

PAPER

# Statistical features of persistence and long memory in mortality data

Gareth W. Peters<sup>1\*</sup>, Hongxuan Yan<sup>2</sup> and Jennifer Chan<sup>3</sup>

<sup>1</sup>Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh EH14 4AS, USA; <sup>2</sup>Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China; and <sup>3</sup>School of Mathematics and Statistics, The University of Sydney, Camperdown, NSW 2006, Australia

\*Corresponding author. E-mail: [g.peters@hw.ac.uk](mailto:g.peters@hw.ac.uk)

(Received 31 December 2019; revised 22 March 2021; accepted 25 March 2021; first published online 11 May 2021)

## Abstract

Understanding core statistical properties and data features in mortality data are fundamental to the development of machine learning methods for demographic and actuarial applications of mortality projection. The study of statistical features in such data forms the basis for classification, regression and forecasting tasks. In particular, the understanding of key statistical structure in such data can aid in improving accuracy in undertaking mortality projection and forecasting when constructing life tables. The ability to accurately forecast mortality is a critical aspect for the study of demography, life insurance product design and pricing, pension planning and insurance-based decision risk management. Though many stylised facts of mortality data have been discussed in the literature, we provide evidence for a novel statistical feature that is pervasive in mortality data at a national level that is as yet unexplored. In this regard, we demonstrate in this work a strong evidence for the existence of long memory features in mortality data, and second that such long memory structures display multifractality as a statistical feature that can act as a discriminator of mortality dynamics by age, gender and country. To achieve this, we first outline the way in which we choose to represent the persistence of long memory from an estimator perspective. We make a natural link between a class of long memory features and an attribute of stochastic processes based on fractional Brownian motion. This allows us to use well established estimators for the Hurst exponent to then robustly and accurately study the long memory features of mortality data. We then introduce to mortality analysis the notion from data science known as multifractality. This allows us to study the long memory persistence features of mortality data on different timescales. We demonstrate its accuracy for sample sizes commensurate with national-level age term structure historical mortality records. A series of synthetic studies as well a comprehensive analysis of real mortality death count data are studied in order to demonstrate the pervasiveness of long memory structures in mortality data, both mono-fractal and multifractal functional features are verified to be present as stylised facts of national-level mortality data for most countries and most age groups by gender. We conclude by demonstrating how such features can be used in kernel clustering and mortality model forecasting to improve these actuarial applications.

**Keywords:** Persistence; Long memory; Hurst exponent; Mortality projection

## 1. Stylised Features Found in National-Level Mortality Data

Mortality is studied in many areas such as demography, insurance and social policy. It is crucial for governments and insurance companies to obtain accurate forecasts of mortality rates to provide population projections for demographers and cash flow projections for actuaries (Koissi & Shapiro, 2008). There will be adverse effects on pension funds and annuity insurers if the mortality forecasts are not sufficiently accurate. Both systematic under or over estimation of mortality rates

will cause detrimental effects: as either high-risk profiles for pension funds and annuity insurers; or as excess premium costs for policyholders (Giacometti *et al.*, 2012).

Numerous authors have studied attributes of mortality data, for instance, Cairns *et al.* (2008) proposed several stylised facts of mortality data. They found that mortality rates have fallen dramatically at all ages. In particular, the rate of decrease in mortality has varied over time and by age group. Furthermore, absolute decreases have varied by age group, and that aggregate mortality rates have significant volatility year on year, especially for both young and old ages.

However, we demonstrate in this paper an important stylised fact of national-level mortality data that is consistently present at the vast majority of age groups, gender and country that was not previously studied. This is the presence of persistence or long memory in the time series structure of mortality data. We argue that such mortality data features if adequately incorporated into models will improve the ability to forecast mortality and enhance the accuracy of life table calculations. We further extend this concept to also show that such persistence can be shown to behave differently for a given age group, gender and country when studied over multiple timescales, through data science methods known as a multifractality estimation. We argue that such multifractal structures, when summarised appropriately, can be used in data science for both improving modelling and as discriminatory features in clustering and partition algorithms. We demonstrate in this paper how to utilise and interpret such long memory on different timescales as a statistical feature.

Two real data studies have been implemented in this paper to verify the conjecture we hold that long memory features are consistently present in the mortality data. The first study considers annual long memory persistence features for 16 countries, by age and gender. The second study considers the presence of different strengths of long memory persistence in mortality data at different timescales, for instance, quarterly, half-yearly, yearly, biyearly, decadal and so forth. Since the majority of publicly available data are not available at such a wide variety of countries by age and gender neither for long periods of time and not accessible on intra-yearly timescales; therefore, classical methods utilised to study long memory structures in mortality data on such timescales would fail. To counter this issue, we have introduced to this literature the notion of multifractal long memory analysis, which allows us to modify the estimators of long memory to capture these time-scaling effects on different timescales, whilst using annual data that is readily available. We believe the introduction of these methods and the results they produce are significant novel additions to the mortality and demographics literature.

To illustrate these statistical features at a national-level study, we consider the commonly utilised Human Mortality Database (HMD) (Shkolnikov, 2017). Moreover, in each group of gender, age and country, the prevalence of such features has been confirmed by a range of estimation procedures using Hurst exponent estimators adopted from statistics. To ensure the accuracy of our findings, we present a range of different estimators that capture the long memory structures in mortality data, to show that these are in agreement, giving further significance to the validity of our findings.

### 1.1 Contribution and structure

The main contribution of this work is to present evidence for the presence of persistence or long memory in mortality data and to study the behaviour of such features in mortality rates from 16 countries cross-classified by gender and age groups. There are limited numbers of very recent studies that explore persistence in mortality data in the actuarial literature, see for instance, Yan *et al.* (2020), and in demographics and population studies literature, see Yaya *et al.* (2019) and Yaya & Gil-Alana (2020). Our work in this paper provides a comprehensive analysis of both monofractal and multifractal long memory features in national-level death count data to establish such stochastic structures as a stylised fact when investigating mortality data. Various patterns of long memory across different countries, age groups, genders and time periods are demonstrated to be present and statistically significant to validate this claim for national-level mortality data.

Our second contribution is to present a simple class of three estimators to study the long memory structures in mortality data. These are constructed from non-parametric measures of the so-called Hurst exponent  $H$ , which is present in fractional Brownian motion. We apply the estimators of the Hurst exponent to investigate the long memory pattern in these national-level mortality data records. The relationship between the representation of long memory through a fractional derivative of the time series back shift operator with order  $d$  and the Hurst exponent  $H$  is introduced.

To introduce these concepts to the actuarial audience, we provide a guide to commonly used estimators of Hurst exponent, namely rescaled range analysis (R/S), detrended fluctuation analysis (DFA) and periodogram regression (PR) methods. We note that such estimators are commonly implemented in statistical packages in R and Python, which makes them readily accessible to practitioners.

A series of synthetic studies are implemented to test the performance of these estimators under different data lengths, different long memory strengths, different missing value settings, different aggregation types and different roundings. All of these attributes and data transformations are commonly encountered when studying national-level mortality data.

We then proceed with introducing the notion of multifractal long memory structures, by explaining how to develop a class of estimators that are applicable to mortality data. Such estimators produce a characteristic signature feature of particular age, gender and country mortality data that is valuable for discrimination and forecasting model development. We then apply these estimators to a variety of countries to illustrate this point on real data. We note that all estimators we explain and introduce to actuarial data science practice are readily available in other communities in R and Python packages, so we tested these improving some erroneous implementations before applying the final corrected package codes to our analysis.

We conclude the paper by demonstrating how the functional feature extraction of multifractal Hurst exponent curves can be used to perform clustering of mortality experience by age, country and gender in an advanced non-linear discriminant method based on kernel k-means, a modern machine learning variant of traditional k-means, in which function space clustering is applied. We demonstrate that such decompositions allow us to group countries and age groups into three classes of mortality structures as characterised by their long memory characteristics. Lastly, we also demonstrate then when one selects populations from each cluster group and then builds a mortality forecasting model that incorporates long memory properties, the model with such features will outperform in forecast accuracy and precision the classical Lee–Carter period effect type mortality models widely used. This demonstrates how such findings of stylised features of long memory in mortality data have direct implications for actuarial practice.

The paper is organised as follows. Section 2 defines Hurst exponent that is derived from fractional Brownian motion and reviews three popular Hurst exponent estimators. Section 3 performs synthetic studies to assess the properties of long memory estimators. Section 4 analyses the empirical long memory properties amongst 16 countries. Section 5 introduces the methodological extension of long memory analysis to multiple timescales or multi-resolution approaches, by introducing the generalisation of DFA mono-fractal analysis from section 2 to a multifractal MF DFA persistence analysis framework. Section 6 applies the MF DFA multifractal extension to the mortality data. Section 7 introduces practical use cases of the stylised long memory structure observed in the country, age and gender death count data time series to perform practical actuarial tasks of kernel k-means feature-based clustering and Bayesian model-based mortality forecasting. Section 8 concludes the paper.

## 2. Defining Long Memory and its Estimation

The statistical feature of long memory basically refers to the strength of statistical dependence between lagged observations in a time series. The stronger the relationship between observations

as one looks progressively back in time the more persistent is the time series (Graves *et al.*, 2014). When talking about the presence of long memory, we are particularly interested in how rapidly the time series seems to forget its past history. Naturally, one will then think about lagged autocorrelation functions and how rapidly such autocorrelation functions decay with time. In cases in which the rate that such lagged dependency decreases slower than exponential decay, one refers to such a feature as long memory structure in the time series.

Given a stationary time series process for death counts,  $Y \equiv \{Y_t\}_{t=1:T}$ , with  $Y \in (\mathbb{N} \cup \{0\})^T$ , Beran (1994) defined a condition for a long memory stationary process in terms of the divergence of the autocorrelation function (ACF) for  $Y_t$  and  $Y_{t+j}$  at lag  $j$ , such that

$$\lim_{n \rightarrow \infty} \sum_{j=-n}^n |\rho(j)| \rightarrow \infty \text{ where } \rho(j) = \frac{\text{Cov}(Y_t, Y_{t+j})}{\sqrt{\text{Var}(Y_t), \text{Var}(Y_{t+j})}}$$

Various representations have been proposed to describe the long memory feature of time series. The Fractional Gaussian Noise model is the first model with long-range dependence introduced by Mandelbrot & Wallis (1969). To incorporate such long memory property into the autoregressive-moving-average (ARMA) model, Granger & Joyeux (1980) and Hosking (1981) extended the classical autoregressive integrated moving average (ARIMA) model to autoregressive fractional integrated moving average (ARFIMA) model. This is achieved by allowing the time series differencing operator, typically used to remove polynomial time trends in non-stationary series, to be of a fractional-order ( $0 < d < 1/2$ ) instead of the typical integer derivative. This produces the class of fractionally integrated processes referred to as the ARFIMA(0,  $d$ , 0) model which describes a type of long memory stationary process with a hyperbolic decay of the ACF as compared to the geometric decay for an ARIMA model. This fractional parameter  $d$  captures the persistence for a given long memory model. Other types of long memory may also be explored such as oscillatory long memory autocorrelation decay patterns that arise from Gagenbauer long memory structures (GARMA), of which the ARFIMA model is a special limiting case when oscillation disappears.

Directly developing non-parametric estimators for long memory fractional derivative power  $d$  is not trivial; instead we will work with the estimation of a related concept the Hurst exponent. We will demonstrate the mathematical relationship between Hurst exponent  $H$  and the fractional derivative power  $d$  that will allow us to adopt existing estimators from the statistical literature to detect long memory features in mortality data.

