-Saharan Africa (Caporale & Gil-Alana, 2016). Multivariate methods would instead be required to analyse the relationship between life expectancy and variables such as economic development, mortality inequality, and changing levels of economic development (Katz et al., 1983; Murray and Lopez 1996; Oeppen & Waupel, 2002; Peltzman, 2009; Preston, 1975; Roser et al., 2013; World Bank Group, 2018), which is beyond the scope of the present study.

Figures 1 and 2 display the two series of interest, namely the annual one (1842–2019) and the 5-year average (1543–2018), respectively. While negative shocks such as World War I and II (see Fig. 1) and the Anglo-Spanish Wars in the XVI and XVIII centuries (see Fig. 2) have had an impact on life expectancy, an overall upward trend is visible. This trend can be attributed to various factors, including improvements in healthcare, sanitation, immunizations, access to clean

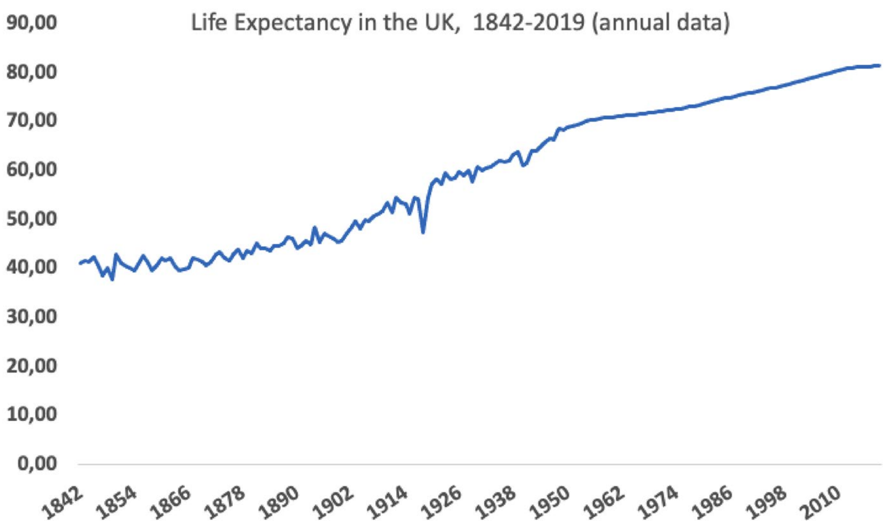

Fig. 1 Life expectancy in the UK: 1842-2019

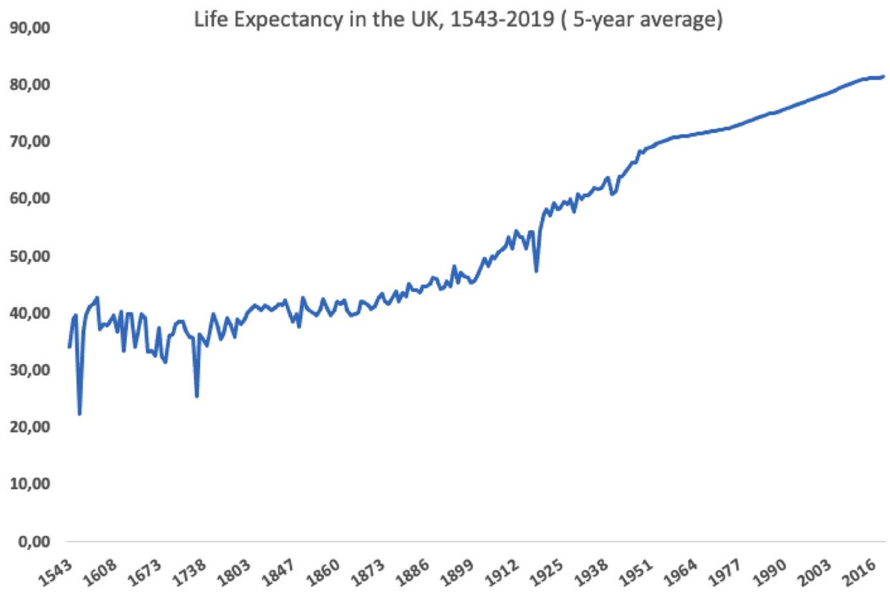

Fig. 2 Life expectancy in the UK: 1543-2019

water, and better nutrition, which are highlighted in Riley (2005). Measuring active life expectancy, or the number of years a person can expect to live without disability or limitations in daily activities, as in Katz et al. (1983) can provide a more comprehensive understanding of health outcomes. The World Bank's data (2018) on global life expectancy trends further support the notion of an overall upward trend. Furthermore, Oeppen and Waupel (2002) challenge the idea that there is a biological limit to the human lifespan, suggesting that continued improvements in health outcomes and life expectancy are possible, and Murray and Lopez (1996) emphasize the need for continued efforts to improve health outcomes and reduce mortality and disability from diseases, injuries, and risk factors.

The degree of persistence of these series is examined by estimating the following regression model:

$$y_t = \beta_0 + \beta_1 t + x_t, \qquad t = 1, 2, ..., \tag{1}$$

where y_t stands for the series of interest, β_0 and β_1 denote the intercept and the coefficient on a linear trend, respectively, and x_t is the error term, which is assumed to be integrated of order d:

