Astin Bulletin (2025), **55**, pp. 97–120
doi:10.1017/asb.2024.38

International Actuarial Association
Association Actuarielle Internationale

RESEARCH ARTICLE

Forecasting mortality rates with functional signatures

Zhong Jing Yap[1], Dharini Pathmanathan[1,2,3] and Sophie Dabo-Niang[4,5]

[1]Institute of Mathematical Sciences, Faculty of Science, Universiti Malaya, 50603, Kuala Lumpur, Malaysia
[2]Universiti Malaya Centre for Data Analytics, Universiti Malaya, 50603, Kuala Lumpur, Malaysia
[3]Center of Research for Statistical Modelling and Methodology, Faculty of Science, Universiti Malaya, 50603, Kuala Lumpur, Malaysia
[4]UMR8524–Laboratoire Paul Painlevé, Inria-MODAL, University of Lille, CNRS, Lille, 59000, France
[5]CNRS–Université de Montréal, CRM–CNRS, Montréal, Canada
Corresponding author: Dharini Pathmanathan; Email: dharini@um.edu.my

Received: 22 July 2024; **Revised:** 12 December 2024; **Accepted:** 12 December 2024; **First published online:** 9 January 2025

Keywords: Hyndman-Ullah model; Lee-Carter model; principal component analysis; functional data analysis

Abstract

This study introduces an innovative methodology for mortality forecasting, which integrates signature-based methods within the functional data framework of the Hyndman–Ullah (HU) model. This new approach, termed the Hyndman–Ullah with truncated signatures (HUts) model, aims to enhance the accuracy and robustness of mortality predictions. By utilizing signature regression, the HUts model is able to capture complex, nonlinear dependencies in mortality data which enhances forecasting accuracy across various demographic conditions. The model is applied to mortality data from 12 countries, comparing its forecasting performance against variants of the HU models across multiple forecast horizons. Our findings indicate that overall the HUts model not only provides more precise point forecasts but also shows robustness against data irregularities, such as those observed in countries with historical outliers. The integration of signature-based methods enables the HUts model to capture complex patterns in mortality data, making it a powerful tool for actuaries and demographers. Prediction intervals are also constructed with bootstrapping methods.

1. Introduction

Mortality modeling and forecasting are quintessential to the disciplines of actuarial science, demography, and public health, serving as critical tools for socioeconomic planning and risk management. Their primary objective is to provide reliable estimates of future mortality rates and life expectancies, which are essential for shaping policy and financial strategies. For government agencies, anticipating mortality trends is fundamental for effective resource allocation in healthcare, social security, and public welfare sectors. Precise mortality forecasts enable informed decisions on retirement ages, healthcare funding, and pensions, ensuring that social security systems remain robust amid demographic changes.

In the insurance and pension industries, mortality modeling takes part in the pricing of life insurance, annuities, and managing pension funds. Actuaries rely on these models to calculate premiums, reserves, and pension contributions that align with predicted mortality rates, thus safeguarding institutions against longevity risks (Deprez *et al*., 2017). Accurate mortality projections are also critical in addressing global challenges posed by increasing life expectancy and declining fertility rates which shift the age structure of populations. These demographic trends have profound implications for the economic sustainability of pension systems and public health policies (Bjerre, 2022). Consequently, the practice of mortality modeling and forecasting not only supports financial and actuarial operations but also facilitates broader strategic planning and policy-making to accommodate the evolving demographic landscapes, ultimately ensuring the economic and social stability of societies.

© The Author(s), 2025. Published by Cambridge University Press on behalf of The International Actuarial Association.

https://doi.org/10.1017/asb.2024.38 Published online by Cambridge University Press

1.1. Mortality modeling and forecasting

Mortality models are broadly categorized into three approaches: expectation, explanation, and extrapolation (Booth and Tickle, 2008). Each serves distinct purposes in understanding and predicting mortality rates. The expectation approach relies on expert opinions to forecast mortality, emphasizing informed conjectures based on past experiences. Meanwhile, explanation models link mortality rates to various risk factors such as lifestyle or socioeconomic status, using regression techniques to uncover the underlying causes of mortality patterns. The primary focus, however, is on extrapolation, the most actively researched and commonly used approach, particularly by national statistical bureaus. This method projects future mortality rates based on the continuation of observed historical trends and age patterns. By leveraging time series analysis and other statistical techniques that emphasize trend continuity, extrapolative models provide robust forecasts grounded in the regularity of past data, making them indispensable for demographic analysis and policy planning.

The seminal work of Lee and Carter (1992) was undoubtedly one of the most influential extrapolative models to be introduced to forecast mortality rates. While the Lee–Carter (LC) model was initially applauded for its simple structure and straightforward implementation, it did have its fair share of limitations which were subsequently addressed in various extensions of the LC model (Lee and Miller, 2001; Booth *et al*., 2002). Extensions were also not limited to adjustments to the original LC model as seen in the deployment of generalized linear models (Renshaw and Haberman, 2003) and Bayesian statistics (Czado *et al*., 2005). Readers are directed to Basellini *et al*. (2023) for a more recent review of LC methods since the introduction of the LC model.

The LC extension of interest in this paper involves the functional data paradigm (Ramsay and Silverman, 2005). The Hyndman–Ullah (HU) model (Hyndman and Ullah, 2007) is a generalization of the LC model that uses concepts from functional data analysis to better model mortality rates by exploring variations within the data from a functional perspective. The use of nonparametric smoothing techniques on the age-specific mortality rates allowed for the removal of the LC model's homoskedasticity assumption and realize an underlying smooth curve that can be decomposed using functional principal component analysis (FPCA) with their orthonormal basis. Component wise it is similar to the LC model except it allows for multiple principal components to be included, accounting more variation into the model which Hyndman and Ullah (2007) noted is the reason for its superior performance for longer-term forecast. Comparisons also show that the HU model outperformed the LC, Lee–Miller, and Booth–Maindonald–Smith models. Recent extensions of the HU model have been for the modeling of subpopulations (Shang and Hyndman, 2017; Shang *et al*., 2022; Jiménez-Varón *et al*., 2024). In light of the substantial contributions made by the HU model, there is still a growing interest to explore new modifications to improve its efficacy and adaptability.

With the rise of network approaches in modeling, there is growing interest in extending age-period models using deep learning techniques. Various aspects of the LC model have been coupled with deep learning applications, both in parameter estimation (Hainaut, 2018; Richman and Wüthrich, 2021; Schnürch and Korn, 2022; Scognamiglio, 2022; Miyata and Matsuyama, 2022) and in forecasting procedures (Nigri *et al*., 2019; Perla *et al*., 2021; Bjerre, 2022; Marino *et al*., 2023). Deep learning models fundamentally process streams of input data to approximate relationships. While they perform well, they can lack replicability and often require increased computational resources. Signatures offer a potential solution to these limitations as they too possess universality to approximate relationships arbitrarily well, reproducible, and efficient. Therefore, this paper aims to integrate signatures that better capture the nonlinear patterns often present in single-population mortality data into the HU model's functional data framework.

1.2. Signatures from rough path theory

The signature is a sophisticated mathematical tool that is capable of summarizing complex, high-dimensional data paths through their iterated integrals in an informative and compact package as a power

https://doi.org/10.1017/asb.2024.38 Published online by Cambridge University Press

series that contains the tensor coefficients. Originally introduced by Chen (1957) and further developed in the context of rough path theory by Lyons *et al*. (2007), they have found extensive applications across multiple fields, particularly where the analysis of dynamic systems is crucial. Their effectiveness has been demonstrated in various applications such as sepsis detection (Morrill *et al*., 2020), identifying Chinese handwriting (Xie *et al*., 2018), and in diagnosing psychological disorders (Wang *et al*., 2020). The increasing interest in signatures stems primarily from their ability to effectively condense information from complex sequences without losing critical details, making them exceptionally useful for detection work. This efficiency in capturing the essence of data paths while minimizing complexity is what drives their growing popularity, especially in areas dealing with large volumes of sequential data such as in the field of quantitative finance (Lyons *et al*., 2020; Kalsi *et al*., 2020; Cuchiero and Möller, 2024; Jaber and Gérard, 2024).

Signatures possess a unique mathematical property known as universal nonlinearity, which enables the application of regression techniques directly on these signature representations, known as signature regression. This method leverages the rich algebra provided by the signatures to model and predict dependent variables from a sequence of inputs. The practical implication of this is that it can construct predictive models that are capable of handling data with complex temporal dynamics and interactions. In other words, the universality of signatures means that any continuous function on the space of signatures can approximate continuous functions on the space of paths to an arbitrary degree of accuracy. This broad capability makes signature-based models exceptionally versatile and powerful, providing a robust framework for tackling problems where conventional regression models might struggle with the complexities of the data.

Signature regression is further highlighted through various applications across different fields. For instance, Levin *et al*. (2016) utilized the universal properties of signatures to characterize functional relationships within datasets, leading to the development of the expected signature model. This model leverages linear regression on signature features to summarize the conditional distribution of responses effectively. It has demonstrated superior forecasting abilities compared to autoregressive and Gaussian Process models, especially in scenarios involving large datasets, thereby highlighting its efficiency in data stream summarization and significant dimension reduction.

Moreover, the adaptability and simplicity of modeling with signature regression have been elaborated by Cohen *et al*. (2023), who noted their effectiveness in simplifying time series modeling. This approach involves merely computing signatures from observed paths and applying regression, significantly reducing the complexity associated with time series analysis. The flexibility of signature methods also extends to handling missing data and irregular sampling, making them applicable for diverse forecasting and analysis tasks. Notably, when applied to nowcast the gross domestic product growth of the United States, signature regression methods have outperformed dynamic factor models, showcasing its potential in economic forecasting.

Building on these insights, Fermanian (2022) was the first to apply the concept of signatures to functional data analysis by introducing the signature linear model within a functional regression framework. Fermanian (2022) noted that signatures are naturally adapted to vector-valued functions, unlike other functional methods, and require little assumption on the regularity of functions to encode nonlinear geometric information. This model operates by mapping functions to their signatures and employing ridge regression for prediction, showing enhanced capabilities in handling multidimensional data. By surpassing established methods like functional linear models and functional principal component regression, the signature linear model eliminates the need for certain assumptions typically required within functional settings, streamlining the modeling process. This lays the foundation that suggests that signature regression is a good candidate to extend the functional data framework of the HU model.

In the field of actuarial science, the use of signatures has started to make inroads, particularly in areas such as validating economic scenarios (Andrès *et al*., 2024) and option pricing (Cuchiero *et al*., 2023, 2024). However, the application of signatures in the life and mortality side of actuarial science remains less explored, indicating a gap for research and potential development. The introduction of signature

https://doi.org/10.1017/asb.2024.38 Published online by Cambridge University Press

regression into mortality forecasting presents a novel methodology that could improve the accuracy of predictions by leveraging the universality of signatures to effectively model complex, nonlinear relationships inherent in mortality data. It is therefore the aim of this article to introduce a functional mortality model by which is decomposed using truncated signatures in place of FPCA that also utilizes the extrapolation method to produce forecasts. The model will from here on be referred to as the HU with truncated signatures (HUts) model and will be used to obtain point and interval forecasts of log mortality rates of several countries that are readily available from the Human Mortality Database (2024).