**2.1 Hurst exponent and fractional Brownian motion**

The Hurst exponent  $H$ , also known as the index of long-range dependence, was proposed by Hurst (1951). It is a classical self-similarity parameter that measures the long memory feature in a time series (Millen & Beard, 2003). Since it is robust with few assumptions about the underlying system, it has been widely applied to many fields (Qian & Rasheed, 2004).

In order to proceed to mathematically describe the Hurst exponent, we briefly review the concept of fractional Brownian motion. Mandelbrot & Van Ness (1968) defined fractional Brownian motion

$$W_H = \{W_H(t) : 0 \leq t < \infty, W_H(t) \in \mathbb{R}\}$$

by its stochastic representation given below.

**Definition 2.1 (fractional Brownian motion).** Let  $W_H(t)$  be a random process defined as the solution to following diffusion

$$W_H(t) := \frac{1}{\Gamma(H + \frac{1}{2})} \left( \int_{-\infty}^0 ((t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}}) dW(s) + \int_0^t (t-s)^{H-\frac{1}{2}} dW(s) \right)$$

where the Gamma function  $\Gamma(\alpha) := \int_0^\infty x^{\alpha-1} \exp(-x) dx$ ,  $W(t)$  denotes standard Brownian motion and  $H \in (0, 1)$ .

Fractional Brownian motion of exponent  $H$  is a moving average of  $dW(t)$ , in which past increments of  $W(t)$  are weighted by the kernel  $(t - s)^{H-\frac{1}{2}}$ . The “span of interdependence” between fractional Brownian motion increments can be said to be infinite. In addition,  $W_H(t)$  will reduce to a standard Brownian motion  $W(t)$  when  $H = 0.5$ . A fractional Brownian motion  $W_H(t)$  is Gaussian distributed with stationary increments. Given the initial value  $W_H(0) = 0$  at  $t = 0$ , the expected value and variance are given by

$$\mathbb{E}(W_H(t)) = 0 \text{ and } \text{Var}(W_H^2(t)) = t^{2H}, \text{ for } t > 0$$

respectively. The covariance function derived by Nualart (2006) is given by

$$\text{Cov}_H(t, s) = \mathbb{E} [W_H(s)W_H(t)] = \frac{1}{2} (t^{2H} + s^{2H} - (t - s)^{2H}), \text{ for } 0 < s \leq t$$

An increment of the process in an interval  $[s, t]$  has a normal distribution with zero mean and variance

$$\mathbb{E} [(W_H(t) - W_H(s))^2] = |t - s|^{2H}$$

For  $H = 1/2$ , the covariance  $\text{Cov}_{1/2}(t, s) = \min(s, t)$  and increments of the process in disjoint intervals are independent. For  $H \neq 1/2$ , increments are not independent. Consider a unit increment discrete time process  $\{\Xi_n := W_H(n) - W_H(n - 1), n \in \mathbb{N}\}$ , which gives a Gaussian stationary sequence with unit variance and ACF

$$\begin{aligned} \rho_H(n) &= \frac{1}{2} [(n + 1)^{2H} + (n - 1)^{2H} - 2n^{2H}] \\ &\approx H(2H - 1)n^{2H-2} \rightarrow 0 \text{ as } n \rightarrow \infty \end{aligned}$$

For  $\frac{1}{2} < H < 1$  and large enough  $n$ , we have  $\rho_H(n) > 0$  and  $\sum_{n=1}^\infty \rho_H(n) \rightarrow \infty$ . For  $0 < H < \frac{1}{2}$  and large enough  $n$ , we have  $\rho_H(n) < 0$  and  $\sum_{n=1}^\infty |\rho_H(n)| < \infty$ . Hence, a value of  $H$  in the range  $(\frac{1}{2}, 1)$  indicates long memory in a time series, meaning that a high value in the series will more likely be followed by another high value and such an effect is likely to continue a long time into the future. A value in the range  $(0, \frac{1}{2})$  indicates a time series which is more likely to switch between high and low values in adjacent pairs, and such anti persistence will last a long time into the future. This process can be used to model sequences with intermittency. A value of  $H = \frac{1}{2}$  can indicate a standard Brownian motion which is a short memory process.

**2.2 Relationship between long memory  $d$  and Hurst exponent  $H$**

This section explains the relationship between  $H$  and  $d$ . To understand this relationship, one can show a weak convergence result for the discrete time fractional time series and its convergence to a fractional Brownian motion. To achieve this, we define a fractional difference characterising process

$$Q_t = (1 - B)^{-d} \varepsilon_t, \quad \varepsilon_t \stackrel{\text{i.i.d}}{\sim} N(0, 1), \quad |d| < 1/2$$

then the  $Q_t$  is a stationary and weakly dependent process. Now consider the following scaled partial sum constructed from this process,

$$Z_T(\xi) := \frac{\sum_{t=1}^{[T\xi]} Q_t}{\sigma_T}, \quad 0 \leq \xi \leq 1$$

where  $\sigma_T^2 = \mathbb{E} \left[ \left( \sum_{t=1}^T Q_t \right)^2 \right]$  and  $[\cdot]$  denotes integral part. Using this partial sum, Davidson & De Jong (2000) showed the following theorem.

**Theorem 2.1.** *The following weak convergence holds for all  $|d| < 1/2$  and all  $0 \leq \xi \leq 1$ :*

$$Z_T(\xi) \xrightarrow{d} V_d^{-\frac{1}{2}} W_d(\xi), \text{ as } T \rightarrow \infty$$

with the scale constant

$$V_d = \frac{1}{\Gamma(d+1)^2} \left( \frac{1}{2d+1} + \int_0^\infty ((1+\tau)^d - \tau^d)^2 d\tau \right)$$

chosen to ensure that  $\mathbb{E} [W_d(1)^2] = 1$ . For  $|d| < 1/2$ ,  $W_d(\xi)$  is a fractional Brownian motion with the following representation:

$$W_d(\xi) = \frac{1}{\Gamma(d+1)} \left( \int_{-\infty}^0 \left( (\xi-s)^d - (-s)^d \right) dW(s) + \int_0^\xi (\xi-s)^d dW(s) \right)$$

where  $W$  is the standard Brownian motion.

Now, we can compare this integral representation for  $W_d(\cdot)$  to the standard Weyl integral representation of fractional Brownian motion for  $W_H(\cdot)$  (Definition 2.1) to find the relationship between  $d$  and  $H$ , which is then given by  $d = H - 0.5$ . As a consequence, one can use existing estimators of the Hurst exponent  $H$  to also estimate the long memory parameter  $d$ . In the next section, we will present three examples of such estimators.

### 2.3 Statistical estimators of persistence and anti persistence

Many estimators of long-range dependence have been proposed in the literature. In particular, we take advantages of these different estimators to show that they all can detect long memory structure in human mortality data, obtained from the HMD (Shkolnikov, 2017). We provide overwhelming statistical evidence via three estimators of long memory: R/S, DFA and PR methods from the mortality data sets of many countries. We briefly outline how these estimators are constructed and their properties before illustrating the evidence of long memory in mortality data.

#### 2.3.1 Rescaled range analysis

To measure the intensity of long-range dependence in a time series, the first Hurst exponent estimator was developed by Hurst (1951) in the reranged scale R/S analysis. Given a time series  $Y_{t \in \{1,2,3,\dots,T\}}$ , the sample mean and the standard deviation process are given by

$$\bar{Y}_T = \frac{1}{T} \sum_{j=1}^T Y_j \text{ and } S_t = \sqrt{\frac{1}{t-1} \sum_{j=1}^t (X_j)^2} \tag{1}$$

where the mean adjusted series  $X_t = Y_t - \bar{Y}_T$ . Then a cumulative sum series is given by  $Z_t = \sum_{j=1}^t X_j$  and the cumulative range based on these sums is

$$R_t = \text{Max} (0, Z_1, \dots, Z_t) - \text{Min} (0, Z_1, \dots, Z_t) \tag{2}$$

The following proposition proposes an estimator of  $H$  as derived in Mandelbrot (1975).

**Proposition 2.1.** Consider a time series  $Y_t \in \mathbb{R}$  and define  $S_t$  and  $R_t$  in equations (1) and (2), respectively, then  $\exists C \in \mathbb{R}$  such that the following asymptotic property of the rescaled range  $R/S$  holds

$$[R/S](T) = \frac{1}{T} \sum_{t=1}^T R_t/S_t \sim CT^H, \text{ as } T \rightarrow \infty$$

In addition, for small sample size  $T$ , the rescaled range  $R/S$  can also be approximated by the following equation (Annis & Lloyd, 1976):

$$[R/S](T) = \begin{cases} \frac{T-1/2}{T} \frac{\Gamma((T-1)/2)}{\sqrt{\pi}(T/2)} \sum_{j=1}^{T-1} \sqrt{\frac{T-j}{j}}, & \text{for } T \leq 340 \\ \frac{T-1/2}{T} \frac{1}{\sqrt{T\pi/2}} \sum_{j=1}^{T-1} \sqrt{\frac{T-j}{j}}, & \text{for } T > 340 \end{cases}$$

where the  $\frac{T-1/2}{T}$  term was added by Peters (1994). The  $H$  estimate can be obtained by a simple linear regression

$$\log R/S(T) = \log C + H \log T$$

Hence, we have the following definition for the estimator of  $H$ .

**Definition 2.2 (Estimator  $\hat{H}$  by  $R/S$ ).** The estimator  $\hat{H}$  based on the  $R/S$  analysis is given by

$$\hat{H}_{R/S} = \frac{T(\sum_{t=1}^T \log R/S(t) \log t) - (\sum_{t=1}^T \log R/S(t))(\sum_{t=1}^T \log t)}{T(\sum_{t=1}^T (\log t)^2) - (\sum_{t=1}^T \log t)^2} \tag{3}$$

The empirical confidence interval of  $\hat{H}$  given in equation (3) with sample size  $T = 2^N$  (Weron, 2002) is

$$(0.5 - \exp(-7.33 \log(\log N) + 4.21), \exp(-7.20 \log(\log N) + 4.04) + 0.5)$$

In this paper, we apply the R package called *pracma* to estimate the value of  $H$  adopting the  $R/S$  analysis.

**Remark 2.1.** A drawback of the  $R/S$  analysis is the fact that no asymptotic distribution theory has been derived for the estimated Hurst parameter  $H$ , but we can apply bootstrap methods to find related properties to test for statistical significance of the estimates in order to detect the long memory properties.

### 2.3.2 Detrended Fluctuation Analysis

DFA proposed by Peters (1994) is a powerful tool to test the presence of long memory in a time series and to quantify such long-range dependence. As implied by its name, it was conceived as a method for detrending local variability in a time series which will then provide insight into long-term variations in the data sets. As such, it is a method for determining the statistical self-affinity of a signal and finds wide spread use when analysing time series that appear to exhibit diverging correlation time, such as power law decaying autocorrelation functions. The estimator obtained from DFA may also be applied to signals whose underlying statistics such as the mean and the variance, or dynamics are non-stationary. Typically, it is argued that the reason to employ the DFA method is to avoid spurious detection of correlations that are artifacts of non-stationarity in the time series.