$$(1 - B)^d x_t = u_t, \qquad t = 1, 2, ..., \tag{2}$$

Using a Binomial expansion, the left-hand side of (2) can be expanded, with B being the backshift operator, such that $B^k x_t = x_{t\text{-}k}$, and u_t is I(0) (as in Granger & Joyeux, 1980 and Hosking, 1981), to obtain the following expression:

Table 1 Estimates of the differencing parameter

Series	No regressors	An intercept	A linear time trend
Annual	0.91 (0.81, 1.04)	0.75 (0.71, 0.80)	**0.65 (0.58, 0.73)**
5-years	0.86 (0.76, 1.00)	0.75 (0.69, 0.84)	**0.73 (0.66, 0.83)**

White noise errors

The values inside the parentheses are the 95% confidence intervals for the non-rejection values of *d*, while the coefficients from the selected models are in bold

Table 2 Estimated coefficients from the selected model in Table 1

Series	No regressors	Intercept (*t*-value)	Time trend (*t*-value)
Annual	0.65 (0.58, 0.73)	**39.681 (37.88)**	**0.242 (13.54)**
5-years	0.73 (0.66, 0.83)	**33.243 (9.94)**	**0.448 (3.67)**

In bold in columns 3 and 4 the *t*-values of the corresponding coefficients

$$(1-B)^d = \sum_{j=0}^{d} \binom{d}{j} (-1)^j B^j = 1 - dB + \frac{d(d-1)}{2} - \ldots \tag{3}$$

where the parameter d is a measure of persistence and sheds light on the properties of the process being modelled. Specifically, if $d=0$ this exhibits short memory, whilst $d>0$ implies long memory; if $d<0.5$, it is covariance stationary and mean reverting; if $0.5 \leq d < 1$ it is nonstationary but mean reversion still occurs; if $d \geq 1$, the process is explosive. By allowing d to be any real value, including fractional ones, this framework encompasses the classical unit root case as a special one when $d=1$, and is preferable to the simple approach of estimating linear trends in Eq. (1) under the assumption that $d=0$ in (2).

We employ the testing procedure proposed by Robinson (1994) that is based on the Lagrange Multiplier (LM) principle and includes a version of the Whittle function in the frequency domain, where the null is the following:

$$H_o : \quad d \quad = \quad d_o, \tag{4}$$

Note that in Eqs. (1) and (2) d can be any real number, including decimals from the nonstationary range ($d \geq 0.5$), but the limit distribution of the test statistc is standard $N(0, 1)$ (for its functional form see Gil-Alana & Robinson, 1997).