The remainder of this paper is structured as follows. Section 2 provides a description of the methodology, detailing the HU model and the integration of signature regression. Section 3 presents the empirical analysis, including data preparation and results of applying the proposed model to mortality data across various countries. In Section 4, we discuss the implications of these results, and Section 5 concludes.

2. Methodology

2.1. The Hyndman-Ullah (HU) model

The HU model is a generalized extension of the LC model, employing the functional data paradigm (Ramsay and Silverman, 2005) and nonparametric smoothing to improve the forecasting of mortality rates. This model is distinguished by its use of multiple principal components, capturing a broader variation within the data, and applying more advanced time series forecasting methods than the random walk with drift model typically used in the LC model (Hyndman and Ullah, 2007).

At its core, the HU model assumes that log mortality rates $y_t(x)$ observed at discrete ages x and years t have an underlying smooth function $f_t(x)$, observed with error:

$$y_t(x) = f_t(x) + \sigma_t(x)\epsilon_t,$$

where ϵ_t is the standard normal error term and $\sigma_t(x)$ denotes the observational standard deviation which is allowed to vary with age. This lets the HU model handle data heterogeneity and ensures demographic consistency by applying monocity constraints when using penalized regression splines to smooth the mortality curves (Hyndman and Ullah, 2007).

FPCA is used to decompose the smoothed mortality rates into principal components. By applying FPCA to the centered smooth mortality would yield:

$$f_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + e_t(x).$$

Here, $\mu(x)$ represents the mean function, $\phi_k(x)$ are orthonormal basis functions, $\beta_{t,k}$ are time-dependent coefficients, and $e_t(x)$ is the error function. Forecasts of mortality rates are done through the time-dependent coefficients, $\beta_{t,k}$ which are fitted into univariate ARIMA models and extrapolated h-steps-ahead.

A critical consideration in applying FPCA is its sensitivity to outliers. Extreme values can disproportionately influence the principal components derived from the data. This sensitivity may lead to biased or misleading results, especially when modeling phenomena such as mortality rates where spikes might occur due to extraordinary events like epidemics or wars. To address these issues, various adaptations of the HU model have been put forth. These include the robust HU (HUrob) model with robust FPCA (Hyndman and Ullah, 2007), which seeks to minimize the impact of outliers by using the L_1-median as the location measure, paired with the reflection-based algorithm for principal components analysis (Hubert *et al*., 2002) for a robust set of principal components; the weighted HU (wHU) model with weighted FPCA (Hyndman and Shang, 2009), which adjusts the influence of data points by placing more weight on recent data with the measure of location being a weighted average using geometrically decreasing weights. These modifications aim to enhance the stability and reliability of the model under conditions where data anomalies are present. Having the same objective in mind to make the model less

https://doi.org/10.1017/asb.2024.38 Published online by Cambridge University Press

sensitive to outliers, we extend the literature by exploring the use of signature regression on the signature of age-specific mortality rates as an alternative to enhancing the robustness of mortality forecasts.

2.2. The signature of a path

2.2.1. The concept of signatures

In statistical analysis, particularly when dealing with time series or sequential data, understanding the concept of a "path" is crucial. A path $X : [0, 1] \to \mathbb{R}^d$ is defined as a continuous mapping from an interval to a topological space, which is assumed to be piecewise differentiable and of bounded variation. Unlike time series or stochastic processes, which are explicitly governed by time, a path emphasizes its geometrical movements through space, making the term "path" more suitable. From a computational perspective, think of a path as input data with d dimensions, analogous to multidimensional time series data.

Signatures are mathematical tools used to encode the information contained in a path. Using Chen's identity (Chen, 1957), the signature of a piecewise linear path is computed by taking the tensor product of exponentials of the linear segments of the path. Thus, the signature of a path, or simply signatures, is a collection of all iterated integrals between coordinates of a path. This collection is often described as an abstract summary of the input signal's variability.

Definition 2.1 *(Signature of a path (Fermanian, 2022)). Let $X : [0, 1] \to \mathbb{R}^d$ be a path of bounded variation, and let $I = (i_1, \dots, i_k) \in \{1, \dots, d\}^k$, where $k \in \mathbb{N}^*$, be a multi-index of length k. The signature coefficient of X corresponding to the index I is defined by:*

$$\begin{aligned} S^I(X) &= \int_{0 \le u_1 < \dots < u_k \le 1} \cdots \int dX_{u_1}^{i_1} \dots dX_{u_k}^{i_k} \\ &= \int_0^1 \int_{u_1}^1 \int_{u_2}^1 \dots \int_{u_{k-1}}^1 dX_{u_k}^{i_k} \dots dX_{u_2}^{i_2} dX_{u_1}^{i_1}, \end{aligned}$$

where $S^I(X)$ is called the signature coefficient of order k. The signature of X is then the sequence containing all the signature coefficients:

$$S(X) = (1, S^{(1)}(X), \dots, S^{(d)}(X), S^{(1,1)}(X), \dots, S^{(i_1,\dots,i_k)}(X), \dots).$$

However, from a practical standpoint, it is not feasible to use all the signature coefficients in the infinite sequence of the signature of X due to computational limits and also due to the factorial decay of the coefficients as the order increases (Bonnier *et al*., 2019). Therefore, it is often sufficient to consider the signature of a path up to a truncated order of m, denoted as:

$$S^m(X) = (1, S^{(1)}(X), \dots, S^{(d)}(X), S^{(1,1)}(X), \dots, S^{(i_1,\dots,i_m)}(X)),$$

which contains $s_d(m) = \frac{d^{m+1}-1}{d-1}$ number of signature coefficients. Note that the number of signature coefficients grows exponentially with the truncation order.

A fundamental attribute of the signature is its ability to encode the geometric properties of a path. The first-order terms, $S^{(i)}(X)$, are straightforward, representing the total change along each dimension of the path or the increments observed in each variable. The second-order terms, $S^{(i_1,i_2)}(X)$, provide insights into the interaction between pairs of dimensions, similar to capturing the area enclosed by the path as it travels through these dimensions, and higher-order interactions are particularly valuable for understanding complex dependencies and dynamics within the data.

2.3. Embedding with signatures

Embedding paths before computing its signature is a standard practice used to maximize the potential of signatures by addressing inherent invariances such as translation and time reparametrization that can

https://doi.org/10.1017/asb.2024.38 Published online by Cambridge University Press

result in significant information loss. These embedding transformation ensure that all relevant information are preserved and encoded into the signature and in turn enrich the signature's capability to reflect the complex dynamics within the path.

There are various embedding techniques that can be used to augment paths, each suited to specific applications based on the aspects of the data they capture. Different embedding choices highlight different features of the data, making them appropriate for particular tasks (Fermanian, 2021). In this paper, we focus mainly on three types of embeddings: basepoint augmentation, time augmentation, and the lead-lag transformation.

The basepoint augmentation (Kidger and Lyons, 2020) involves inserting a fixed initial point to the path data. By anchoring the sequence, basepoint augmentation allows the signature to overcome the translation invariance of the signature.

Time augmentation counteracts the signature's invariance to time reparametrization, by introducing a monotone coordinate that serves as a virtual timestamp to each point in the path. This embeds precise temporal information into the path, making the signature sensitive to the sequence's timing. Besides, this also guarantees the uniqueness of the signatures computed (Hambly and Lyons, 2010).

To further aid the sequential learning capabilities of the signature, the lead-lag transformation extends the path's dimensionality by interleaving the original path with its lagged version, which enriches the data representation, allowing the signature to capture not only the path's immediate direction but also its past states. Fermanian (2021) noted the lead-lag transformation improves a model's predictive capabilities and is beneficial for tasks involving sequential data.

In the proceeding sections, we will use these embedding transformation to prepare our data paths before computing its signature so as to enhance the signatures' ability to capture and reflect the complex dynamics of the sequences.

2.4 Signature regression

One of the many practical applications of signatures are its usage as feature maps in a regression framework. This rich representation allows for the modeling of complex relationships within functional data. The ability of signatures to compactly encode information about paths makes them particularly powerful for handling high-dimensional and complex datasets, enabling the application of regression techniques in scenarios where conventional methods might be inadequate due to the complexity and dimensionality of the data. These are motivated and guaranteed by many results as pointed out by the shuffle identity, the signature approximation theorem (Levin *et al*. 2016), and the universal nonlinearity of paths (Bonnier *et al*., 2019).

Definition 2.2 *(Universal nonlinearity (Bonnier et al., 2019)). Let f be a real-valued continuous function on continuous piecewise smooth paths in $\mathbb{R}^d$ and let κ be a compact set of such paths. Assume that $X_0 = 0, \forall X \in \kappa$, let $\epsilon > 0$, there exists a linear functional L such that,*

$$|f(X) - L(S(X))| < \epsilon.$$

Kiraly and Oberhauser (2019) then gathered that a linear combination of signature features are capable of approximating continuous functions arbitrarily well, meaning signatures are able to linearize functions of a path, further emphasizing the universality (Levin *et al*., 2016) of signatures. To further clarify, consider a classical regression setting with a path X that maps some space into $\mathbb{R}^d$ and Y a response, we have

$$Y = f(X) + \epsilon,$$

where f is some functional that characterizes the relationship between the response and the path. Assuming f is nonlinear, it is then generally challenging to determine the function f accurately due to the increased complexity in parameter estimation, model interpretation, and computational resources required for nonlinear regression or possibly neural networks, compared to linear regression. However,

https://doi.org/10.1017/asb.2024.38 Published online by Cambridge University Press

if the signature of path X is considered as the predictors instead, we can redefine the relationship with a linear functional g such that

$$Y = g(S(X)) + \epsilon.$$

Although in theory linear regression is sufficient for estimating g, in practice, regularization techniques are used to handle the multicollinearity of signature features (Levin *et al*., 2016) as linear regression models can become unstable when variables are not independent. This instability often results in large variances in the estimated coefficients, making the model unreliable.

2.5. The Hyndman–Ullah model with truncated signatures

We are now prepared to integrate signatures into the HU model, which we will refer to as the HU with truncated signatures (HUts) model. The integration of signatures with functional data models have been documented by Fermanian (2022) and Frévent (2023). However, these studies primarily applied signature regression in a functional regression framework with scalar responses and functional covariates, where signatures of the functional covariates were used directly. In contrast, the HU model operates more like a functional regression model with functional responses, making the direct application of signature regression more complex. We plan to consider using the signature of cross-sectional data from these functional covariates to apply signature regression at each point throughout the functional domain and then reconstruct a functional response. The HUts model is presented as follows:

$$y_t(x) = f_t(x) + \sigma_t(x)\epsilon_t,$$

$$f_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k} Z_k(x) + e_t(x),$$

where $Z_k(x)$ is the kth principal component of the signature of the transformed path of age-specific rates, and its respective scores $\beta_{t,k}$.