A time series  $Y_{t \in \{1,2,3,\dots,T\}}$  is divided into  $K$  non-overlapped subseries  $Y_{l,k}$  of length  $L$  for  $l = 1, 2, \dots, L$  and  $k = 1, 2, \dots, K$ . Given a new cumulative time series  $Z_{j,k} = \sum_{l=1}^j Y_{l,k}$  with  $j = 1, \dots, L$ , the sample average of the root mean square fluctuation for all  $K$  subseries of length  $L$  is obtained by

$$\bar{F}(L) := \frac{1}{K} \sum_{k=1}^K \sqrt{\frac{1}{L} \sum_{j=1}^L (Z_{j,k} - a_k j - b_k)^2} \tag{4}$$

where  $a_k j + b_k$  is the least squares line  $\tilde{Z}_k(j) = a_k j + b_k$  fitted to points  $(j, Z_{j,k}), j = 1, \dots, L$  (Weron, 2002). The following proposition proved by Taquq *et al.* (1995) has been used to estimate  $H$ .

**Proposition 2.2.** Consider a time series  $Y_t$  and  $\bar{F}(L)$  defined in equation (4),  $\exists C \in \mathbb{R}$  such that the following asymptotic property for  $\bar{F}(L)$  holds:

$$\bar{F}(L) \sim CL^H, \text{ as } L \rightarrow \infty$$

As a consequence of this asymptotic result, the estimated value of  $H$  can be obtained by running a linear regression

$$\log(\bar{F}(L)) = \log(C) + H \log(L)$$

and this leads to an alternative statistical estimator for  $H$ .

**Definition 2.3 (Estimator  $\hat{H}$  by DFA).** The estimator  $\hat{H}$  developed by adopting DFA approach is given by

$$\hat{H}_{DFA} = \frac{L(\sum_{l=1}^L \log \bar{F}(l) \log l) - (\sum_{l=1}^L \log \bar{F}(l))(\sum_{l=1}^L \log l)}{L(\sum_{l=1}^L (\log l)^2) - (\sum_{l=1}^L \log l)^2} \tag{5}$$

The empirical confidence interval of  $\hat{H}$  given in equation (5) with sample length  $T = 2^N$  for  $N \in [10, 50]$  (Weron, 2002) is

$$(0.5 - \exp(-2.33 \log N + 3.25), \exp(-2.46 \log N + 3.38) + 0.5)$$

and the empirical confidence interval for  $\hat{H}_{DFA}$  when  $N > 50$  is

$$(0.5 - \exp(-2.93 \log N + 4.45), \exp(-3.10 \log N + 4.77) + 0.5)$$

**Remark 2.2.** There is no asymptotic distribution theory derived for the DFA statistics. Hence, explicit hypothesis testing can not be performed however again one can apply bootstrap procedures.

### 2.3.3 Periodogram regression

The spectral density function of a general fractionally integrated model with parameter  $d$  is identical to that of a fractional Gaussian noise with Hurst exponent  $H = d + 0.5$ . Hence, PR was proposed by Geweke & Porter-Hudak (1983) to estimate  $H$ . A periodogram of time series  $Y_{t \in \{1,2,3,\dots,T\}}$  with frequencies  $\omega_j = \frac{j}{T}$  and  $j = 1, \dots, [T/2]$  can be represented as

$$I_T(\omega_j) = \frac{1}{T} \left| \sum_{t=1}^T Y_t e^{-2\pi i(t-1)\omega_j} \right|^2 \tag{6}$$



A simple linear regression at low frequencies  $\omega_j, j = 1, \dots, J \leq [T/2]$  can be applied to estimate  $H$  according to the linear model,

$$\log(I_T(\omega_j)) = C - d \log\left(4 \sin^2\left(\frac{\omega_j}{2}\right)\right) + \varepsilon_j$$

where  $C$  is a constant and  $Y_t = \log\left(4 \sin^2\left(\frac{\omega_j}{2}\right)\right)$  (Weron, 2002). Consequently, the least squares estimate of the slope yields another estimator for  $d$ .

**Definition 2.4 (Estimator  $\hat{d}$  by PR).** Consider a time series  $Y_t$  with the periodogram  $I_T(\omega_j)$  defined in equation (6), the estimator  $\hat{d}$  developed by adopting PR is given by

$$\hat{d} = \frac{J \left[ \sum_{j=1}^J \log(I_T(\omega_j)) \log\left(4 \sin^2\left(\frac{\omega_j}{2}\right)\right) \right] - \left[ \sum_{j=1}^J \log(I_T(\omega_j)) \right] \left[ \sum_{j=1}^J \log\left(4 \sin^2\left(\frac{\omega_j}{2}\right)\right) \right]}{J \sum_{j=1}^J \left( \log\left(4 \sin^2\left(\frac{\omega_j}{2}\right)\right) \right)^2 - \left[ \sum_{j=1}^J \log\left(4 \sin^2\left(\frac{\omega_j}{2}\right)\right) \right]^2} \tag{7}$$

with asymptotic distribution

$$\hat{d} \sim N\left(d, \frac{\pi^2}{6 \sum_{t=1}^T (Y_t - \bar{Y}_t)^2}\right)$$

Hence, an estimate of  $H$  using PR can be calculated by  $\hat{H} = \hat{d} + 0.5$  and the 95% empirical confidence interval for  $\hat{d}$  calculated in equation (7) with the sample size  $L = 2^N$  and  $J = [L^{0.5}]$  (Weron, 2002) is

$$(0.5 - \exp(-0.71 \cdot N^{2/3} + 2.04), \exp(-0.68 \cdot N^{2/3} + 1.78) + 0.5)$$

**Remark 2.3.** The PR approach is the only approach with known asymptotic properties.

### 3. Synthetic Studies: Assessing the Properties of Long Memory Estimation

In this section, we design a range of synthetic studies in order to determine if the estimation of long memory fractional derivative parameter  $d$  via the Hurst exponent estimators presented in section 2.3 are accurate. In particular, we are interested in setting up time series with long memory structures in which the attributes of the time series reflect the characteristics typically encountered in mortality data records. All studies in this section involve 1,000 generated sets of time series data trials, which are used for the various studies. More specifically, we are interested in understanding and addressing the following core requirements:

- Mortality data sets at a national level in the HMD are typically relatively short between 80 and 100 time points in length. Therefore, we are interested to determine in a controlled setting how accurately one can detect different strengths of long memory (i.e. different values of  $d$ ) for time series of this length.
- Mortality data are typically aggregated by age groups, or by regions or by gender, etc. Therefore, we are interested to understand if the aggregation of individual count time series generated with long memory structures will maintain long memory structure.
- Mortality data often can contain missingness and outliers in the time series records; therefore, we are interested to see what influence this has on the robustness of our different classes of long memory estimators.

In the following synthetic case studies, we evaluate the performances of R/S, DFA and PR methods in the estimation of  $H$ . In order to produce synthetic data records with desired known long

**Table 1.** Estimate  $H$  by R/S, DFA and PR methods from the data generated by the ARFIMA model with  $d = 0.45$  and  $n = 50, 100$  and  $300$ . Empirical confidence intervals at 95% were found to be less than 5% in deviation from the point estimator in all cases

$H = d + 0.5$			
$d = 0.45$	$n = 50$	$n = 100$	$n = 300$
R/S	0.954	0.951	0.951
DFA	0.946	0.952	0.958
PR	0.955	0.945	0.944

memory structure, we simulate data via the ARFIMA model with different values  $d$  and different sample sizes  $n$ .

First, we study the effect of the length of the data on the ability to estimate long memory. Second, we study the strength of persistence on the ability to detect weak and strong persistence (i.e. weak and strong long memory). Third, we study the influence of missing data on the ability to detect long memory. As in mortality settings, there are often missing data in some age group or year or country or gender. We also study the effect of aggregation and transformation of mortality data because people often collect data stratified by single age group, but study them by 5 years or 10 years stratification. Lastly, we study the effects of both scaling and rounding off to measure the sensitivity of these estimators.

We start this section by noting that in the accompanying supplementary appendix, hosted as Supplementary Material on the AAS website, we provide results for preliminary analysis of the performance of the three estimators RSA, DFA and PR obtained using known model parameters and synthetic data for a range of increasing  $d$  parameter and sample size case studies.

### 3.1 Influence of sample size on detecting long memory

In real data, for different countries, the mortality data sets often vary in length and are typically relatively short between 80 and 120 years of samples. It is, therefore, important to test the sensitivity of the Hurst exponent estimators to sample length in order to make sure that the strength of long memory can be measured accurately for settings of relevance to mortality modelling.

Given  $d = 0.45$ , Table 1 shows that these three methods are not sensitive to the sample size when operating in a range of 50–300 samples. This is good news as it means one should be able to utilise such estimators for real mortality data applications.

### 3.2 Influence of long memory strength on accuracy of long memory estimation

For different mortality data sets, the strength of long memory can vary from very weak (namely,  $H$  is around 0.6 or  $d$  is round 0.1 because  $H = d + 0.5$ ) to very strong (namely,  $H$  is around 0.9 or  $d$  is round 0.4). The mortality data with strong long memory intuitively is easy to estimate because the long memory feature is more pronounced. Consequently, we need to test if these estimators are also sensitive when the long memory feature is not significant.

Given sample size  $n = 150$  that is similar to the length of real world mortality data, Table 2 shows that these three methods are able to detect the different strength of long memory.

### 3.3 Influence of missing data in estimation of long memory

Missing data commonly arises in mortality data sets for death counts by country, gender and age. Therefore, it is practically important to see if the three estimators proposed are sensitive to missingness in the time series records. In practice, there may be varying types of missingness mechanism that could arise in real data for different countries, genders and age groups over

**Table 2.** Estimate  $H$  by R/S, DFA and PR methods from the data generated by the ARFIMA model with  $n = 150$  and  $d = 0.15, 0.25$  and  $0.35$ . Empirical confidence intervals at 95% were found to be less than 5% in deviation from the point estimator in all cases

$H = d + 0.5$			
$n = 150$	$d = 0.15$	$d = 0.25$	$d = 0.35$
R/S	0.652	0.753	0.862
DFA	0.647	0.749	0.862
PR	0.658	0.739	0.869

**Table 3.** Estimate  $H$  by R/S, DFA and PR methods from the data generated by the ARFIMA model with  $n = 150$ ,  $d = 0.45$  and 5%, 10% and 30% missing values. Empirical confidence intervals at 95% were found to be less than 5% in deviation from the point estimator in all cases

$H = d + 0.5$			
$n = 150, d = 0.45$	Missing 5%	Missing 10%	Missing 30%
R/S	0.954	0.957	0.964
DFA	0.948	0.944	0.958
PR	0.946	0.945	0.945

time. For instance, these may include the non-exhaustive list of censoring, truncation or Missing Completely at Random (MCAR), Missing at Random (MAR) through to Missing Not at Random (MNAR). In general, we will not know exactly which of these mechanisms arose to produce the missingness observed for each country and each age group and furthermore, we will not be sure that only one type of missingness is present in real data, most likely it will be a combination of different missingness mechanisms. As a consequence, in the analysis in this case study, we chose to explore the simplest case to produce, in which data are missing at random.

Table 3 tests the sensitivity of the three estimation methods on a different percentage of missing values. The missing value data sets are generated by randomly dropping a portion of the data simulated by the ARFIMA model. Under the settings of  $n = 150$  and  $d = 0.45$ , the accuracy of these three methods to detect long memory strength with different amount of missing values does not decrease.

### 3.4 Influence of aggregation on long memory structure

It is often the case that mortality data are aggregated when it is studied at the national level. For instance, one may typically consider aggregated total death count data time series for multiple age groups combined together. For example, single age group mortality data  $\{0, 1, 2, 3, 4\}$  can be aggregated into  $\{0 - -4, 5 - -9, 10 - -14, 15 - -19\}$  or  $\{0 - -10, 11 - -20, 21 - -30, 31 - -40, 41 - -50\}$ . Table 4 shows that different aggregations do not change the long memory feature which can be detected by these three methods.