Empirical Results

Tables 1, 2, 3, 4, 5, 6, 7, and 8 report the estimated values of d alongside their 95% confidence intervals for three different model specifications: (i) without deterministic terms, (ii) with an intercept only, and (iii) with an intercept as well as a

Table 3 Estimates of the differencing parameter

Series	No regressors	An intercept	A linear time trend
Annual	0.85 (0.66, 1.08)	0.88 (0.82, 0.97)	**0.82 (0.72, 0.96)**
5-years	0.93 (0.76, 1.12)	1.00 (0.87, 1.15)	**1.00 (0.85, 1.17)**

Autocorrelated errors

The values inside the parentheses are the 95% confidence intervals for the non-rejection values of d, while the coefficients from the selected models are in bold

Table 4 Estimated coefficients from the selected model in Table 3

Series	No regressors	Intercept (t-value)	Time trend (t-value)
Annual	0.82 (0.72, 0.96)	**40.480 (34.29)**	**0.236 (6.09)**
5-years	1.00 (0.85, 1.17)	**33.940 (9.96)**	–

In bold in columns 3 and 4 the t-values of the corresponding coefficients

Table 5 Estimates of the differencing parameter

Series in logs	No regressors	An intercept	A linear time trend
Annual	0.97 (0.88, 1.10)	0.72 (0.68, 0.77)	**0.61 (0.55, 0.69)**
5-years	0.95 (0.82, 1.13)	0.62 (0.55, 0.70)	**0.57 (0.49, 0.67)**

White noise errors

The values inside the parentheses are the 95% confidence intervals for the non-rejection values of d, while the coefficients from the selected models are in bold

Table 6 Estimated coefficients from the selected model in Table 5

Series in logs	No regressors	Intercept (t-value)	Time trend (t-value)
Annual	0.61 (0.55, 0.69)	**3.689 (179.98)**	**0.00422 (13.66)**
5-years	0.57 (0.49, 0.67)	**3.498 (43.65)**	**0.00809 (4.29)**

In bold in columns 3 and 4 the t-values of the corresponding coefficients

Table 7 Estimates of the differencing parameter

Series in logs	No regressors	An intercept	A linear time trend
Annual	0.93 (0.78, 1.14)	0.84 (0.77, 0.93)	**0.77 (0.67, 0.90)**
5-years	0.89 (0.67, 1.15)	0.81 (0.69, 0.97)	**0.78 (0.64, 0.96)**

Autocorrelated errors

The values inside the parentheses are the 95% confidence intervals for the non-rejection values of d, while the coefficients from the selected models are in bold

Table 8 Estimated coefficients from the selected model in Table 7

Series	No regressors	Intercept (t-value)	Time trend (t-value)
Annual	0.77 (0.67, 0.90)	**3.703 (155.23)**	**0.00411 (6.43)**
5-years	0.78 (0.64, 0.96)	**3.520 (37.59)**	**0.00859 (2.12)**

In bold in columns 3 and 4 the t-values of the corresponding coefficients

linear time trend. The coefficients in bold are those from the models selected on the basis of the statistical significane of the regressors.

Tables 1, 2, 3, and 4 display the results for the original data, while Tables 5, 6, 7, and 8 show the corresponding ones for the logged transformed series, in both cases under the alternative assumptions of white noise and autocorrelated residuals; the latter are modelled using the exponential spectral approach of Bloomfield (1973) that approximates well AR structures in the context of Robinson's (1994) tests (see Gil-Alana, 2004). Concerning the raw data, with white noise residuals the time trend is significant and positive for both series (Table 2); as for d, the estimated values are 0.65 and 0.73, respectively, for the annual and 5-year average series (Table 1). Moreover, the confidence intervals do not include 1, which supports the hypothesis of mean reversion. When allowing for autocorrelated residuals (Tables 3 and 4) the time trend is again statistically significant and positive for both series, and the estimates of d are now 0.82 and 1.00 (i.e. they are much higher than in the previous case) and mean reversion ($d < 1$) is only found in the case of the annual series.