The smoothing approach of the mortality rates are remained as the original HU model, which is with the constrained weighted penalized regression splines to obtain $\hat{f}_t(x)$. This ensures the model retains the capacity to account for handling heteroskedasticity in mortality data as variability is different across different ages or time periods. Although more advanced smoothing techniques such as bivariate smoothing (Dokumentov *et al*., 2018) have been put forth, they also introduce additional complexity and computational demands.

Additionally, the estimation of the mean function is also kept the same as in the HU model, by averaging all the smoothed curves:

$$\hat{\mu}(x) = \frac{1}{n} \sum_{t=1}^{n} \hat{f}_t(x).$$

This method provides a straightforward and effective way to capture the central tendency of the functional responses. Now, we will decompose the centered curves, $f_t^*(x) = \hat{f}_t(x) - \hat{\mu}(x)$ with signature regression over a fine grid of q equally spaced values $x_1^*, \ldots, x_q^* \in [x_1, x_p]$.

Consider the time series of the centered smoothed mortality rates for age x_i, $\{f_t^*(x_i)\}_{t=1}^n$. Combining the basepoint augmentation, time augmentation, and the lead-lag transformation, the augmented path can be transformed as follows:

$$\tilde{X}_i = \Big((t_0, 0, 0), (t_1, f_1^*(x_i), f_1^*(x_i)), (t_2, f_2^*(x_i), f_1^*(x_i)), (t_3, f_2^*(x_i), f_2^*(x_i)),$$
$$(t_4, f_3^*(x_i), f_2^*(x_i)), \ldots, (t_n, f_n^*(x_i), f_n^*(x_i))\Big),$$

where t is an ordered sequence within the interval $[0,1]$ and $t_0 = 0 < t_1 < t_2 < \ldots < t_n = 1$, marking the times of observations. This structured augmentation ensures that each data point x_i is not only represented by its current value but also linked to its immediate past, enhancing the path's dimensionality and

https://doi.org/10.1017/asb.2024.38 Published online by Cambridge University Press

capturing both the direction and history within the data, thereby enriching the data representation for signature computation. The transformed time series $\tilde{X}_i$ is then used to compute the truncated signature $S^m(\tilde{X}_i)$.

To estimate $\beta_{t,k}$, we then set up our signature regression framework:

$$f_t^*(x_i) = g_t(S^m(\tilde{X}_i)) + \hat{e}_t(x_i)$$

for some linear function g_t, which we will use principal component regression to solve. Applying principal component analysis (PCA) helps by reducing the dimensionality of the signature coefficient feature space to a smaller set of k uncorrelated principal components denoted as $Z_k(x_i)$. This reduction not only simplifies the model but also enhances its stability and efficiency by focusing on the most informative aspects of the dependency within the signatures. Additionally, using PCA aids in orthogonalizing the terms, which simplifies the calculation of forecast variance later and resembles the approaches used in both the original HU and LC models.

For computational simplicity, we define the matrices:

$$S = \begin{bmatrix} S^m(\tilde{X}_1) \\ \vdots \\ S^m(\tilde{X}_q) \end{bmatrix}_{q \times s_d(m)}, \qquad Y_t = \begin{bmatrix} f_t^*(x_1^*) \\ \vdots \\ f_t^*(x_q^*) \end{bmatrix}_{q \times 1}. \tag{2.1}$$

We then normalize S to obtain S^* before applying PCA to obtain the orthogonal principal component scores by singular value decomposition, $S^* = U\Sigma V^T$ where the first K principal component scores are the first K columns of $Z = U\Sigma$. Then, to obtain $\hat{\beta}_t$ is equivalent to solving

$$\hat{\beta}_t = (Z_K^\top Z_K)^{-1} Z_K^\top Y_t,$$

where Z_K is a $q \times K$ matrix where the ith row contains the principal component of S^* of order m of age x_i.

By combining everything together, we have

$$\hat{f}_t(x) = \hat{\mu}(x) + \sum_{k=1}^{K} \hat{\beta}_{t,k} \hat{Z}_k(x) + \hat{e}_t(x),$$

where $\hat{Z}_k(x)$ is obtained via interpolation.

Remark 2.3. As time augmentation leads to identical values for the signature coefficients $S^{(1)}, S^{(1,1)}, S^{(1,1,1)}, \ldots$, along each row of S in Equation (2.1), certain columns of S become constant. During normalization, these zero-variance columns are excluded from being centered and scaled, ensuring a well-defined normalized matrix S^*. Consequently, when performing PCA on S^*, one of the principal components is expected to be constant.

2.6. Forecasting with the HUts model

2.6.1. Point forecast

Similar to that of the HU model, the time-dependent coefficients are extrapolated by fitting them into univariate time series $\{\hat{\beta}_{t,k}\}_{t=1}^n$. We preserve the choice of using univariate ARIMA models as in the HU model to forecast each coefficient h-steps-ahead, $\hat{\beta}_{n+h|n,k}$, for each $k = 1, \ldots, K$. Subsequently, the forecasted log mortality rates can be calculated with:

$$\hat{y}_{n+h|n}(x) = \hat{\mu}(x) + \sum_{k=1}^{K} \hat{\beta}_{n+h|n,k} \hat{Z}_k(x).$$

Since each term is constructed to be orthogonal to one another, the forecast variance can also be approximated, assuming all sources of error are normally distributed, by summing the component variances

https://doi.org/10.1017/asb.2024.38 Published online by Cambridge University Press

similar to that of the forecast variance of the HU model (Hyndman and Ullah, 2007):

$$\mathrm{Var}(y_{n+h|n}(x)) \approx \hat{\sigma}_{\mu}^{2}(x) + \sum_{k=1}^{K} \hat{Z}_{k}^{2}(x) u_{n+h|n,k} + v(x) + \sigma_{n+h|n}^{2}(x), \tag{2.2}$$

where $\hat{\sigma}_{\mu}^{2}(x)$ is the variance obtained from estimating the mean function, $u_{n+h|n,k}$ is the variance from the time series used to forecast $\hat{\beta}_{n+h|n,k}$, $v(x)$ are model fit errors estimated by averaging $\hat{e}_{t}(x)$, and finally $\sigma_{n+h|n}^{2}(x)$, the observational variance.

2.6.2. Interval forecast

Obtaining prediction intervals is essential because they provide a measurable indication of the uncertainty in forecasts. Prediction intervals quantify the range within which future observations are likely to occur, at a specified confidence level. This allows evaluation of the reliability of predictions when model uncertainty is significant, as different models may offer varying forecasts performance. Chatfield (1993) emphasized the importance of prediction intervals for evaluating future uncertainty, preparing for various scenarios, comparing different forecasting methods, and exploring outcomes under different assumptions. By providing a range of possible outcomes, interval forecasts allow policymakers and researchers to prepare for a variety of future scenarios, thus enhancing the reliability of their strategies against unexpected changes. For example, understanding the variability in future mortality rates can profoundly impact the structuring of insurance (Hyndman and Shang, 2009). Interval forecasts helps insurers assess the risk associated with different demographic scenarios, which is then translated into premiums to reflect the uncertainty of their projections. This further frames the concept of interval forecasts as a fundamental component in the field of demography.

Constructing prediction intervals is straightforward under the assumption that all sources of error are uncorrelated and normally distributed. Using the standard normal distribution, these intervals can be easily derived by applying quantile values to scale the forecast standard deviation obtained in Equation (2.2) which is straightforward to implement. However, its accuracy heavily depends on the validity of the normality assumption. If the normality assumption is violated, the prediction intervals may not be accurate, suggesting that a nonparametric bootstrap procedure would be more practical.

2.6.3. Bootstrap prediction interval

Since the HUts model, like the wHU model, do not satisfy the normality assumption (see Appendix A in the Supplementary Material), we employ the same method used to construct bootstrap prediction intervals for the wHU model in Hyndman and Shang (2009) to construct prediction intervals for the HUts model.

The procedure begins by utilizing univariate time series models to generate multi-step-ahead forecasts for the scores of principal components, $\{\beta_{t,k}\}_{t=1}^{n}$. Subsequently, for $t = h+1, \ldots, n$, we calculate the forecast errors for h-step-ahead forecasts as:

$$\hat{\xi}_{t,h,k} = \hat{\beta}_{t,k} - \hat{\beta}_{t|t-h,k}.$$

To create a bootstrap sample of $\beta_{n+h,k}$, we randomly resample these forecast errors with replacement, designated as $\hat{\xi}_{*,h,k}^{(l)}$, and add them to the h-step-ahead forecast, $\hat{\beta}_{n+h|n,k}$, yielding:

$$\hat{\beta}_{n+h|n,k}^{(l)} = \hat{\beta}_{n+h|n,k} + \hat{\xi}_{*,h,k}^{(l)},$$

for $l = 1, \ldots, L$.

The model's residual, $\hat{e}_{n+h|n}^{(l)}(x)$ is bootstrapped from the residuals $\{\hat{e}_1(x), \ldots, \hat{e}_n(x)\}$ by sampling with replacement. Similarly, smoothing errors $\hat{\epsilon}_{n+h,i}^{(l)}$ can be bootstrapped by resampling with replacement from $\{\hat{\epsilon}_{1,i}, \ldots, \hat{\epsilon}_{n,i}\}$.

https://doi.org/10.1017/asb.2024.38 Published online by Cambridge University Press

Table 1. *Mortality data by country.*

Country	Commencing year	Age range
Australia	1921	0-100+
Belgium	1920	0-100+
Bulgaria	1947	0-100+
Denmark	1899	0-99+
Finland	1899	0-96+
France	1899	0-100+
Ireland	1950	0-100+
Italy	1899	0-100+
Japan	1947	0-100+
Netherlands	1899	0-99+
Norway	1899	0-100+
United States	1933	0-100+

When constructing the L variants for $\hat{y}^{(l)}_{n+h|n}(x)$, all possible components of variability are combined, assuming they are uncorrelated to each other by using:

$$\hat{y}^{(l)}_{n+h|n}(x_i) = \hat{\mu}(x_i) + \sum_{k=1}^{K} \hat{\beta}^{(l)}_{n+h|n,k}\hat{Z}_k(x_i) + \hat{e}^{(l)}_{n+h|n}(x_i) + \hat{\sigma}_{n+h}(x_i)\hat{\epsilon}^{(l)}_{n+h,i}.$$

Subsequently, the $(1-\alpha)100\%$ prediction intervals, $[l_h(x), u_h(x)]$ can then be obtained from the $(1-\alpha/2)$ and $(\alpha/2)$ quantiles of these bootstrap variants. We also consider the adjustment made to the bootstrap prediction intervals in Hyndman and Shang (2009):

$$[0.5\{l_h(x)+u_h(x)\} - \{u_h(x)-l_h(x)\}p(x), 0.5\{l_h(x)+u_h(x)\} + \{u_h(x)-l_h(x)\}p(x)]$$

to account for the uncertainty in our forecast by calibrating the intervals using one-step-ahead forecast errors. Here, $p(x) = d(x)/[u_1(x) - l_1(x)]$, where $d(x)$ is the empirical differences between the $(1-\alpha/2)$ and $(\alpha/2)$ quantiles of the in-sample one-step-ahead forecast errors $\{\hat{f}_{t+1}(x) - \hat{f}_{t+1|t}(x); t = r, \cdots, n-1\}$ and r is the smallest number of observations used to fit the model.