### 3.5 Influence of quantisation of long memory time series data

It is often the case that one will quantise to a given unit in mortality data. Therefore, it is interesting to study if such quantisation transformations of data demonstrating long memory will significantly affect the ability to detect the long memory features in the underlying un-quantised data.

For example, the number 123 rounded off by one decimal place is 120, and the number 123 rounded off by two decimal places is 100. Table 5 shows that these three methods are not sensitive to different decimal rounding off.

**Table 4.** Estimate  $H$  by R/S, DFA and PR methods from the data generated by the ARFIMA model with  $n = 150$ ,  $d = 0.45$  and aggregate every 5 and 10 values. Empirical confidence intervals at 95% were found to be less than 5% in deviation from the point estimator in all cases

$H = d + 0.5$			
$n = 150, d = 0.45$	No aggregation	Aggregate every 5	Aggregate every 10
R/S	0.952	0.952	0.958
DFA	0.952	0.956	0.960
PR	0.954	0.957	0.959

**Table 5.** Estimate  $H$  by R/S, DFA and PR methods from the data generated by the ARFIMA model with  $n = 150$ ,  $d = 0.45$  and rounding off one and two decimal places. Empirical confidence intervals at 95% were found to be less than 5% in deviation from the point estimator in all cases

$H = d + 0.5$			
$n = 150, d = 0.45$	No rounding	Rounding off one decimal place	Rounding off two decimal place
R/S	0.952	0.952	0.953
DFA	0.952	0.950	0.954
PR	0.954	0.953	0.955

**Table 6.** Estimate  $H$  by R/S, DFA and PR methods from the data generated by the ARFIMA model with  $n = 150$ ,  $d = 0.45$  and scaling every 10, 100 and 1,000. Empirical confidence intervals at 95% were found to be less than 5% in deviation from the point estimator in all cases

$H = d + 0.5$			
$n = 150, d = 0.45$	Scaling by 10	Scaling by 100	Scaling by 1,000
R/S	0.952	0.953	0.952
DFA	0.952	0.952	0.951
PR	0.954	0.954	0.954

### 3.6 Influence of scaling long memory data

For computation purpose, mortality date may be scaled by a certain amount. It is important to test if the long memory feature can be disturbed by the different amount of scaling. For example, 1,000 can be scaled into 100 by 10. Table 6 shows that these three methods are not sensitive to scaling.

## 4. Evidence that Long Memory Features in National-Level Mortality Data are a Stylised Fact

In this section, we demonstrate the prevalence of a long memory feature in the death count data from four perspectives by studying the pattern of long memory across different countries, age groups, genders and time periods. To measure the strength of long memory, we estimate  $H$  using all three estimation methods presented (R/S, DFA and PR) and we only report the results of the R/S in section 2.3.1 since the other two methods provide very similar results, and this is evidenced in the hosted Supplementary Material on the AAS website.

### 4.1 Mortality data

We considered mortality data sets of 16 countries, obtained from the HMD (Shkolnikov, 2017), which provides detailed mortality and population data to the public. Table 7 reports the data length for each country together with their abbreviated name.

**Table 7.** Data length (years) and abbreviation for country names

Full name	Australia	Belgium	Canada	Denmark	Finland	France	Iceland	Italy
Abbrev.	AU	BE	CA	DK	FI	FR	IS	IT
Data length	94	170	91	180	138	199	176	141
Full name	Netherlands	Norway	Spain	Sweden	Switzerland	UK	USA	Japan
Abbrev.	NL	NO	ES	SE	CH	GB	US	JP
Data length	163	169	107	264	139	92	83	68

**Table 8.** Estimated Hurst exponent for three types of aggregation. Age group represents different aggregation for different types, for example, age group 2 for type 1 is 1, for type 2 is 1–4 and for type 3 is 11–20. Empirical confidence intervals at 95% were found to be less than 5% in deviation from the point estimator in all cases

Aggregated types	Age group 1	Age group 2	Age group 3	Age group 4	Age group 5
Type 1	0.7905	0.7954	0.7905	0.7901	0.7882
Type 2	0.7905	0.7942	0.7825	0.7888	0.7713
Type 3	0.7913	0.7855	0.8089	0.8017	0.7731

Each data set contains the number of death  $y_{x,t}^*$  cross-classified by genders and years and they are aggregated into 21 age groups, namely age 0 group, age 1–4 groups, age 5–10 groups, age 11–15 groups, etc. The  $y_{x,t}^*$  are scaled according to the following equation (8):

$$y_{x,t} = \lceil y_{x,t}^* / c_x \rceil \quad (8)$$

where  $\lceil \cdot \rceil$  represents rounding to the nearest integer, and the scaling factor  $c_x$  is determined by

$$c_x = 10^k \text{ if } \min y_{x,t}^* \in (10^{k+1}, 10^{k+2})$$

such that the range of  $y_{x,t}$  is approximately within (0,100) and  $k$  is an integer selected to achieve this. This re-scaling of data only changes the unit and preserves the features.

Furthermore, we find comprehensively, statistically significant evidence for the presence of long memory features which is present irrespective of how the data sets are aggregated or disaggregated. To illustrate this, we take Australia female death count as an example. There are different ways to aggregate age groups. For example, taking five age groups for each type, one can form single age  $\{0, 1, 2, 3, 4\}$  for type 1 or alternatively  $\{0, 1-4, 5-9, 10-14, 15-19\}$  for type 2 or  $\{0-10, 11-20, 21-30, 31-40, 41-50\}$  for type 3. Table 8 shows that the estimated  $H$  for the three types of aggregation are similar. This feature, also being tested in the synthetic study in section 3.4, is important to demonstrate since, unlike weak serial correlation in a time series, which may be diminished or removed via aggregation, stronger persistence in the form of long memory prevails as a structural feature in aggregation typically.

We note that since the three estimators for  $H$  provide similar estimates as demonstrated in all synthetic studies in section 3, we report only estimator R/S in all subsequent real data analyses.

#### 4.2 Long memory pattern across age groups, gender and countries

The first finding we highlight is that all of the 16 countries studied demonstrate some degree of long memory across all age groups. We provide statistical evidence of this data feature in each countries' mortality data via a heat map depiction, given in Figure 1, which displays the estimated  $H$  for females and males in 21 ages across 16 different countries. From this analysis, a few key observations are discernible from the statistical data structures captured by these feature summaries.

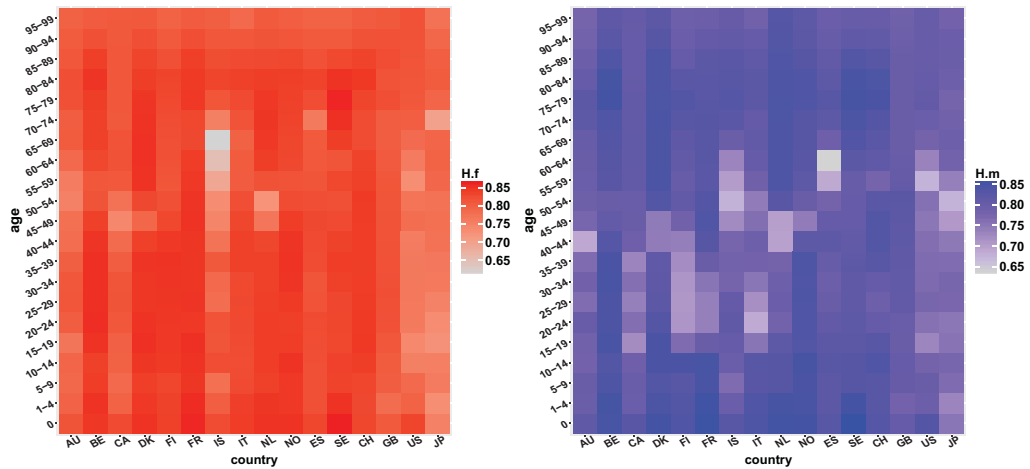


Figure 1. Heat map of estimated  $H$  across countries and age groups for female (left) and male (right).

First, there seems to be a clear gender effect. For instance, for male populations between ages 15 and 64, the estimated  $H$  is lower than females in most of the countries analysed with the exceptions of Belgium, Sweden, Switzerland and the UK.

Second, for the US and Japan, both female and male mortality features show generally weaker mortality persistence than other countries. This is interesting since the US and Japan have historically vastly different populations in terms of migration patterns and ethnic diversity that may be expected to affect persistence in a population's mortality experience. We conjecture that in the case of the US perhaps, this feature can be explained by the multicultural and diverse population that has arisen over multiple decades of immigration that produces diverse cohort structures in the mortality data of the US. For Japan, which has had very little immigration relative to the US, one possible explanation is that the length of mortality data available on public record is too short to provide enough information to estimate  $H$  in the nonparametric setting. This is generally because the records are only available publicly from the 60s, making it by far the shortest data set studied, and we suspect this small sample size makes it difficult to accurately estimate the persistence.

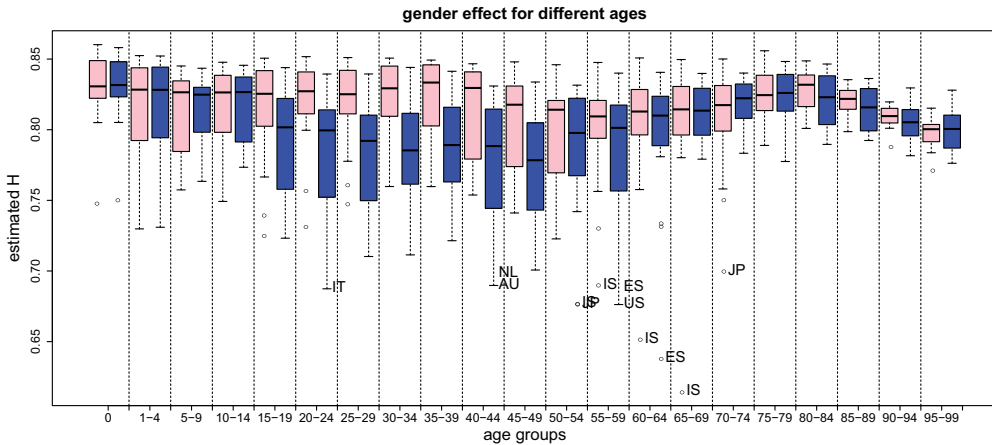
There also seems to be an exception for Icelandic females, compared with other countries, the mortality persistence (estimated  $H$ ) is relatively lower between ages 25 and 74. It may be a result of a smaller population size such that the mortality process becomes incapable to sustain a stable mortality feature consistently over time.

### 4.3 Long memory pattern across age group by gender

After an overview of how the three attributes of age, gender and country interact in the long memory patterns, we establish as present in the mortality data, we then investigate the factors of age and country separately. In this section, we look at age effect by gender on the long memory pattern after aggregating over countries.

Figure 2 below displays the estimated  $H$  from both female and male groups, respectively, in different ages across 16 countries. This perspective of the long memory patterns in mortality data clearly demonstrates that there is a trend variation in mortality experienced by age group, which is generally universal for most countries and consistent in marked median age term structure trend variation between the sexes.

For instance, it was not a surprise to observe 2 key features: in early age groups in developed countries such as those studied in this collection of 16 countries, we expect fairly consistent mortality experience by each country and by gender as birth and early childhood medical practice is



**Figure 2.** Boxplot of estimated  $H$  across age groups aggregated over countries to show gender effect. For each age group, the first box plot is female (pink) and the second is male (blue).

very advanced, leading to fairly consistent patterns in infant and early childhood mortality, and subsequently strong persistence in historical time series of mortality in these age groups that is in agreement across all countries with lowest fluctuations by country.