Concerning the logged values (Tables 5, 6, 7, and 8), the time trend is significant for both series regardless of the error term specification, and the estimated values of d are now 0.61 (annual) and 0.57 (5-years) with white noise errors, and 0.77 and 0.78 with autocorrelation, whilst the hypothesis of mean reversion cannot be rejected in any single case, which implies that the effects of shocks will gradually die away. In addition, in all cases the hypothesis of short memory or I(0) behaviour is rejected, which implies that estimating the linear trends using only the model given by Eq. (1) will produce misleading results.

Conclusions

This paper examines persistence in historical UK data on life expectancy using a fractional integration approach. The advantage of this method is that it is a very general and flexible framework encompassing a variety of cases including those when the data are nonstationary but still display mean reversion; moreover, the estimated fractional differencing parameter provides a direct estimate of the degree of persistence of the series under examination. The data analysed are annual observations for the period from 1842 to 2019 and the 5-year average from 1543 to 2018. The results indicate that the former series exhibits a statistically significant and positive time trend as well as a high degree of persistence but nevertheless it is mean reverting, and thus shocks only have transitory effects; the findings are similar for the latter series, though its degree of persistence is higher. Nonstationary (though mean reverting) behaviour is also found in the case of the logged series. This evidence represents useful information for policy makers whose task is to take appropriate measures to increase life expectancy. Policy makers can implement several measures to increase life expectancy. These include improving healthcare access and quality, investing in preventive healthcare, addressing socioeconomic determinants of health, promoting healthy behaviours (focussing on mental health where necessary), fostering social connections, ensuring healthy ageing, supporting relevant research with data collection and fostering collaboration across sectors. These measures can enhance healthcare services, reduce risk factors, tackle social inequalities, promote healthy lifestyles, and create supportive environments that have a positive impact on health outcomes and overall life expectancy.

One limitation of the current study is that it focuses exclusively on the UK case. Unfortunately, extending the analysis to other countries is constrained by data availability issues; also, different countries have different data collection methods and reporting systems, which can affect the accuracy and comparability of the results. The findings for the UK are robust to using alternative methods, including both parametric (Sowell, 1992) and semiparametric (Geweke & Porter-Hudak, 1983) methods as, for instance, in Caporale et al. (2016) (these results are not reported for reasons of space). Future work could also investigate possible non-linearities by including non-linear deterministic components such as Chebyshev polynomials in time (Cuestas & Gil-Alana, 2016), Fourier functions in time (Caporale et al., 2022), or neural network approximations (Yaya et al., 2021), all within a fractional integration framework, and assess the empirical adequacy of more complex models of life expectancy (Caporale & Gil-Alana, 2016; Cutler et al., 2006; Katz et al., 1983; Peltzman, 2009; Riley, 2001; Roser et al., 2013).

While each country has its unique context and challenges, the similarities in lifestyle and development between the UK and other developed countries make the conclusions from this study relevant and applicable to a wider set of countries. Policymakers can use the insights provided by our analysis to tailor policies that are suitable for their own populations, promoting healthier lifestyles, improving healthcare systems, and ultimately increasing life expectancy. Overall, while the immediate focus of the present study is on the UK, its findings contribute more generally to the existing knowledge body of knowledge on life expectancy and provide new insights that can inform discussions and policy decisions in other regions and countries around the world.

Acknowledgements Luis A. Gil-Alana gratefully acknowledges financial support from the MINEIC-AEI-FEDER ECO201785503-R project from 'Ministerio de Economía, Industria y Competitividad' (MINEIC), `Agencia Estatal de Investigación' (AEI) Spain and `Fondo Europeo de Desarrollo Regional' (FEDER). An internal Project from the Universidad Francisco de Vitoria is also acknowledged. Comments from the Editor and three anonymous reviewers are gratefully acknowledged.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. The funding was received from Fundación Biodiversidad, MINEIC-AEI-FEDER ECO2017-85503-R, Luis Gil-Alana.

Data availability Data used in this research are available from the authors upon request.

Declarations

Conflict of interest There is no conflict of interest with the publication of the present manuscript.