3. Empirical analysis

3.1. Data description

This study utilized datasets from 12 different countries, all sourced from the Human Mortality Database (2024). Of the 12 countries considered, data spanning all available years up to 2015 were extracted, with age groups ranging from zero to a carefully determined maximum age to avoid zero or missing values in elderly data. For ages beyond this established maximum, they were combined into an upper age group. In addressing intermediate missing values within the data for Denmark, Finland, and Norway, a uniform approach was applied. A constant, represented by the smallest death rate value in the dataset, was added to each death rate prior to taking its logarithm. This adjustment facilitated a consistent treatment of missing values across the dataset. The selected countries are tabulated in Table 1, along with the commencing years and age groups.

3.2. Point forecast

The log mortality rates of 12 countries were forecasted with the HUts model using a truncation order $m = 2$ (see Appendix B in the Supplementary Material). It is therefore intuitive to compare the forecast accuracy of the proposed method with the HU model, HUrob model, and the wHU model, which

https://doi.org/10.1017/asb.2024.38 Published online by Cambridge University Press

are commonly used on mortality rates of a single population. The selection of the number of principal components K can impact the forecasting performance of the HUts model and the HU model variants. They can be chosen based off a threshold function that measures the proportion of variance explained by each principal component, or by minimizing the integrated squared forecast error (Hyndman and Ullah, 2007). However, Hyndman and Booth (2008) noted that the HU models are relatively insensitive to the choice of K as long as it is sufficiently large, and that setting $K = 6$ is more than adequate to capture the essential features of the data. For consistency across all models, we adopt $K = 6$ for both the HU model variants and the proposed HUts model. The models were fitted using an expanding window approach, incorporating data from the respective commencing years of each country as outlined in Table 1. The initial forecasting period covers the most recent two decades which ends at 2015. We fit this data using both models and produce forecasts up to 10 years ahead. The accuracy of these forecasts are then evaluated. Subsequently, the fitting period is then extended by a year and the forecasts for the same horizons are recalculated. This procedure of expanding the training window continues until it includes data up to the year 2014.

In order to gauge the accuracy of the point forecast of the HUts model, the mean squared errors (MSE) are calculated and compared with those of the forecasts of the HU model variants. MSE quantifies the average squared difference between the predicted and observed values, providing a measure of the overall accuracy and precision of predictions with lower values indicating better forecasting performance. The MSE is computed using:

$$MSE(h) = \frac{1}{pq} \sum_{t=1}^{q} \sum_{i=1}^{p} (y_t(x_i) - \hat{y}_{t|t-h}(x_i))^2, \tag{3.1}$$

where h is the forecasting horizon, p is the number of ages, q is the number of years considered in the forecasting period, $y_t(x_i)$ is the observed log mortality, and $\hat{y}_{t|t-h}(x_i)$ is the predicted value.

Figure 1 presents the plot of the root MSE of $h = 1, 5, 10$ forecasts across the different models. The square roots of the MSE values are plotted instead of the MSE themselves to ensure the plot scale accommodates all data. Readers are directed to Appendix D in the Supplementary Material for the tabulated MSE values of all countries. The results indicate that while the wHU model often achieves the lowest MSE values for one-step-ahead forecasts, the HUts model still performs competitively, frequently achieving values close to the best-performing models. In comparison, the HU model shows higher error rates, followed by the HUrob model.

For five-step-ahead forecasts, the HUts model demonstrates more competitive performance. Figure 1(b) shows that overall the HUts model achieves lower error metrics compared to the other models, suggesting an improvement in the HUts model's forecasting ability as the forecast horizon increases.

The results for ten-step-ahead forecasts, shown in Figure 1(c), further confirms the strong performance of the HUts model. It consistently achieves lower or comparable MSE, reinforcing its effectiveness in long-term mortality rate forecasting. In most countries, the HUts model has lower MSE values, with exceptions to a few cases such as in the United States, where the HU model occasionally performs better. This consistent performance across a majority of countries highlights the robustness of the HUts model in capturing long-term trends in mortality rates.

Across different forecasting horizons, the best model is generally a choice between the HUts and wHU models, as these models tend to have the lowest MSE values. The wHU model performs better at shorter forecast horizons, as seen in Figure 1(a). However, as the forecasting horizon increases, the HUts model demonstrates improved performance and eventually surpasses the wHU model. This shift indicates that the HUts model's accuracy and reliability become more pronounced with longer-term forecasts.

Notably, in the cases of France and Ireland, the HUts model performs consistently well across all forecasting horizons, maintaining lower MSE values compared to other models. Conversely, for Australia,

https://doi.org/10.1017/asb.2024.38 Published online by Cambridge University Press

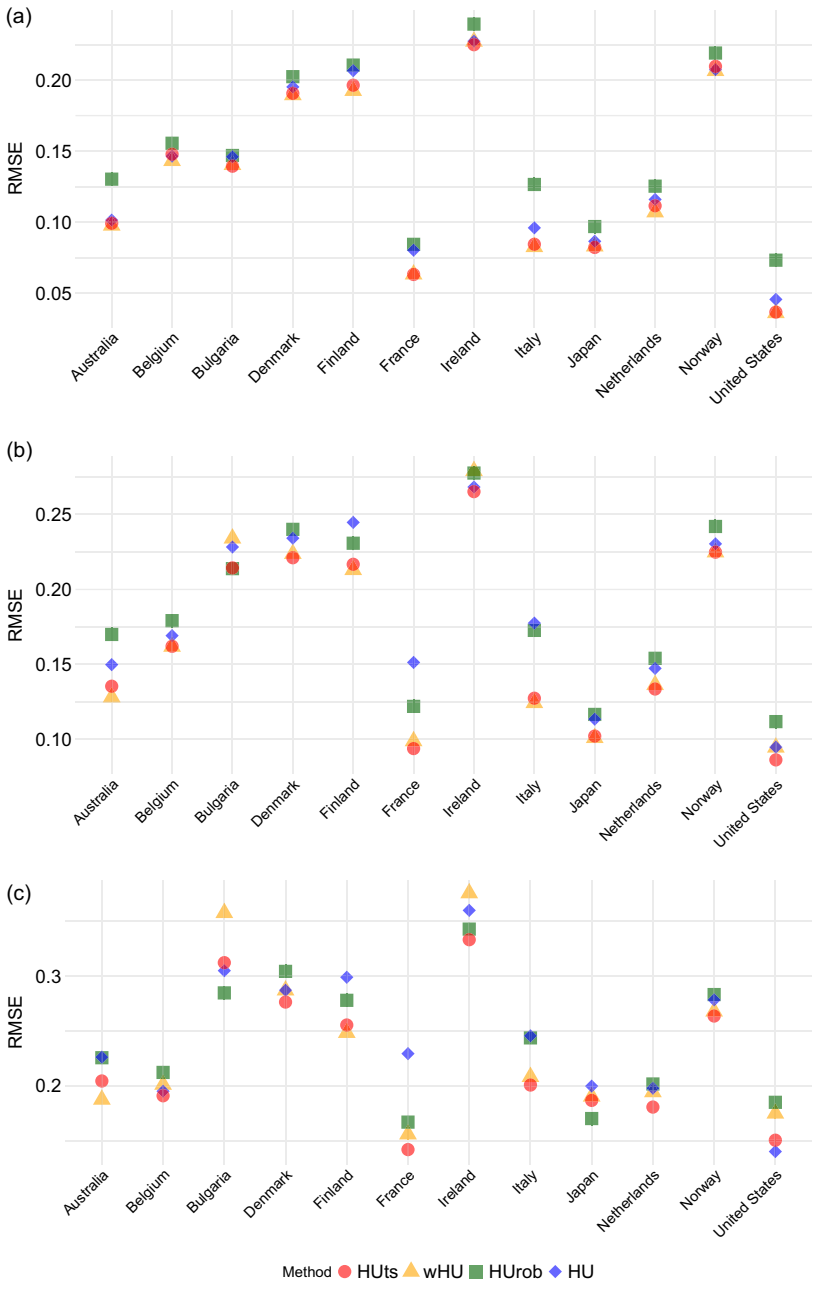


Figure 1. *Root MSE of (a) one, (b) five, and (c) ten-step-ahead point forecasts of log mortality rates by model and country.*

the HUts model does not appear to be the most favorable at any point across the forecast horizons, with other models, such as the wHU model, often yielding better results. The following section will take a closer look at the mortality rate forecasts for France and Australia, exploring the nuances and potential reasons behind the differing performances of the HUts model against the HU model variants.

https://doi.org/10.1017/asb.2024.38 Published online by Cambridge University Press

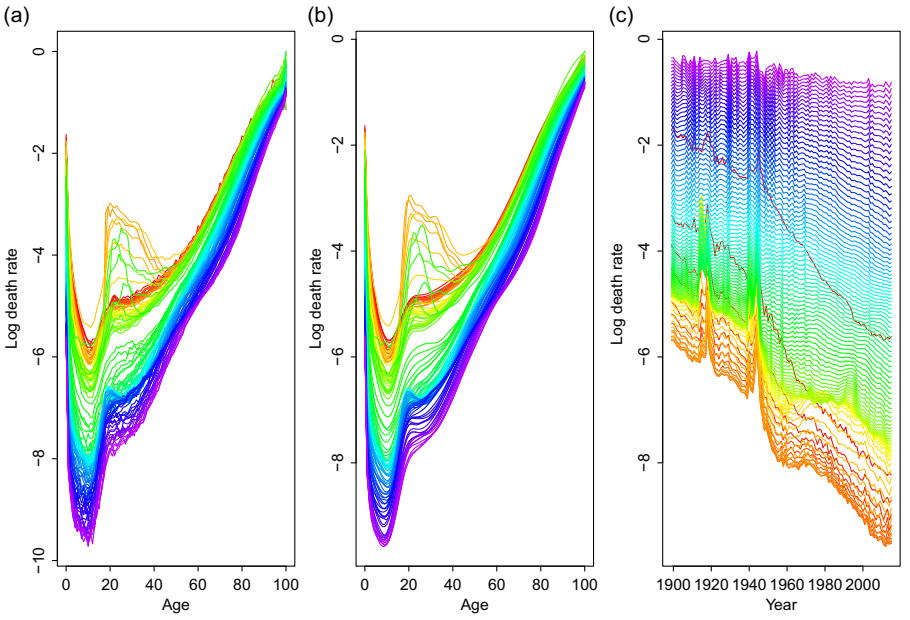


Figure 2. *France (1899–2015); (a) observed log mortality rates; (b) smoothed log mortality rates; (c) times series for ages.*

3.2.1. A closer look at the French and Australian mortality data

To assess the robustness of the HUts model, we examine the mortality rates of France and Australia more closely. Figures 2(a) and 3(a) include the observed log mortality rates, and Figures 2(b) and 3(b) show the smoothed log mortality rates with red representing the earlier years. Figures 2(c) and 3(c) the smoothed log mortality rates viewed as a time series with red representing the earlier ages. France's mortality data, notably influenced by periods of war and disease outbreaks, presents distinct outliers which are evident in the volatility observed in their mortality trends, as shown in Figure 2 which are most visible around the 1920s and 1940s. This provides an oppurtunity to test for the HUts model's ability to handle anomalies in data. In contrast, Australia's mortality data, starting from 1921 appears to have a smoother trend, without any significant outliers as depicted in Figure 3.