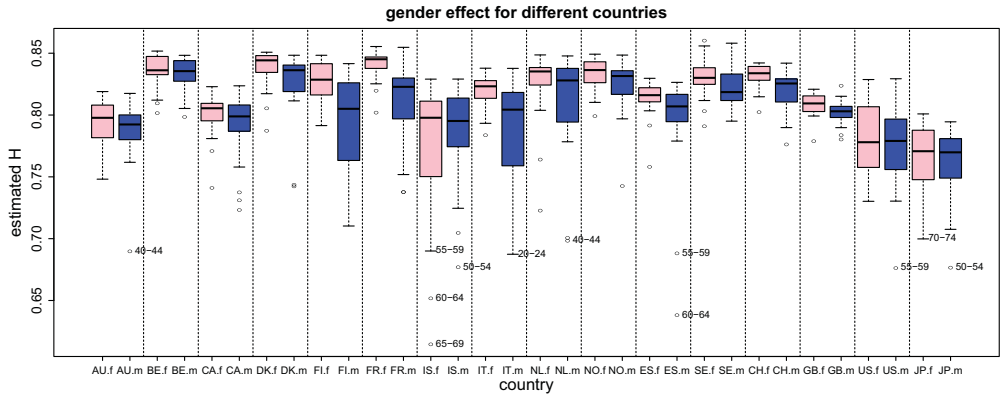
We also expect that at older age groups as we just pass the typical life expectancy for males and females in developed countries, then one would expect again less fluctuations between countries and stronger persistence in mortality experience as rates of mortality significantly increase year on year for such subpopulations. This variation is also expected to be more pronounced in males than females as statistically, they tend to die earlier than females and so smaller sample sizes in increasing age groups past age 85 lead to wider uncertainty.

We also observe that for the senior age groups ranging between 80 and 99, there is a consistent downward trend of estimated  $H$  for both genders. This feature is due to the fact that the mortality persistence is more difficult to detect in a smaller population of seniors. Furthermore, such population sizes in this age bracket have rapidly changed throughout the period of analysis of this time series for such age groups, which is evident if one observes the change in life expectancy over time, see an example of this change in the England and Wales population in the online technical appendix accompanying this paper hosted as Supplementary Material on the AAS website. There is also less variation in estimated  $H$  across countries because the health deterioration for seniors is a common feature for all countries.

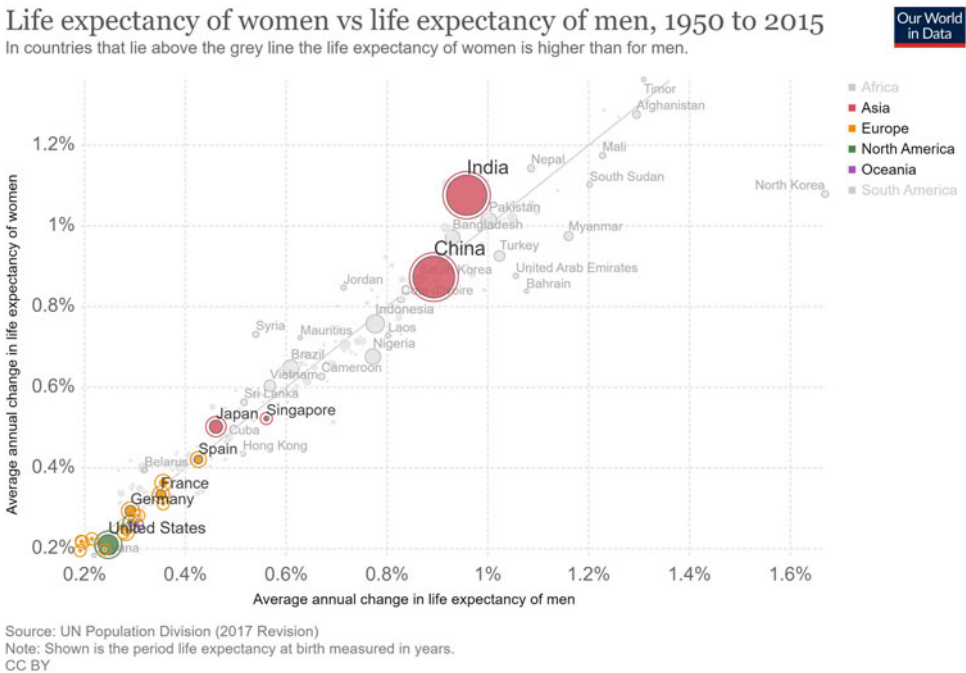
What is most interesting in this analysis is what we observe in the age groups between 15 and 64 years of age. Several important statistical data features are present, foremost of these is that we observe a marked difference in male versus female mortality persistence. The median global persistence in mortality for males is noticeably lower than female mortality at all age groups in this bracket of 15–64 years. Second, the mortality persistence in female age groups is significantly skewed towards high persistence structure with lower variation in such statistical features compared to males in each age group when compared globally.

The way to interpret this is as follows, it means that the mortality patterns in females in these age groups display a very consistent year on year autocorrelations even at very long time lags. For instance, practically, it is analogous to stating the female mortality in these age groups has much stronger second order autocorrelation relationships over the decades compared to males globally. Such a stylised feature can now be incorporated to improve mortality projection, it is clear that based on these findings models of mortality for males and females and in different age bands should be enhanced to incorporate these differences in long memory persistence patterns in order to better improve the mortality forecasting and life table construction.





**Figure 3.** Boxplot of estimated  $H$  across countries aggregated over ages to show gender effect. For each age group, the first box plot is female (pink) and the second is male (blue).



**Figure 4.** Source: Max Roser and Esteban Ortiz-Ospina (2019) Life expectancy by sex. Published online at OurWorldInData.org. Retrieved from: <https://ourworldindata.org> [Online Resource].

We do, however, note that this significant difference, for the data set of 16 countries considered here cannot, however, be simply explained by the supposition that males annual changes in life expectancy differ significantly to women so much so that the additional variation would reduce the long memory persistence, at least not in the recent times of 1950–2019, as indicated in the plot of global female versus male historical average annual change in life expectancy per year by country, see Figure 4. In this plot, we see that though the rates of change vary significantly year on year for some countries in our sample ranging from Iceland with around 0.2% average annual changes in life expectancy for females and males compared to Japan with a slight tendency towards higher rates of change in life expectancy and noticeably higher rates of change for females compared to males over this time window. These country differences may be indicative of the variations in

persistence seen globally and may partially account for why Japan is a notable exception for female mortality at ages 50–74 for females versus males.

Another possible reason to account for some differences here could clearly relate to wars in the early part and middle of last century which clearly would have produced a cohort effect which is more pronounced in male populations in those eras in which they took place. Analysis of this component is provided in detail in the online technical appendix accompanying this manuscript and hosted as Supplementary Material on the AAS website.

#### 4.4 Long memory pattern across countries by gender

Figure 3 below compares the estimated  $H$  across 16 countries by gender aggregated over 21 age groups, which indicate mortality persistence over the years. Again, the estimated  $H$  has lower values and wider spreads for male populations. The exceptions of Iceland, Australia and Japan are marginal. This difference in the spread is most apparent for Finland, France, Italy and Netherlands. Some countries like Belgium and Denmark display a high level of mortality persistence over the years, whereas Iceland, US and Japan typically exhibit lower levels of mortality persistence.

Furthermore, the results show that the spread of  $H$  is wider for male than female in Finland, France, Italy, the Netherlands and Spain. A possible reason is the effect of the two world wars which affect the mortality of European male more. The next section studies the effect of wars in more details. Again, the spread of  $H$  for Iceland is more, and the spread is wider for female than male.

## 5. Multiple Timescales of Long Memory: Multifractal Persistence

In the following section, we aim to demonstrate how to extract a functional characterisation of long memory. These functional characterisations will be available for each gender, country and age groups death count processes. As such, we will demonstrate how to construct a class of functional features that we argue can act as a discriminatory signature for mortality time series data by gender and country to aid classification and forecasting analysis.

The resulting functional features will be curves constructed from an extension of mono-fractal long memory persistence analysis to multifractal long memory persistence. In practical terms, this is analogous to multiple timescale analysis, even though we only have annual data. This allows us to study differences in patterns of long memory at different time resolutions, which we show is useful for feature based cluster analysis and forecast model design.

### 5.1 Motivation for multifractal persistence functional feature extraction

In previous sections, we have studied what is known as mono-fractal long memory analysis. It assumes that the presence of long memory exists to the same degree of strength at all timescales of the time series data, and there is a scalar valued parameter  $H$  (or  $d$ ) characterising the strength of such persistence. For instance, the data generating process, in our case death counts would not differ in its persistence properties at timescales of daily, weekly, monthly, quarterly, annually, biannually, etc. In other words, the same long memory strength of persistence would be present at all timescales in a mono-fractal process. However, in practice, this is often not the case, and in fact, often long memory properties may either vary over time or they can vary over timescales or time resolutions.

In this section, we study the generalisation of mono-fractal long memory to multifractal long memory in which one assumes that the data generating process will admit persistence properties in which the strength of long memory persistence is allowed to vary as a function of the timescale or resolution, now we have a functional parameter or curve for  $H(\cdot)$  to estimate versus resolution.

In other words, now we allow for the situation where the strength of long memory may depend on the timescale that you observe the process, it can either be constant, reducing back to the mono-fractal case studied or it can be allowed to vary over timescales.

Intuitively, in practice such an analysis is equivalent to a generalisation from mono-fractal persistence analysis to a setting in which different strengths of persistence may be exhibited by a process depending on the time sampling that it is observed. For instance, daily data, weekly data, monthly data and annual data records do not necessarily have to demonstrate the same persistence properties. To accommodate this fact, one often employs feature extraction methods related to multifractality, i.e. the presence of different persistence strengths for different timescales.

The unique thing about multifractal analysis as it is performed in this section is that it only requires the observation of the data set at a single sample rate, i.e. in the mortality setting of the HMD data these are annual data for death counts. However, even though the data are observed annually, the methods in this section still allow one to extract feature information related to the variation of persistence strength versus timescale. This can be extracted from a time series at a fixed sample rate, but we can still learn about persistence at different sampling rates from this time series. The resulting long memory characterisation is a functional feature extraction, where instead of estimating for a given death count time series a single Hurst exponent, we now extract a functional curve of Hurst exponent versus timescale.

The reason why such generalised feature extraction is of relevance is that Hurst exponent functional features obtained from the multi-timescale multifractal analysis can be very informative regarding persistence properties over multiple timescales. This is important for both forecast modelling as well as feature based classification of multi-population mortality experience by age and gender. We have included a detailed analysis of how such functional features can be used for classification of mortality experience by age and country in the following sections.

## 5.2 Multifractal functional Hurst exponent curve via extensions of DFA

In the mono-fractal analysis, we effectively explored the presence of scale invariant structure in the time series, where the persistence structure repeats itself on subintervals of the mortality time series. Formally, the mortality time series, say generically  $Y(t)$  are scale invariant when  $Y(ct) = c^H Y(t)$ .