References

Bloomfield, P. (1973). An exponential model in the spectrum of a scalar time series. *Biometrika, 60*, 217–226.

Caporale, G. M., & Gil-Alana, L. A. (2016). Estimating time trends in life expectancy in Sub-Saharan Africa. *Population Review*. https://doi.org/10.1353/prv.2016.0004

Caporale, G.M., Gil-Alana, L.A., and Yaya, O.S. (2022). Modelling persistence and non-linearities in the US Treasury 10-year bond yields. *CESifo Working Paper No. 9554. Economics Bulletin, forthcoming*.

Cuestas, J. C., & Gil-Alana, L. A. (2016). A nonlinear approach with long range dependence based on Chebyshev polynomials. *Studies in Nonlinear Dynamics and Econometrics, 20*, 57–94.

Cutler, D., Deaton, A., & Lleras-Muney, A. (2006). The determinants of mortality. *Journal of Economic Perspectives, 20*(3), 97–120.

Geweke, J., & Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis, 4*(4), 221–238.

Gil-Alana, L. A. (2004). The use of the Bloomfield model as an approximation to ARMA process in the context of fractional integration. *Mathematical and Computer Modelling, 39*(4), 429–436.

Gil-Alana, L. A., & Robinson, P. M. (1997). Testing of unit roots and other nonstationary hypotheses in macroeconomic time series. *Journal of Econometrics, 80*, 241–268.

Granger, C. W. J., & Joyeux, R. (1980). An introduction to long memory time series and fractional differencing. *Journal of Time Series Analysis, 1*, 15–29.

Hosking, J. R. M. (1981). Fractional Differencing. *Biometrica, 68*, 168–176.

Katz, S., Branch, L. G., Branson, M. H., Papsidero, J. A., Beck, J. C., & Greer, D. S. (1983). Active life expectancy. *The New England Journal of Medicine, 1983*(309), 1218–1224. https://doi.org/10.1056/NEJM198311173092005

Li, J. (2013). A Poisson common factor model for projecting mortality and life expectancy jointly for females and males. *A Journal of Demography, 67*(1), 111–126.

Murray, C. J. L., & Lopez, A. D., World Health Organization, World Bank & Harvard School of Public Health. (1996). The Global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020: summary / edited by Christopher J. L. Murray, Alan D. Lopez. World Health Organization. https://apps.who.int/iris/handle/10665/41864

Oeppen, J., & Waupel, J. W. (2002). Broken limits to life expectancy. *Science, 296*(5570), 1029–1031. https://doi.org/10.1126/science1069675

Our world in data. (2023). https://ourworldindata.org/organization

Peltzman, S. (2009). Mortality Inequality. *Journal of Economic Perspectives, 23*(4), 175–219.

Preston, S. H. (1975). The changing relation between mortality and level of economic development. *Population Studies, 29*(2), 231–248. https://doi.org/10.2307/2173509

Riley, J. C. (2001). *Rising life expectancy: A global history*. Cambridge University Press.

Riley, J. C. (2005). Estimates of regional and global life expectancy, 1800–2001. *Population and Development Review, 31*(3), 537–543.

Robinson, P. M. (1994). Efficient tests of nonstationary hypotheses. *Journal of the American Statistical Association, 89*, 1420–1437.

Roser, M., Ortiz-Ospina, E., and Ritchie, H. (2013). Life Expectancy. *Published online at OurWorldInData.org*. Retrieved from: 'https://ourworldindata.org/life-expectancy' [Online Resource]

Sowell, F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics, 53*(1–3), 165–188.

World Bank Group (2018). Retrieved 17 March 2018 from https://data.worldbank.org/indicator/SP.DYN.LE00.IN?locations=ZG

Yaya, O., Ogbonna, A.E., Furuoka, F., & Gil-Alana, L.A. (2021). A New Unit Root Test for Unemployment Hysteresis Based on the Autoregressive Neural Network. *Oxford Bulletin of Economics and Statistics, forthcoming.*