The ability of the HUts model to effectively handle anomalies in French mortality is illustrated in Figure 4(a), where the model demonstrates a substantial reduction in MSE across various forecast horizons when compared to the HU model. This significant improvement shows the HUts model's robustness against data irregularities, making it a reliable choice for forecasting in scenarios with potential outliers. Additionally, the HUts model showcases lower MSE than both the HUrob and wHU models, further affirming its superior performance in handling French mortality. Conversely, Australia presents a more stable dataset with mortality rates following a consistent trend and lacking significant outliers. The MSE plot for Australia in Figure 4(b) shows less pronounced differences compared to France. While the HUts model demonstrates superior performance over the HU and HUrob models, affirming its effectiveness even in more uniform datasets, the wHU model performs better than the HUts model. This suggests that the wHU model might be more suited for datasets with consistent trends and fewer irregularities. This comparative analysis highlights the importance of selecting appropriate models based on the characteristics of the dataset, with the HUts model excelling in more variable datasets and the wHU model in more stable ones. The MSE plots of the remaining countries can be found in Appendix E in the Supplementary Material.

https://doi.org/10.1017/asb.2024.38 Published online by Cambridge University Press

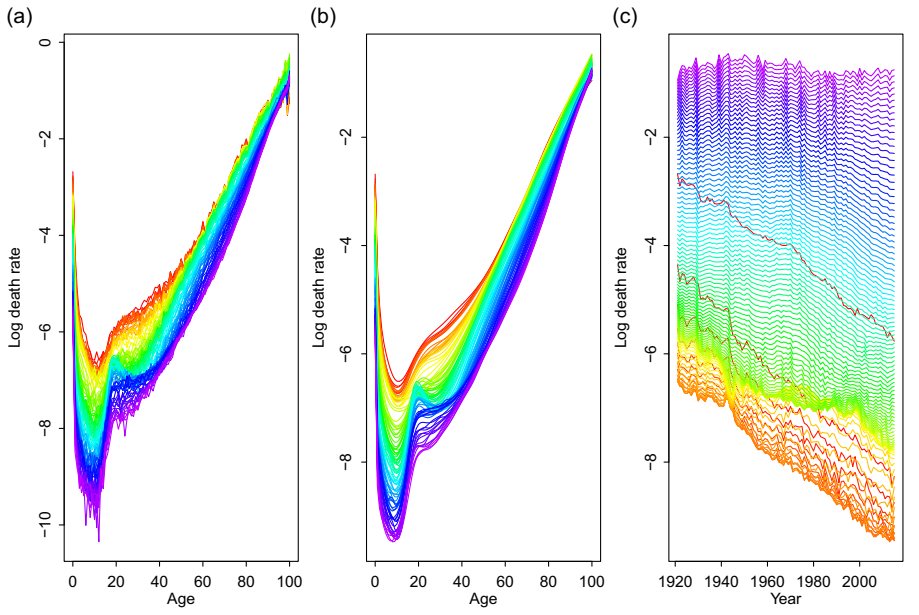


Figure 3. *Australia (1921–2015); (a) observed log mortality rates; (b) smoothed log mortality rates; (c) times series for ages.*

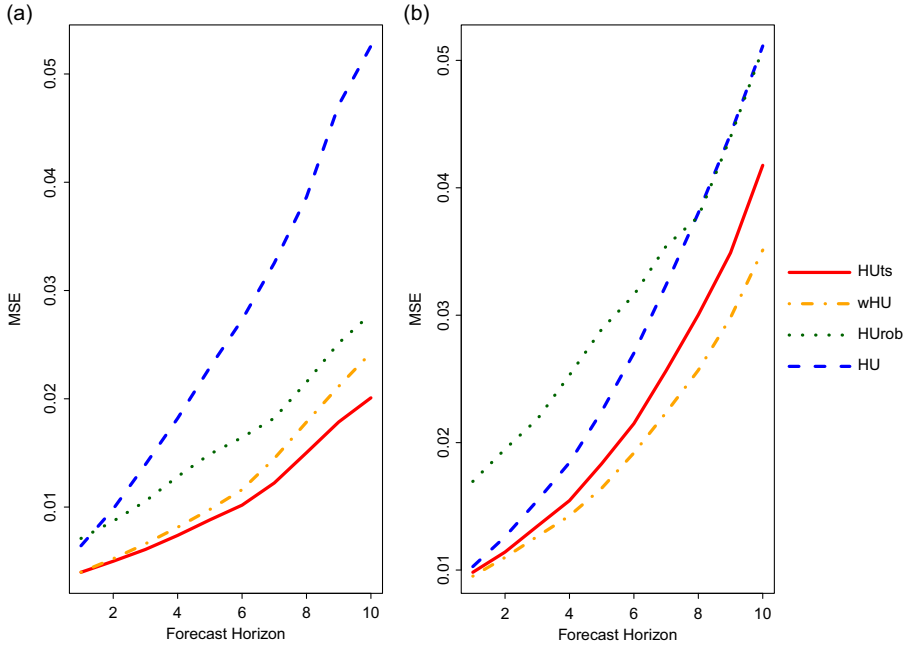


Figure 4. *MSE plots for (a) French and (b) Australian mortality of the HUts, wHU, HUrob, and HU models.*

https://doi.org/10.1017/asb.2024.38 Published online by Cambridge University Press

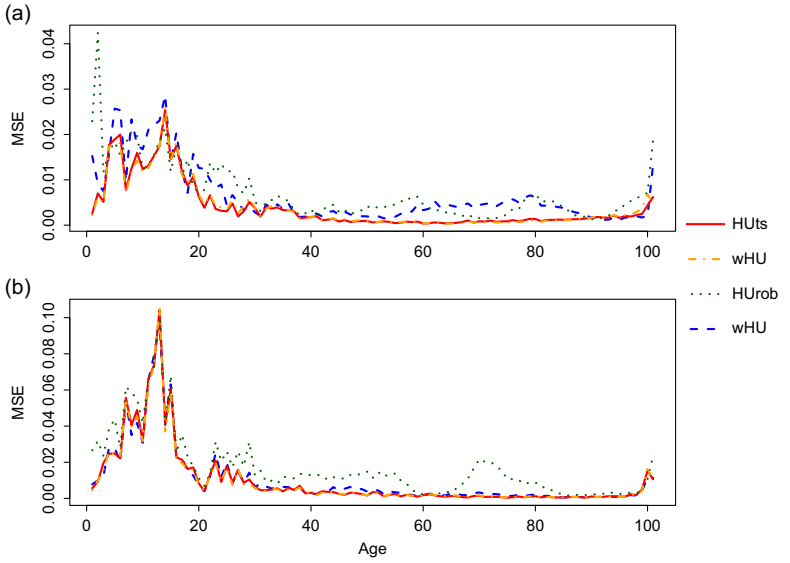


Figure 5. *One-step-ahead forecast MSE averaged over years in forecasting period of (a) French and (b) Australian mortality for the HUts, wHU, HUrob, and HU models.*

To gain deeper insights into the model performance, the MSE values are considered across the age spectrum. Instead of using Equation (3.1), the MSE values are averaged only over the years of the forecasting period. This approach will allow us to observe how the models perform at different age groups, providing a more granular understanding of the forecast accuracy. The MSE across ages are computed with the following:

$$MSE(h, x_i) = \frac{1}{q}\sum_{t=1}^{q}(y_t(x_i) - \hat{y}_{t|t-h}(x_i))^2.$$

Upon a closer inspection of the MSE across different age groups of the French mortality data in Figures 5(a), 6(a), and 7(a), the HUts model generally surpasses every other model in terms of forecasting performance. For one-step-ahead forecasts (Figure 5(a)), the HUts and wHU models show similar performance in terms of MSE, with both models outperforming the HU and HUrob models substantially. For five-step-ahead forecasts (Figure 6(a)), the HUts model continues to perform well, maintaining lower MSE values across most age groups compared to the other models. The HUts model shows improved performance at middle ages (30–50 years) over the wHU model, while the HU and HUrob models, although better at these ages, struggle during early and older ages, contributing to their higher overall error metric. For ten-step-ahead forecasts (Figure 7(a)), the HUts model's performance remains robust, achieving lower MSE values across most age groups. However, it still faces competition from the wHU model, which performs better in certain age ranges. Overall, these observations suggest that while the HUts model performs exceptionally well, particularly as the forecast horizon increases, the wHU model also shows strong performance, especially at shorter forecast horizons. The choice between these models often depends on the specific age group and forecasting horizon, with the HUts model becoming more favorable as the horizon increases.

Similarly, MSE for the Australian mortality forecasts in Figures 5(b), 6(b), and 7(b) highlight the competitive nature of the HUts model, particularly at shorter forecast horizons. For one-step-ahead forecasts (Figure 5(b)), the HUts model performs comparably to the wHU model, with both showing nearly identical MSE values. This indicates that the HUts model is effective at short-term forecasting. At the five-step-ahead horizon (Figure 6(b)), the HUts model continues to perform well but begins to show

https://doi.org/10.1017/asb.2024.38 Published online by Cambridge University Press

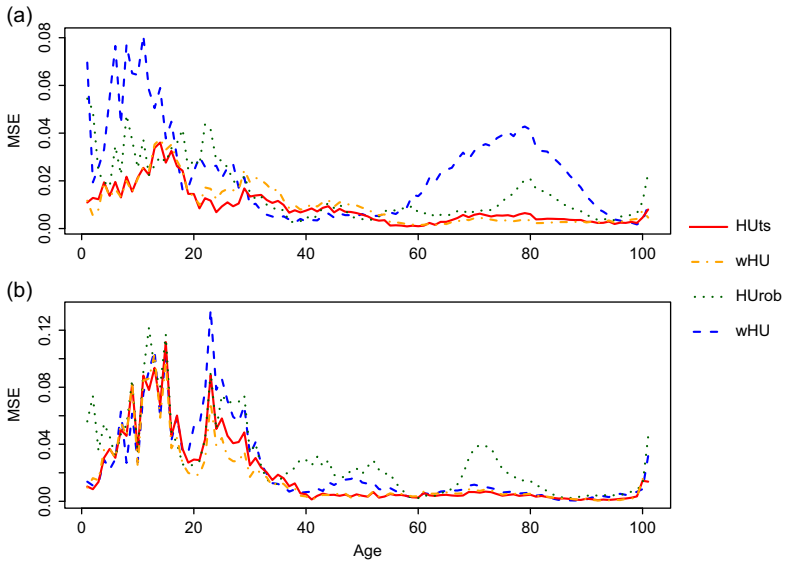


Figure 6. *Five-step-ahead forecast MSE averaged over years in forecasting period of (a) French and (b) Australian mortality for the HUts, wHU, HUrob, and HU models.*