In this section, we explore if different scaling behaviours can be observed for many interwoven fractal subsets of the time series, i.e. if long memory or persistence varies by observation time, i.e. quarter yearly, semi-annual, annual, biannual, decadal, etc. If such variations exist at these different sampling rates, we will require a multitude of scaling exponents for a full description of the scaling behaviour, and a multifractal analysis can be utilised, we will focus on multifractal DFA (MF DFA), the natural multi-resolution extension from the DFA analysis previously presented.

Recall, in mono-fractal DFA estimation of scalar  $H$ , we assume the fluctuation asymptotic relationship characterised for a time series  $Y_t$  and fluctuation  $\bar{F}(L)$ , which is given by

$$\bar{F}(L) \sim CL^H, \text{ as } L \rightarrow \infty$$

for scalar valued Hurst exponent  $H$ .

In Kantelhardt *et al.* (2002) they discuss two types of multifractal process, we will be exploring the case of Type I multifractality. The concept of multifractal DFA is to consider extending the asymptotic fluctuation relationship to a power scaling function rather than a single scalar exponent, as detailed by

$$\bar{F}_q(L) \sim CL^{H(q)}, \text{ as } L \rightarrow \infty$$

where the long memory power exponent that dictates the strength of autocorrelation is now transformed to a function  $H(q)$ . Here, the argument  $q$  is used to modify the norm of the fluctuation calculation as follows:

$$\bar{F}_q(L) := \left\{ \frac{1}{4K} \sum_{k=1}^{2K} [\bar{F}^2(L)]^{q/2} \right\}^{1/q} \quad (9)$$

which uses the previously defined fluctuation sample average of the root mean square fluctuation for all  $K$  subseries of length  $L$ . The value of  $q$  can take any real value except zero, where as  $q \rightarrow 0$  one must switch to a logarithmic averaging procedure to ensure the averaging procedure is well defined, as given by

$$\bar{F}_0(L) := \exp \left\{ \frac{1}{4K} \sum_{k=1}^{2K} \ln [\bar{F}^2(L)] \right\} \quad (10)$$

This quantity is calculated repeatedly for all timescales  $L$  to determine the relationship between  $\bar{F}_q(L)$  and  $L$ . Typically,  $\bar{F}_q(L)$  is an increasing function of  $L$ . The steps to the estimation are then analogous to those already discussed, with the modified fluctuation and a function  $H(q)$  is then obtained, typically over a grid of values of  $q$  and then a smooth interpolation is applied to produce the function  $H(q)$ , see details of the steps to the algorithm for estimation in Kantelhardt *et al.* (2002).

We note that one may interpret the value of  $q$  as varying the resolution at which one is looking at the persistence in the time series, analogous to applying mono-fractal analysis on different sampling rates, hence why such a functional characterisation is commonly referred to as multi-timescale or multi-resolution long memory characterisation.

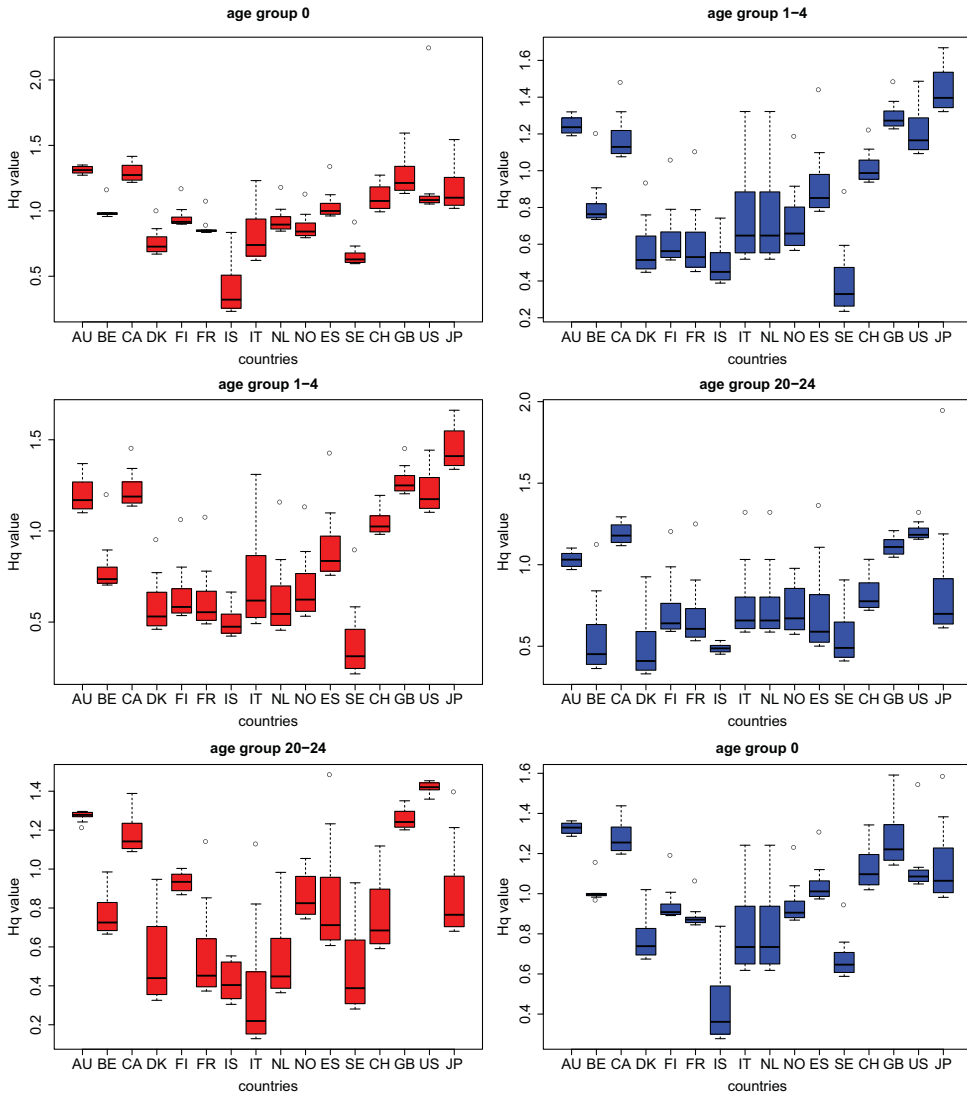
Furthermore, it is useful to note that  $q = 2$  corresponds to the mono-fractal analysis in the first part of the manuscript. We utilise the {MF DFA} package in R that allows one to perform the estimation of the function  $H(q)$  for mortality data. We confirm its performance for various synthetic data studies to ensure it is accurately performing the estimation of  $H(q)$ . We briefly outline a small selection of these synthetic studies in the accompanying online technical appendix hosted as Supplementary Material on the AAS website.

## 6. Multi-Timescale Long Memory Features in National-Level Mortality Data

In this section, we apply the MF DFA procedure to mortality data analogously to the studies done previously in a mono-fractal setting. We provide just a selection of all the results produced to provide evidence for the utility of this method as a characterising signature for long memory patterns in mortality by age and country and gender. To assist the reader with interpreting the next sets of results, we provide the following guidance on understanding the plots. Each box plot represents the multifractal signature of the mortality experience of the population sample labelled. This is obtained by fitting for a grid of  $q$  values between 1 and 10 and obtaining from the MF DFA estimation procedure estimates of  $H(q)$ , each estimate is combined for that time series into the signature box plot for that time series.

### 6.1 MF DFA estimated Hurst exponent curves by ages, gender and country

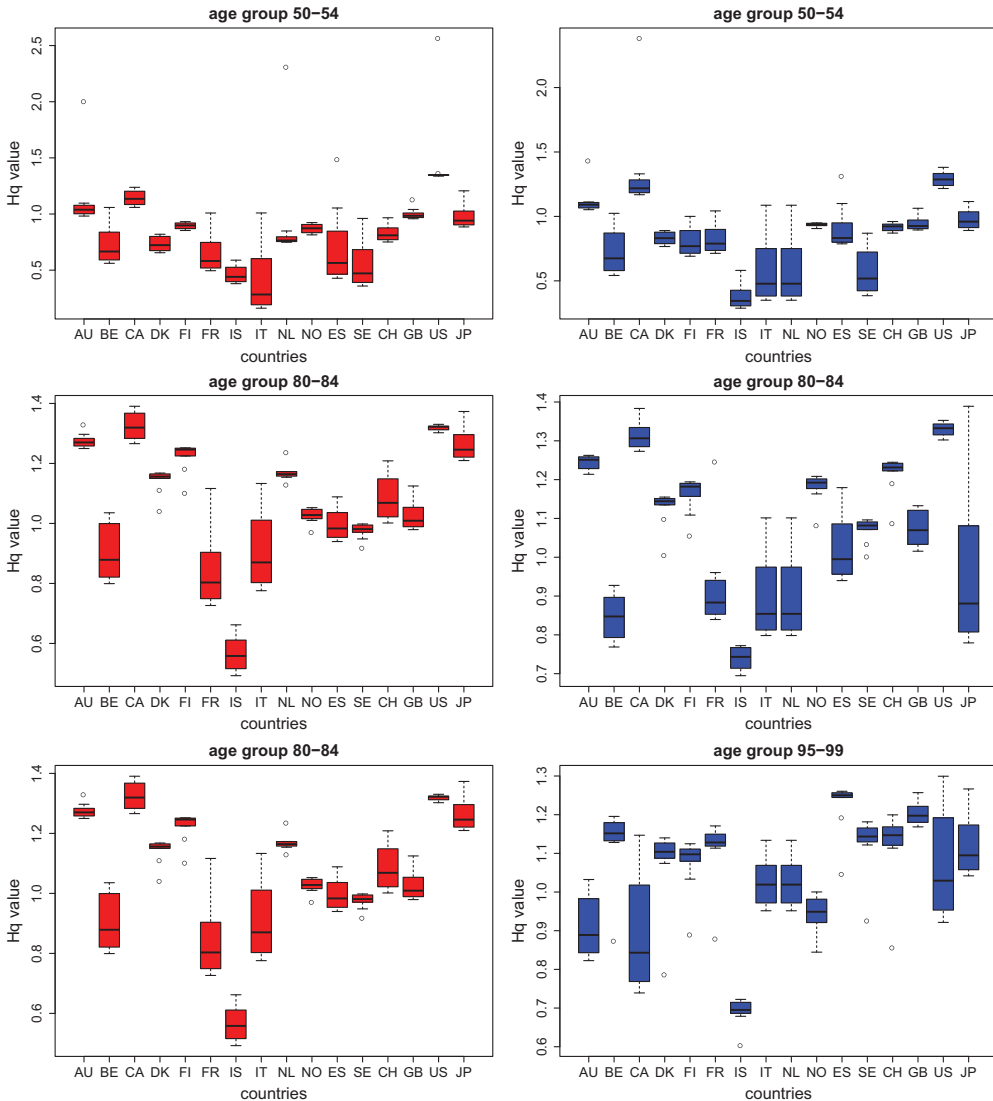
Figures 5 and 6 below compares the estimated  $H(q)$  by MF DFA across the same 16 countries as discussed earlier, again split by gender, country and age. In this case, the box plot is calculated over the estimators of  $H(q)$  for a range of  $q$  values from 1 to 10. Therefore, each box plot represents a summary of the Hurst exponent curves  $H(q)$  evaluated at values of  $q \in \{1, \dots, 10\}$  corresponding to varying time resolutions. Furthermore, for ease of presentation, the results have been displayed over a selection of six representative age groups characterising patterns observed for young, middle and elderly aged.



**Figure 5.** Boxplots of estimated  $H$  by MFDFA across age groups aggregated over countries to show gender effect. For each age group, the left panel is female (pink), and the right panel is male (blue).