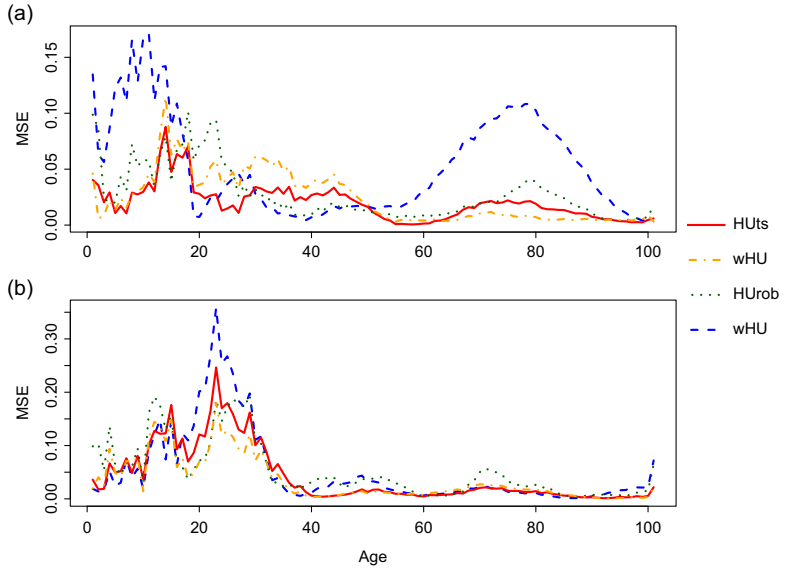


Figure 7. *Ten-step-ahead forecast MSE averaged over years in forecasting period of (a) French and (b) Australian mortality for the HUts, wHU, HUrob, and HU models.*

higher MSE values in the 20–40 age range compared to the wHU model. This underperformance in mid-age groups becomes more evident, reflecting some potential biases that start to emerge at this forecast horizon. By the ten-step-ahead forecasts (Figure 7(b)), the HUts model, although still competitive, shows further increased MSE values around ages 20–40 years. This trend highlights the model's underestimation (see Appendix F of the Supplementary Material) in this specific age group, significantly affecting the overall error metric. Despite this, the HUts model maintains reasonable performance across other

https://doi.org/10.1017/asb.2024.38 Published online by Cambridge University Press

Table 2. *Empirical coverage probabilities of bootstrapped prediction intervals.*

	HUts		wHU	
	95%	Adjusted 95%	95%	Adjusted 95%
Australia	92.70%	94.77%	88.30%	89.22%
Belgium	97.36%	96.09%	91.05%	81.13%
Bulgaria	91.89%	85.44%	82.51%	79.47%
Denmark	92.99%	95.84%	78.06%	91.17%
Finland	97.39%	98.80%	94.68%	98.97%
France	100.00%	100.00%	99.95%	100.00%
Ireland	92.81%	82.62%	73.54%	80.54%
Italy	99.84%	99.94%	99.42%	99.86%
Japan	97.08%	97.39%	95.84%	79.40%
Netherlands	99.61%	99.95%	98.15%	99.94%
Norway	94.30%	97.98%	88.26%	98.15%
USA	96.60%	99.11%	92.39%	85.70%

age groups, yet the wHU model generally achieves better results at these longer horizons, suggesting that the wHU model might be better suited for datasets with more consistent trends and fewer irregularities.

The findings of French and Australian mortality data present a comparative analysis of the HUts model's performance under differing demographic conditions and historical events influencing mortality data. While the HUts model is able to perform well in stable demographic conditions, its advantages are particularly noticeable when confronting data with potential outliers, making it a robust tool for forecasting when compared with the HU model variants.

3.3. Interval forecast

To evaluate the interval forecasts, the empirical coverage probability is calculated using the following expression (Hyndman and Shang, 2009):

$$\frac{1}{qph}\sum_{t=1}^{q}\sum_{j=1}^{h}\sum_{i=1}^{p}\mathbf{1}\left(\hat{y}_{t+j|t}^{(\alpha/2)}(x_i) < y_{t+j}(x_i) < \hat{y}_{t+j|t}^{(1-\alpha/2)}(x_i)\right),$$

where $\mathbf{1}(\cdot)$ is the indicator function and $\hat{y}_{t+j|t}^{(\alpha/2)}(x_i)$ is the $\alpha/2$-quantile from the bootstrapped samples. It quantifies whether the intervals are adequately wide to encompass the actual observations at a specified nominal coverage of $(1-\alpha)100\%$. 95% prediction intervals of the HUts model and the wHU model are constructed for all the 12 countries and tabulated in Table 2. The minimum observations required to obtain the one-step-ahead forecast errors used for the adjustment, r is set to 30.

The bootstrapping procedure generally performs well in obtaining prediction intervals that are close to the nominal 95% coverage level. Across various countries and forecasting methods, the unadjusted 95% prediction intervals yield empirical coverage probabilities that align reasonably well with the expected nominal level. These results demonstrate that the bootstrap approach reliably captures the underlying variability in mortality data, ensuring that the prediction intervals accurately achieve the desired 95% coverage.

However, when applying the Hyndman and Shang (2009) adjustment to these prediction intervals, the effectiveness of the adjustment varies across different countries and methods. This adjustment aims to refine the intervals based on one-step-ahead forecast errors, potentially bringing the empirical coverage probabilities closer to the nominal level. Yet, its impact is not uniformly beneficial and depends on the specific characteristics of the mortality data and the forecasting models used.

https://doi.org/10.1017/asb.2024.38 Published online by Cambridge University Press

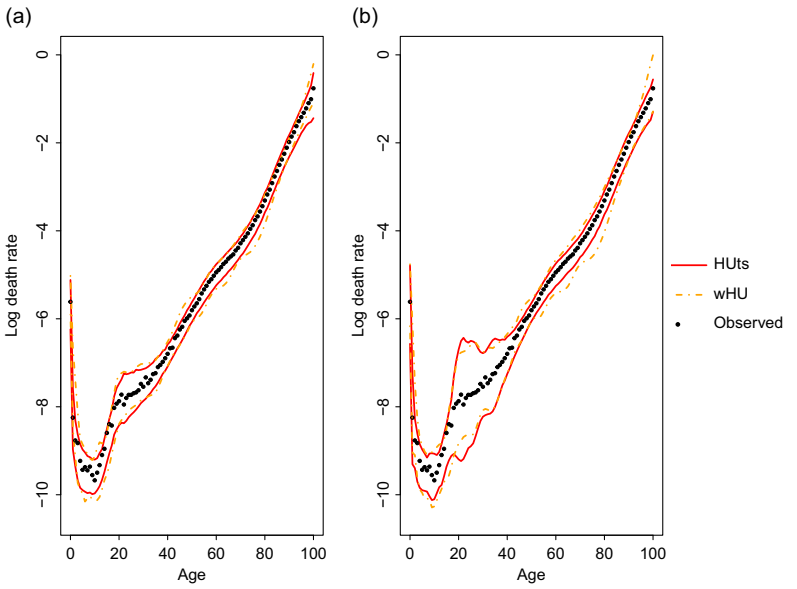


Figure 8. *Bootstrapped (a) 95% (b) adjusted 95% prediction intervals of one-step-ahead forecast for French mortality (1899–2014) for the HUts model and the wHU model.*

In the case of Australia, both the prediction intervals of the HUts and wHU models show improvements in coverage probabilities after the adjustment. For the HUts model, the empirical coverage increased from 92.70% to 94.77%, moving closer to the nominal 95% level. Similarly, for the wHU model, the coverage improves from 88.30% to 89.22%. This indicates that the adjustment effectively enhances the reliability of the prediction intervals in these instances. A similar positive effect is observed for Denmark with the wHU model, where the coverage probability significantly increases from 78.06% to 91.17% after adjustment, as well as from 92.99% to 95.84% with the HUts model.

On the other hand, the adjustment does not always lead to better coverage probabilities. For example, in Belgium using the wHU model, the coverage probability decreases from 91.05% to 81.13% after adjustment, moving further away from the nominal level. Similarly, in Bulgaria with the HUts model, the coverage dropped from 91.89% to 85.44% post-adjustment. These instances indicate that the adjustment may not universally improve the prediction intervals possibly due to the nature of the forecast errors of the fitted models.

A notable case is France, where the empirical coverage probabilities reaches 100% for both unadjusted and adjusted intervals with the HUts model. This phenomenon is caused by the prediction intervals absorbing a significant amount of uncertainty from the forecasts, resulting in wider intervals as seen in Figure 8. For the unadjusted prediction intervals, the uncertainty arises from bootstrapping the model errors. Due to the presence of outliers in the French mortality data, these model errors are larger. When the prediction intervals are adjusted, they incorporate additional variability from the in-sample one-step-ahead forecast errors. These forecast errors are also inflated by outliers, resulting in even wider intervals. This further widening makes the predictions more conservative to ensure coverage despite the high variability in the data. The same effect is observed with the wHU model, where the prediction intervals are similarly wide which suggests that outliers can impact the bootstrapping approach, potentially reducing the reliability of the intervals.

In contrast, when examining the prediction intervals for Australia in Figure 9, they are significantly narrower than those for France. Australia's mortality data is more stable, with fewer outliers, leading to less uncertainty in the forecasts. Consequently, the prediction intervals are narrower. This

https://doi.org/10.1017/asb.2024.38 Published online by Cambridge University Press

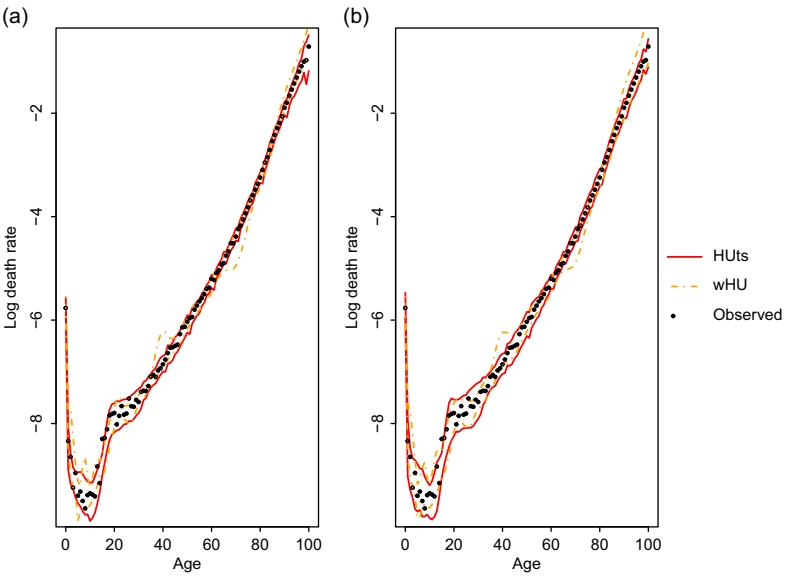


Figure 9. *Bootstrapped (a) 95% (b) adjusted 95% prediction intervals of one-step-ahead forecast for Australian mortality (1921–2014) for the HUts model and the wHU model.*

comparison highlights the point that uncertainty within mortality data leads to wider bootstrapped prediction intervals, as seen in France, while more stable data results in narrower intervals, as observed in Australia.

4. Discussion

In mortality modeling and forecasting, the HUts model introduces a novel methodology by taking advantage of the universal nonlinearity of signatures and applying signature regression to decompose mortality curves. Unlike the HU model, which relies on the covariance function and is sensitive to outliers, the HUts model learns data dependency through signature coefficients using principal component regression or SVD, effectively reducing the impact of outliers. The principal components obtained thus represent the structure and interactions within the data.

The HUts model performs competitively and often surpasses other HU model variants in various scenarios. Notably, the HUts model achieves these results without the need for weighted methods or robust statistics, which are employed by the wHU and HUrob models to improve their performance under specific conditions. This approach of the HUts model delivers robustness and is capable of adapting to diverse datasets. Whether dealing with irregularities as seen in the French mortality data or handling more stable demographic trends like those observed in Australia, the HUts model maintains high accuracy and reliability, making it a highly viable option for actuaries and demographers seeking dependable mortality forecasts without the complexity of additional methodological adjustments.

The HUts model also offers the advantage of easy modification to enhance performance, unlike the HU model's basis expansion. Therefore, suboptimal results are not indicative of inherent limitations but rather reflect its modular nature, allowing for greater adaptability. For instance, the embedding step is not limited to the basepoint, time and lead-lag transformation, Morrill *et al*. (2021) compiled various other augmentations that can be incorporated to improve the model and better represent the geometric characteristics of a path. Besides augmentations, applying the log-signature (Morrill *et al*., 2021) or a randomized signature (Cuchiero *et al*., 2020; Compagnoni *et al*., 2023) (see Appendix G

https://doi.org/10.1017/asb.2024.38 Published online by Cambridge University Press

in the Supplementary Material) over the signature can also be considered to reduce the feature vector dimensions. Morrill *et al*. (2021) also suggests that the scaling of paths or signatures may offer improvements.

It is also important to acknowledge the limitations of this study. The prediction intervals constructed under the HUts model face certain limitations, particularly due to the normality assumptions on error terms. Normality assumptions, when not justified, can lead to inaccuracies in the predicted intervals, necessitating the use of bootstrap methods. Bootstrap methods, although useful in addressing non-normality, come with more computational demand. This computational demand can be a significant drawback in practical applications where computational efficiency is crucial. In addition, bootstrap methods are sensitive to outliers which can affect the resampling process, leading to wider prediction intervals. This is evident in the case of French mortality data, which includes periods of war and disease outbreaks. These outliers introduce greater variability into the bootstrap samples, resulting in more conservative and wider intervals. This phenomenon is less pronounced in the mortality data, where the historical data is more stable and follows a general trend without significant outliers. Steps such as adapting the robust bootstrap procedure from Beyaztas and Shang (2022) can be taken to reduce the influence of outliers by assigning weights to model residuals.

In addition, the performance of the HUts model, like any other model, depends on the country's data and forecast horizon, with some models performing relatively better in certain settings. This emphasizes the influence of country-specific mortality dynamics and model sensitivities. As Shang (2012) pointed out, there is no universal mortality model capable of effectively handling data from all countries. This makes the HUts model a candidate for model averaging approaches. Model averaging combines the strengths of various models to improve forecast accuracy by assigning empirical weights to each model based on their performance at different horizons. Given the HUts model's demonstrated efficacy, especially in long-term forecasts, it can be assigned greater weight in model-averaged forecasts to leverage its strengths (Shang and Booth, 2020; Chang and Shi, 2023). The benefits of model averaging lie in enhancing forecast accuracy through a strategic combination of models. By integrating the HUts model into a model averaging framework, it is possible to achieve more reliable and accurate mortality forecasts, particularly over longer forecast horizons.

Beyond demography, the HUts model extends itself by offering a robust framework to address a range of challenges in functional data analysis, including functional principal component regression (FPCR), functional regression with functional responses, and forecasting functional time series. The flexibility of the HUts model to adapt to various functional data analysis scenarios is primarily due to its ability to integrate advanced mathematical concepts from functional data analysis, such as smoothing and handling of high-dimensional datasets. This integration not only enhances the model's performance in demographic forecasting but also makes it a powerful tool for broader applications in statistical and machine learning fields involving complex, structured data.

5. Conclusion

This paper has introduced the incorporation of signatures into the HU model, resulting in the HUts model. It is able to outperform the HU model variants while exhibiting enhanced robustness, particularly in the presence of outliers and irregularities in the data. The empirical results from the French and other mortality datasets provide compelling evidence of the HUts model's efficacy. In the case of France, with the presence of notable historical outliers due to wars and disease outbreaks, the HUts model showed substantial improvement over the HU model variants. This is evident from the lower MSE values across various forecast horizons. The enhanced performance can be attributed to the model's ability to mitigate the influence of outliers, which are more prevalent in the French data. This robustness is crucial for accurate mortality forecasting in regions with volatile historical trends. The HUts model's superior performance across all age groups, particularly at longer forecast horizons, displays the model's versatility across diverse demographic conditions. Prediction intervals were constructed with the application of

https://doi.org/10.1017/asb.2024.38 Published online by Cambridge University Press

bootstrap procedures. The coverage probabilities of the 95% bootstrapped prediction intervals generally are close to the intended nominal coverage, while the adjustment is able to improve said coverage for certain cases. The presence of outliers can impact the construction of the intervals, which is particularly pronounced in the case of French mortality, where the intervals constructed were more conservative and wide. In contrast, Australian mortality, with fewer outliers, shows smaller coverage deviance with its bootstrapped prediction interval. From a practical standpoint, these findings have significant implications for actuaries and policymakers. The improved accuracy and robustness of the HUts model can lead to more reliable mortality forecasts, which are essential for various applications, including insurance pricing, pension fund management, and public health planning. The model's ability to handle data with outliers and provide accurate long-term forecasts makes it a valuable tool for demography.

Future research should focus on refining the HUts model, potentially integrating more advanced smoothing techniques or exploring alternative statistical methods to enhance its performance further. Additionally, the model's application can be extended to multi-population mortality forecasting as signatures are an excellent dimension reduction tool, and incorporating socioeconomic factors as a binary dimension could provide deeper insights into mortality trends. The following practical extensions can be readily implemented:

1. Gender-specific HUts model

$$\text{Male: } f_{t,M}(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k} Z_k(x) + e_t(x),$$
$$\text{Female: } f_{t,F}(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k} Z_k(x) + e_t(x).$$

   This adaptation allows the model to capture and forecast mortality trends separately for different genders, taking into account potential differences in mortality patterns, improvements, and other gender-specific characteristics.

2. Fitting the HUts model with mortality improvement rates, $z_{x,t} = 2 \times \frac{1-\frac{m_{x,t}}{m_{x,t-1}}}{1+\frac{m_{x,t}}{m_{x,t-1}}}$ instead of raw mortality rates $m_{x,t}$ (Haberman and Renshaw, 2012). Let the underlying smoothed function of $z_{x,t}$ be $\gamma_t(x)$, then using the HUts we can decompose it into:

$$\gamma_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k} Z_k(x) + e_t(x).$$

   Modeling mortality improvement rates helps stabilize the mortality data series by focusing on changes over time, making it stationary and enhancing the accuracy of forecasts.

3. Coherent mortality forecasting using the product ratio method by Hyndman *et al*. (2012) by fitting $p_t(x) = \sqrt{f_{t,M}(x) \cdot f_{t,F}(x)}$ and $r_t(x) = \sqrt{\frac{f_{t,M}(x)}{f_{t,F}(x)}}$ using the HUts model and forecasting it h-steps-ahead. Then the h-steps-ahead forecasts of the gender-specific log mortality rates can be obtained as follows:

$$f_{n+h|n,M}(x) = p_{n+h|n}(x) \times r_{n+h|n}(x),$$
$$f_{n+h|n,F}(x) = \frac{p_{n+h|n}(x)}{r_{n+h|n}(x)}.$$

   This method ensures that the mortality forecasts for different subpopulations remain within realistic and historically observed relationships. Jiménez-Varón *et al*. (2024), Shang *et al*. (2022), Shang and Hyndman (2017) also provide potential directions for extending the HUts model for subpopulation modeling.

https://doi.org/10.1017/asb.2024.38 Published online by Cambridge University Press

Acknowledgments. The authors are grateful for the comments from the reviewers which led to the improvement of the manuscript.

Competing interests. The authors declare none.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/asb.2024.38

References

Andrès, H., Boumezoued, A. and Jourdain, B. (2024) Signature-based validation of real-world economic scenarios. *ASTIN Bulletin*, **54**(2), 410–440. https://doi.org/10.1017/asb.2024.12

Basellini, U., Camarda, C.G. and Booth, H. (2023) Thirty years on: A review of the Lee–Carter method for forecasting mortality. *International Journal of Forecasting*, **39**(3), 1033–1049. https://doi.org/10.1016/j.ijforecast.2022.11.002

Beyaztas, U. and Shang, H.L. (2022) Robust bootstrap prediction intervals for univariate and multivariate autoregressive time series models. *Journal of Applied Statistics*, **49**(5), 1179–1202. https://doi.org/10.1080/02664763.2020.1856351

Bjerre, D.S. (2022) Tree-based machine learning methods for modeling and forecasting mortality. *ASTIN Bulletin*, **52**(3), 765–787. https://doi.org/10.1017/asb.2022.11

Bonnier, P., Kidger, P., Arribas, I.P., Salvi, C. & Lyons, T. (2019) Deep signature transforms. https://doi.org/10.48550/arXiv.1905.08494

Booth, H., Maindonald, J. and Smith, L. (2002) Applying Lee–Carter under conditions of variable mortality decline. *Population Studies*, **56**(3), 325–336. https://doi.org/10.1080/00324720215935

Booth, H. and Tickle, L. (2008) Mortality modelling and forecasting: A review of methods. *Annals of Actuarial Science*, **3**(1–2), 3–43. https://doi.org/10.1017/S1748499500000440

Chang, L. and Shi, Y. (2023) Forecasting mortality rates with a coherent ensemble averaging approach. *ASTIN Bulletin*, **53**(1), 2–28. https://doi.org/10.1017/asb.2022.23

Chatfield, C. (1993) Calculating interval forecasts. *Journal of Business & Economic Statistics*, **11**(2), 121–135. https://doi.org/10.2307/1391361

Chen, K.-T. (1957) Integration of paths, geometric invariants and a generalized Baker–Hausdorff formula. *Annals of Mathematics*, **65**(1), 163–178. https://doi.org/10.2307/1969671