It is important to consider how to read and interpret the data structures and meaning from these signature multifractal data summaries. First, we recognise that the box plot signatures displays the strength of long memory, i.e.  $H(q)$ , over a range of different timescales such as yearly, biannual, 5 yearly and decadal as obtained analogously by varying the resolution parameter  $q \in \{1, \dots, 10\}$ .

Second, one can consider the median on the  $y$ -axis which gives the best estimate for  $H(q)$ , which would be a mono-fractal long memory analysis at a given time resolution and represents the most prevalent value for the strength of long memory across multiple timescales. Consequently, if the box plot is highly compacted such that the interquartile range is very tight, it means the time series is most likely displaying mono-fractal long memory properties. Conversely, if the interquartile range of the box plot is rather wide, then this time series is displaying multiple timescales of long memory and is therefore multifractal in nature.



**Figure 6.** Boxplots of estimated  $H$  by MFDFA across age groups aggregated over countries to show gender effect. For each age group, the left panel is female (pink), and the right panel is male (blue).

Third, if the box plot is symmetric around the median, it means that typically the short timescale persistence is as likely to be present in the mortality time series under study as larger timescale long memory effects. If the box plot signature is right skewed, then the mortality experience has more pronounced small timescale long memory effects compared to larger timescale long memory persistence. Conversely, if the box plot is left skewed then the opposite is true, where the long timescale effects of long memory would be more prevalent in the mortality time series.

We observe several interesting signatures in the mortality data by country and by age which we now summarise from these plots. First, all data sets demonstrate stronger evidence for long memory on short timescales rather than longer timescales since in all case there is significant right skew in the box plots. Second, we see that there is more pronounced multifractal long memory structure in males compared to females as typically the male box plots have markedly wider interquartile ranges in all age brackets. Third, in the 16 countries in the study, one can observe that

they can largely be separated into two groups according to these MF DFA signatures: mono-fractal patterns versus multifractal patterns. For instance, Australia, Canada and the US almost universally have very compact interquartile ranges indicating mono-fractal structure and incidentally they have a relatively strong long memory. Conversely, Italy, Iceland and the Netherlands tend to have relatively very wide interquartile ranges suggesting they have mortality dynamics with multiple timescales of long range dependence. Some additional points are worth noting, Japans elder population of males clearly has a very wide interquartile range, however, this is most likely caused by significant variance increase due to reduced population size in males over the age of 80 versus females.

## 7. Using Multifractal Mortality Hurst Exponent Curves for Population Mortality Clustering

When understanding variations in population dynamics, it is useful to consider if there are any groupings or patterns that may arise in national-level mortality data by age, by gender or by country. If one can find clustering tendency in mortality data, in other words, demonstrate that there is the dissimilarity between particular groups of mortality data, say collections of countries or collections or age groups for a given gender. Then such a finding would assist in further actuarial practice when designing longevity products as well as in performing model development for forecasting mortality, where the models are tailored to particular groups of data with similar statistical characteristics. In this section, we apply advanced versions of functional data analysis curve based clustering to the multifractal curves for  $H(q)$  obtained for each country, each age group and each gender. We will adopt an advanced machine learning approach to functional clustering using kernel k-means, which is a Hilbert space clustering technique. The specification of kernel k-means is provide for the interested reader in the accompanying technical appendix hosted as Supplementary Material on the AAS website.

Kernel k-means is an extension of the standard k-means clustering algorithm that splits the data objects into a required number of preselected clusters amounts (k-clusters) in order to optimise a certain objective function based on a measure of similarity (Peters *et al.*, 2017). In this study, we apply a weighted version of kernel k-means which can cluster data objects using a non-uniform weight to denote its likelihood of occurrence. This approach provides diversity and stability in this cluster assignment, see Filippone *et al.* (2008) for a detailed tutorial exposition of this class of machine learning method.

### 7.1 Application of kernel K-Means clustering to Hurst exponent curves mortality data features

In the following study, we apply kernel k-means on the estimated features of the Hurst exponent curves  $H(q)$  by MF DFA method to classify the long memory features of different mortality data sets across countries, age groups and genders. For the MF DFA feature extraction method, the time resolution scales are chosen for q-order time resolutions from  $-10$  to  $10$ . The kernel k-means method clusters these estimated vectors of  $H(q)$  for different age groups and countries.

We tested the cluster behaviour for different counts of cluster and found a reasonable balance between the cluster quality, the within cluster homogeneity and the between cluster heterogeneity occurred when  $k = 3$  clusters was selected. According to the Figure 1, we demonstrate that when clustering with three clusters, this lines up well with the empirical findings from earlier in the paper that there exists three types of long memory in national-level mortality data, namely, short memory, weak long memory and strong long memory, labelled by cluster labels 1, 2 and 3, respectively. The kernel applied in this study is Gaussian Radial Basis kernel function, selected for the fact that it is a universal kernel. Universal kernels produce a very rich Reproducing Kernel Hilbert Space (RKHS) to perform the functional clustering in the implicit feature space capturing the Hurst exponent curve groupings, see Micchelli *et al.* (2006).



**Table 9.** Kernel k-means clustering assignments for ages 10–14, 35–39 and 70–74 across 16 countries

Age group	AU	BE	CA	DK	FI	FR	IS	IT	NL	NO	ES	SE	CH	GB	US	JP
Female																
10–14	3	3	2	3	2	3	1	1	3	2	1	1	3	2	2	2
35–39	2	1	2	1	2	2	1	3	1	3	1	3	3	3	2	1
70–74	2	3	2	2	3	3	1	1	3	2	1	3	3	3	2	1
Male																
10–14	2	2	2	2	2	1	1	3	3	2	1	1	2	3	3	3
35–39	3	3	1	1	1	3	1	2	2	1	2	2	1	3	2	3
70–74	2	3	1	3	3	2	3	2	2	3	1	2	1	1	3	3

**Table 10.** Kernel k-means clustering assignments for AU, IS, UK, US and JP across all age groups

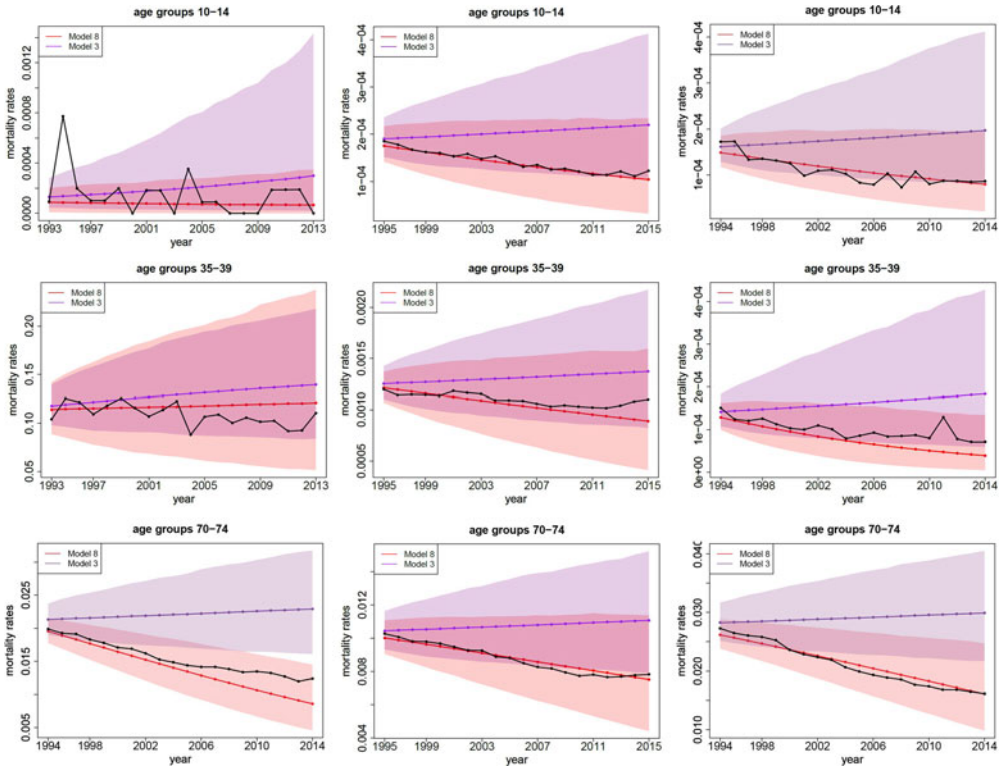
Female												
Country	0	1–4	5–9	10–14	15–19	20–24	25–29	30–34	35–39	40–44	45–49	
AU	3	3	1	1	1	3	3	2	1	1	1	
IS	1	1	3	2	2	2	3	2	1	1	1	
UK	3	1	3	2	3	1	1	1	1	2	3	
US	3	2	2	1	3	3	3	3	3	1	3	
JP	1	1	1	2	2	2	2	2	1	1	2	
Country	50–54	55–59	60–64	65–69	70–74	75–79	80–84	85–89	90–94	95–99		
AU	1	2	2	2	3	3	3	3	3	1		
IS	2	2	1	1	1	2	3	1	2	3		
UK	2	2	1	2	2	2	2	2	1	1		
US	1	1	1	1	2	2	2	2	2	2		
JP	3	3	3	3	2	2	1	1	1	2		
Male												
Country	0	1–4	5–9	10–14	15–19	20–24	25–29	30–34	35–39	40–44	45–49	
AU	2	2	1	1	3	3	3	3	1	1	2	
IS	1	1	2	3	2	3	2	3	3	2	2	
UK	2	1	1	1	1	3	2	3	3	2	3	
US	3	3	2	3	1	3	3	1	2	2	1	
JP	3	1	2	2	3	2	1	2	3	3	3	
Country	50–54	55–59	60–64	65–69	70–74	75–79	80–84	85–89	90–94	95–99		
AU	2	1	1	3	1	1	2	1	3	3		
IS	3	3	2	2	1	1	1	1	3	1		
UK	2	1	1	2	1	3	3	1	1	1		
US	2	3	3	1	2	3	2	2	3	2		
JP	1	2	3	1	1	1	3	3	2	1		

We then performed two sets of kernel cluster analysis. The first is clustering by country across all age groups, see Table 9. The second cluster analysis is in Table 10, which shows the cluster analysis across countries by age group.

In both cases, the resulting cluster memberships are largely consistent with the groupings one may make from the mono-fractal cluster analysis heat maps, as shown in Figure 1. Most countries

**Table 11.** Kernel k-means cluster representative data sets selected by age groups 10–14, 35–39 and 70–74 to be used for forecasting analysis application

Age group	1	2	3
Female			
10–14	IS	US	AU
35–39	IS	US	UK
70–74	IS	US	UK
Male			
10–14	IS	US	AU
35–39	IS	US	UK
70–74	IS	US	UK



**Figure 7.** Observed mortality rates  $\mu_{x,T-19:T}$  (black line) from HMD and forecast  $\hat{\mu}_{x,T-19:T}$  by Lee-Carter model (female in purple) and long memory extended Lee-Carter model (female in red) for age groups (10–14, 35–39 and 70–74).

show strong long memory features except for IS and ES. Table 11 shows the kernel k-means clustering results for age groups 10–14, 35–39 and 70–74 (representing young, middle and senior age groups) across 16 countries. For IS, the strength of long memory is lower between age 25 and 74 compared to other age groups. The long memory features of the US and JP are generally weaker than other countries.

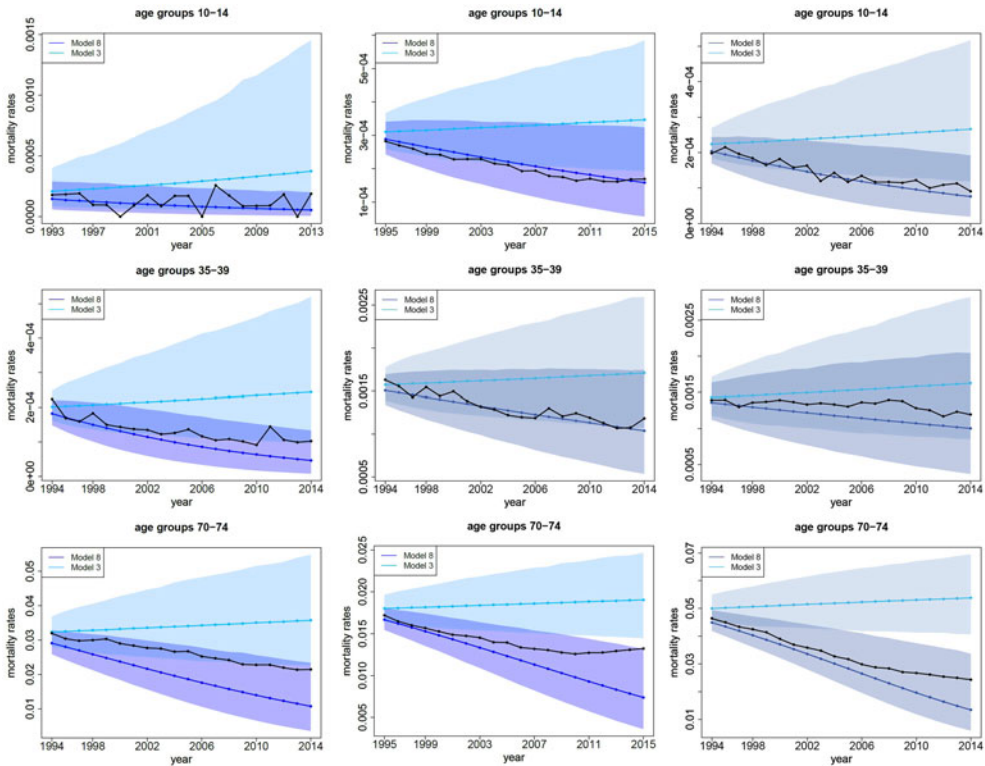


Figure 8. Observed mortality rates  $\mu_{x,T-19:T}$  (black line) from HMD and forecast  $\hat{\mu}_{x,T-19:T}$  by Lee-Carter model (male in skyblue) and long memory extended Lee-Carter model (male in blue) for age groups (10-14, 35-39 and 70-74).

## 7.2 Application of mortality forecasting using long memory features to improve precision and accuracy

In Yan *et al.* (2021, 2020), a set of Bayesian mortality models that extend classic Period-Cohort effect models of Lee-Carter are studied with extensions that include long memory structures. In this section to demonstrate further how the stylised facts of long memory persistence in national-level mortality data by age and gender are important for the actuarial application, we present examples of mortality forecasting. We have taken for this illustration two single age group univariate Bayesian models that are explained in detail in Yan *et al.* (2021). We adopt from this reference, Model 3, which has no long memory and is based on the classical Lee-Carter model with extensions for dispersion effects in the generalised Poisson Observation probability mass function. We compare this to the long memory extended Lee-Carter type Model 8. The specification of these models and their estimation via STAN package MCMC sampling can be found in great detail in the two aforementioned references and are omitted here due to space.

We present results of the application of these two models to study out-of-sample forecast using death count data split into training and testing. The training data for each age group, country and gender is the death count time series records, excluding the last 20 years of data. The forecast testing data are the most recent 20 years of data, which we compare to the model forecasts when incorporating long memory model structure versus leaving such structure out, as is currently the practice in most actuarial applications.

The intention of this short study is to simply demonstrate that in practice incorporating the stylised features of long memory into forecast modelling, as shown in Figures 7 and 8 which plot for male (blue) and female (red) populations the time series of observed mortality rate from the

HMD database and the forecasted mortality rate using the non-long memory Lee–Carter model, compared to the extended long memory Lee–Carter model with 95% credible interval for the 10–14, 35–39 and 70–74 age groups, respectively, improves the forecast performance.

For each age group, a representative example country from each of the three clusters obtained from the kernel k-means clustering corresponding to different long memory features (labelled by 1, 2 and 3) is selected to show the improvements of adopting long memory in the mortality forecasting models.

The three rows represent age groups 10–14, 35–39 and 70–74, respectively, and the three columns represent different long memory features for the cluster groups. For all three age groups and long memory categories, the forecast performance of the long memory Lee–Carter model is significantly better both in out-of-sample forecast accuracy and forecast precision when compared to the traditional Lee–Carter model.

The following two panels of plots demonstrate examples for three sets of age groups, and one country data set per cluster group to be a representative of the forecasting performance when using a classical Lee–Carter model versus the long memory extended Lee–Carter models.

## 8. Conclusion

In this paper, we seek to introduce to actuarial and demographic mortality data analysis an important class of time series data structures not previously studied in such data that we believe can act as an important stylised class of features when developing machine learning methods for mortality. Furthermore, the incorporation of such structures into time series forecasting models can significantly improve the performance of forecasting and consequently life table construction compared to models that fail to accommodate such long memory structures. This paper provides a means to undertake such feature extraction which is model free and allows for significant insights about population mortality dynamics and stylised facts to be explored.

This paper studies the long memory features in mortality data. It first introduces three long memory estimators via Hurst exponent. Through extended synthetic studies, properties of these three estimators are carefully tested. Then in the real data analyses, we find comprehensive evidence for the previously unknown stylised fact that mortality data at the national level in most age groups and genders contains strong long memory structure.

The long memory structure was comprehensively studied using 16 countries' mortality data sets from the HMD. It was found that all 16 countries' mortality data presented strong long memory features consistently for all age groups and genders. This is an important finding as it can now be utilised to improve the estimation of mortality time series by building new classes of mortality models that incorporate long memory structure. It should also aid in improving the accuracy of mortality forecasting by incorporating into model designs this previously unknown stylised fact of mortality data.

We conclude the paper with a multifractal long memory analysis of mortality data, which allows one to discern different timescales of long memory and persistence in mortality time series. We then demonstrate such long memory features can be used for demographic clustering applications and mortality forecasting applications.

**Supplementary materials.** To view supplementary material for this article, please visit <http://doi.org/10.1017/S1748499521000129>.

## References

- Annis, A. & Lloyd, E. (1976). The expected value of the adjusted rescaled hurst range of independent normal summands. *Biometrika*, **63**(1), 111–116.
- Beran, J. (1994). *Statistics for Long-Memory Processes*, vol. **61**. Chapman & Hall/CRC Press, New York.

- Cairns, A.J., Blake, D. & Dowd, K. (2008). Modelling and management of mortality risk: a review. *Scandinavian Actuarial Journal*, **2008**(2–3), 79–113.
- Davidson, J. & De Jong, R.M. (2000). The functional central limit theorem and weak convergence to stochastic integrals ii: fractionally integrated processes. *Econometric Theory*, **16**(5), 643–666.
- Filippone, M., Camastra, F., Masulli, F. & Rovetta, S. (2008). A survey of kernel and spectral methods for clustering. *Pattern Recognition*, **41**(1), 176–190.
- 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.
- Giacometti, R., Bertocchi, M., Rachev, S.T. & Fabozzi, F. J. (2012). A comparison of the Lee–Carter model and AR–ARCH model for forecasting mortality rates. *Insurance: Mathematics and Economics*, **50**(1), 85–93.
- Granger, C.W. & Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, **1**(1), 15–29.
- Graves, T., Gramacy, R.B., Watkins, N. & Franzke, C. (2014). A brief history of long memory. arXiv preprint [arXiv:1406.6018](https://arxiv.org/abs/1406.6018).
- Hosking, J.R. (1981). Fractional differencing. *Biometrika*, **68**(1), 165–176.
- Hurst, H.E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, **116**, 770–808.
- Kantelhardt, J.W., Zschiegner, S.A., Koscielny-Bunde, E., Havlin, S., Bunde, A. & Stanley, H.E. (2002). Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and its Applications*, **316**(1–4), 87–114.
- Koissi, M.-C. & Shapiro, A.F. (2008). The Lee-Carter model under the condition of variables age-specific parameters. In *43rd Actuarial Research Conference*.
- Mandelbrot, B.B. (1975). Limit theorems on the self-normalized range for weakly and strongly dependent processes. *Probability Theory and Related Fields*, **31**(4), 271–285.
- Mandelbrot, B.B. & Van Ness, J.W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review*, **10**(4), 422–437.
- Mandelbrot, B.B. & Wallis, J.R. (1969). Computer experiments with fractional Gaussian noises: Part 1, averages and variances. *Water Resources Research*, **5**(1), 228–241.
- Micchelli, C.A., Xu, Y. & Zhang, H. (2006). Universal kernels. *Journal of Machine Learning Research*, **7**(12), 2651–2667.
- Millen, S. & Beard, R. (2003). Estimation of the Hurst exponent for the Burdekin River using the Hurst-Mandelbrot rescaled range statistic. In *First Queensland Statistics Conference*.
- Nualart, D. (2006). Fractional Brownian motion: stochastic calculus and applications. In *International Congress of Mathematicians*, vol. 3 (pp. 1541–1562). European Mathematical Society
- Peters, E.E. (1994). *Fractal Market Analysis: Applying Chaos Theory to Investment and Economics*, vol. 24. John Wiley & Sons.
- Peters, G., Shevchenko, P.V., Cohen, R. and Maurice, D. (2017). Statistical machine learning analysis of cyber risk data: event case studies. Chapter 3 in Edited Book: *Fintech: Growth and Deregulation*. In Diane Maurice, Jack Freund and David Fairman (Eds.), 2018, Risk Books, London.
- Qian, B. & Rasheed, K. (2004). Hurst exponent and financial market predictability. In *Proceedings of The 2nd IASTED International Conference on Financial Engineering and Applications* (pp. 203–209).
- Shkolnikov, V. (2017). Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at [www.mortality.org](http://www.mortality.org) or [www.humanmortality.de](http://www.humanmortality.de) (downloaded date 01/01/2019).
- Taqqu, M.S., Teverovsky, V. & Willinger, W. (1995). Estimators for long-range dependence: an empirical study. *Fractals*, **3**(04), 785–798.
- Weron, R. (2002). Estimating long-range dependence: finite sample properties and confidence intervals. *Physica A: Statistical Mechanics and its Applications*, **312**(1), 285–299.
- Yan, H., Peters, G. & Chan, J. (2021). Mortality models incorporating long memory for life table estimation: A comprehensive analysis. *Annals of Actuarial Science*, 1–38. doi: [10.1017/S1748499521000014](https://doi.org/10.1017/S1748499521000014).
- Yan, H., Peters, G.W. & Chan, J.S. (2020). Multivariate long-memory cohort mortality models. *ASTIN Bulletin: The Journal of the IAA*, **50**(1), 223–263.
- Yaya, O.S. & Gil-Alana, L.A. (2020). Modelling Long-Range Dependence and Non-linearity in the Infant Mortality Rates of African Countries. *International Advances in Economic Research*, **26**(3), 303–315.
- Yaya, O.S., Gil-Alana, L.A. & Amoateng, A.Y. (2019). Under-5 mortality rates in g7 countries: analysis of fractional persistence, structural breaks and nonlinear time trends. *European Journal of Population*, **35**(4), 675–694.