Cohen, S. N., Lui, S., Malpass, W., Mantoan, G., Nesheim, L., de Paula, Á., Reeves, A., Scott, C., Small, E. and Yang, L. (2023) Nowcasting with signature methods. https://doi.org/10.48550/arXiv.2305.10256

Compagnoni, E.M., Scampicchio, A., Biggio, L., Orvieto, A., Hofmann, T. and Teichmann, J. (2023) On the effectiveness of randomized signatures as reservoir for learning rough dynamics. *2023 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8. https://doi.org/10.1109/IJCNN54540.2023.10191624

Cuchiero, C., Gazzani, G., Möller, J. and Svaluto-Ferro, S. (2024) Joint calibration to SPX and VIX options with signature-based models. *Mathematical Finance*. https://doi.org/10.1111/mafi.12442

Cuchiero, C., Gazzani, G. and Svaluto-Ferro, S. (2023) Signature-based models: Theory and calibration. *SIAM Journal on Financial Mathematics*, **14**(3), 910–957. https://doi.org/10.1137/22M1512338

Cuchiero, C., Gonon, L., Grigoryeva, L., Ortega, J.-P. and Teichmann, J. (2020) Discrete-time signatures and randomness in reservoir computing. *IEEE Transactions on Neural Networks and Learning Systems*, **33**, 6321–6330. https://api.semanticscholar.org/CorpusID:225094517

Cuchiero, C. and Möller, J. (2024) Signature methods in stochastic portfolio theory. https://doi.org/10.48550/arXiv.2310.02322

Czado, C., Delwarde, A. and Denuit, M. (2005) Bayesian Poisson log-bilinear mortality projections. *Insurance: Mathematics and Economics*, **36**(3), 260–284. https://doi.org/10.1016/j.insmatheco.2005.01.001

Deprez, P., Shevchenko, P. and Wüthrich, M. (2017) Machine learning techniques for mortality modeling. *European Actuarial Journal*, **7**, 337–352. https://doi.org/10.1007/s13385-017-0152-4

Dokumentov, A., Hyndman, R.J. and Tickle, L. (2018) Bivariate smoothing of mortality surfaces with cohort and period ridges. *Stat*, **7**(1), e199. https://doi.org/10.1002/sta4.199

Fermanian, A. (2021) Embedding and learning with signatures. *Computational Statistics & Data Analysis*, **157**, 107148. https://doi.org/10.1016/j.csda.2020.107148

Fermanian, A. (2022) Functional linear regression with truncated signatures. *Journal of Multivariate Analysis*, **192**, 105031. https://doi.org/10.1016/j.jmva.2022.105031

Frévent, C. (2023) A functional spatial autoregressive model using signatures. https://doi.org/10.48550/arXiv.2303.12378

Haberman, S. and Renshaw, A. (2012) Parametric mortality improvement rate modelling and projecting. *Insurance: Mathematics and Economics*, **50**(3), 309–333. https://doi.org/10.1016/j.insmatheco.2011.11.005

Hainaut, D. (2018) A neural-network analyzer for mortality forecast. *ASTIN Bulletin*, **48**(2), 481–508. https://doi.org/10.1017/asb.2017.45

Hambly, B. and Lyons, T. (2010) Uniqueness for the signature of a path of bounded variation and the reduced path group. Annals of Mathematics, 109–167. https://www.jstor.org/stable/27799199

https://doi.org/10.1017/asb.2024.38 Published online by Cambridge University Press

Hubert, M., Rousseeuw, P.J. and Verboven, S. (2002) A fast method for robust principal components with applications to chemometrics. *Chemometrics and Intelligent Laboratory Systems*, **60**(1), 101–111. https://doi.org/https://doi.org/10.1016/S0169-7439(01)00188-5

Human Mortality Database. (2024) www.mortality.org

Hyndman, R.J. and Booth, H. (2008) Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting*, **24**(3), 323–342. https://doi.org/10.1016/j.ijforecast.2008.02.009

Hyndman, R.J., Booth, H. and Yasmeen, F. (2012) Coherent mortality forecasting: The product-ratio method with functional time series models. *Demography*, **50**(1), 261–283. https://doi.org/10.1007/s13524-012-0145-5

Hyndman, R.J. and Shang, H.L. (2009) Forecasting functional time series. *Journal of the Korean Statistical Society*, **38**(3), 199–211. https://doi.org/10.1016/j.jkss.2009.06.002

Hyndman, R.J. and Ullah, M.S. (2007) Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis*, **51**(10), 4942–4956. https://doi.org/10.1016/j.csda.2006.07.028

Jaber, E.A. and Gérard, L.-A. (2024) Signature volatility models: Pricing and hedging with fourier. https://doi.org/10.48550/arXiv.2402.01820

Jiménez-Varón, C.F., Sun, Y. and Shang, H.L. (2024) Forecasting high-dimensional functional time series: Application to sub-national age-specific mortality. *Journal of Computational and Graphical Statistics*, 1–15. https://doi.org/10.1080/10618600.2024.2319166

Kalsi, J., Lyons, T. and Arribas, I.P. (2020) Optimal execution with rough path signatures. *SIAM Journal on Financial Mathematics*, **11**(2), 470–493. https://doi.org/10.1137/19M1259778

Kidger, P. and Lyons, T. (2020) Universal approximation with deep narrow networks. *Proceedings of Thirty Third Conference on Learning Theory* (eds. J. Abernethy and S. Agarwal), vol. 125, pp. 2306–2327. PMLR. https://proceedings.mlr.press/v125/kidger20a.html

Kiraly, F.J. and Oberhauser, H. (2019) Kernels for sequentially ordered data. *Journal of Machine Learning Research*, **20**(31), 1–45. http://jmlr.org/papers/v20/16-314.html

Lee, R.D. and Carter, L.R. (1992) Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, **87**(419), 659–671. https://doi.org/10.2307/2290201

Lee, R.D. and Miller, T. (2001) Evaluating the performance of the Lee–Carter method for forecasting mortality. *Demography*, **38**(4), 537–549. https://doi.org/10.2307/3088317

Levin, D., Lyons, T. and Ni, H. (2016) Learning from the past, predicting the statistics for the future, learning an evolving system. https://doi.org/10.48550/arXiv.1309.0260

Lyons, T., Caruana, M. and Lévy, T. (2007) Differential equations driven by moderately irregular signals. In *Differential Equations Driven by Rough Paths: École d'été de probabilités de saint-flour xxxiv - 2004*, pp. 1–24. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-71285-5_1.

Lyons, T., Nejad, S. and Arribas, I.P. (2020) Non-parametric pricing and hedging of exotic derivatives. *Applied Mathematical Finance*, **27**(6), 457–494. https://doi.org/10.1080/1350486X.2021.1891555

Marino, M., Levantesi, S. and Nigri, A. (2023) A neural approach to improve the Lee–Carter mortality density forecasts. *North American Actuarial Journal*, **27**(1), 148–165. https://doi.org/10.1080/10920277.2022.2050260

Miyata, A. and Matsuyama, N. (2022) Extending the Lee–Carter model with variational autoencoder: A fusion of neural network and Bayesian approach. *ASTIN Bulletin*, **52**(3), 789–812. https://doi.org/10.1017/asb.2022.15

Morrill, J., Fermanian, A., Kidger, P. and Lyons, T. (2021) A generalised signature method for multivariate time series feature extraction. https://doi.org/10.48550/arXiv.2006.00873

Morrill, J., Kormilitzin, A., Nevado-Holgado, A. J., Swaminathan, S., Howison, S.D. and Lyons, T.J. (2020) Utilization of the signature method to identify the early onset of sepsis from multivariate physiological time series in critical care monitoring. *Critical Care Medicine*, **48**(10). https://doi.org/10.1097/CCM.0000000000004510

Nigri, A., Levantesi, S., Marino, M., Scognamiglio, S. and Perla, F. (2019) A deep learning integrated Lee–Carter model. *Risks*, **7**(1), 33. https://doi.org/10.3390/risks7010033

Perla, F., Richman, R., Scognamiglio, S. and Wüthrich, M.V. (2021) Time-series forecasting of mortality rates using deep learning. *Scandinavian Actuarial Journal*, **2021**(7), 572–598. https://doi.org/10.1080/03461238.2020.1867232

Ramsay, J.O. and Silverman, B.W. (2005) *Functional Data Analysis*. Springer. http://www.worldcat.org/isbn/9780387400808

Renshaw, A. and Haberman, S. (2003) Lee–Carter mortality forecasting: A parallel generalized linear modelling approach for England and Wales mortality projections. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **52**(1), 119–137. http://www.jstor.org/stable/3592636

Richman, R. and Wüthrich, M.V. (2021) A neural network extension of the Lee–Carter model to multiple populations. *Annals of Actuarial Science*, **15**(2), 346–366. https://doi.org/10.1017/S1748499519000071

Schnürch, S. and Korn, R. (2022) Point and interval forecasts of death rates using neural networks. *ASTIN Bulletin*, **52**(1), 333–360. https://doi.org/10.1017/asb.2021.34

Scognamiglio, S. (2022) Calibrating the Lee–Carter and the Poisson Lee–Carter models via neural networks. *ASTIN Bulletin*, **52**(2), 519–561. https://doi.org/10.1017/asb.2022.5

Shang, H.L. (2012) Point and interval forecasts of age-specific life expectancies: A model averaging approach. *Demographic Research*, **27**(21), 593–644. https://doi.org/10.4054/DemRes.2012.27.21

Shang, H.L. and Booth, H. (2020) Synergy in fertility forecasting: improving forecast accuracy through model averaging. *Genus*, **76**(1), 27. https://doi.org/10.1186/s41118-020-00099-y

Shang, H.L., Haberman, S. and Xu, R. (2022) Multi-population modelling and forecasting life-table death counts. *Insurance: Mathematics and Economics*, **106**, 239–253. https://doi.org/10.1016/j.insmatheco.2022.07.002

https://doi.org/10.1017/asb.2024.38 Published online by Cambridge University Press

Shang, H.L. and Hyndman, R.J. (2017) Grouped functional time series forecasting: An application to age-specific mortality rates. *Journal of Computational and Graphical Statistics*, **26**(2), 330–343. https://doi.org/10.1080/10618600.2016.1237877

Wang, B., Wu, Y., Taylor, N., Lyons, T., Liakata, M., Nevado-Holgado, A. J. and Saunders, K.E. (2020) Learning to detect bipolar disorder and borderline personality disorder with language and speech in non-clinical interviews. *Proc. Interspeech 2020*, pp. 437–441. https://doi.org/10.21437/Interspeech.2020-3040

Xie, Z., Sun, Z., Jin, L., Ni, H. and Lyons, T. (2018) Learning spatial-semantic context with fully convolutional recurrent network for online handwritten Chinese text recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **40**(8), 1903–1917. https://doi.org/10.1109/TPAMI.2017.2732978

https://doi.org/10.1017/asb.2024.38 Published online by Cambridge University Press