Functional regression with randomized signatures: An application to age-specific mortality rates

Zhong Jing Yap[0009−0000−5599−8515],
Dharini Pathmanathan[0000−0002−3279−4031] and
Sophie Dabo-Niang[0000−0002−4000−6752]

Abstract We propose a novel extension of the Hyndman-Ullah (HU) model to forecast mortality rates by integrating randomized signatures, referred to as the HU model with randomized signatures (HUrs). Unlike truncated signatures, which grow exponentially with order, randomized signatures, based on the Johnson–Lindenstrauss lemma, are able to approximate higher-order interactions in a computationally feasible way. Using mortality data from four countries, we evaluate the performance of the novel HUrs model compared to two alternative HU model versions. Our empirical results show that the proposed HUrs model performs well, particularly for Bulgarian and Japanese data.

1 Introduction

Over thirty years since the introduction of the Lee-Carter (LC) model, the mortality forecasting literature has seen a steady rise in new methodologies that adopt its extrapolation approach. Some notable extensions include adjustments to the time-index [2], addition of age-cohort interactions [3], [4] and [5] employed the LC methodology within a generalised linear model framework, and [6] extended the LC model to forecast multiple populations. Recent literature also see the deployment of

Zhong Jing Yap
Institute of Mathematical Sciences, Faculty of Science, Universiti Malaya, 50603, Kuala Lumpur, Malaysia, e-mail: sim190030@siswa.um.edu.my

Dharini Pathmanathan
Institute of Mathematical Sciences, Faculty of Science, Universiti Malaya, 50603, Kuala Lumpur, Malaysia, e-mail: dharini@um.edu.my

Sophie Dabo-Niang
UMR8524–Laboratoire Paul Painlevé, Inria-MODAL, University of Lille, CNRS, Lille, 59000, France;
CNRS–Université de Montréal, CRM–CNRS, Montréal, Canada, e-mail: sophie.dabo@univ-lille.fr

neural networks to forecast mortality as seen in [7, 8, 9]. See [10] for a recent review on the LC model.

Another well-known extension is the Hyndman-Ullah (HU) model [11], which uses the functional data paradigm [12] and applies nonparametric smoothing to view the mortality curves as functional time series, as well as functional principal component analysis (FPCA) to decompose the mortality curves. It distinguishes itself from the LC model by allowing multiple principal components to capture broader variation within the data, relaxes the homoskedasticity assumption of the LC model by allowing the error variance to vary with age, and employs more advanced time series methods in place of the random walk with drift model to forecast the time index. [11] reported superior point forecast accuracy over the LC model when forecasting French mortality. Variants of the HU model that have been proposed include a robust version [11], and a weighted version (wHU) [13].

Recently, [14] integrated signature methods from rough path theory into the HU model by replacing the FPCA decomposition with regression using truncated signatures, which is referred to as the HU with truncated signatures (HUts) model. This was motivated by [15] who noted that signatures are naturally adapted to vector-valued functions, and require mild assumption on the regularity of functions to encode nonlinear geometric information. These properties suggest signature regression to be well suited for extending the HU model's forecasting framework. The HUts model leverages the universality of signatures by representing the mortality data into a series of signature coefficients that capture the underlying linear and nonlinear interactions inherent within the data. As a result, the HUts model demonstrated remarkable robustness in forecasting mortality rates, particularly in situations where data irregularities may arise due to disease outbreaks or periods of war.

In this study, we investigate the utilization of an alternative signature representation based on linear random projections, specifically known as *randomized signatures*.

2 Signatures and Randomized Signatures

Signatures, rooted in the theory of rough paths, provide a systematic way to capture the geometric features of a time series or a functional trajectory through iterated integrals. Given a path $X : [0, T] \to \mathbb{R}^d$, its signature can be thought of as an infinite series of terms.

Definition 1 (Signature of a path [15]) *Let $X : [0, T] \to \mathbb{R}^d$ be a path of bounded variation, and let $I = (i_1, \ldots, i_k) \in \{1, \ldots, d\}^k$, 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*

$$S^I(X) = \int \cdots \int_{0 \le u_1 < \cdots < u_k \le T} dX_{u_1}^{i_1} \ldots dX_{u_k}^{i_k} \tag{1}$$

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:

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

The first term is always 1 (by convention). Subsequent terms are iterated integrals of the path increments, capturing increasingly higher-order interactions.

2.1 Truncated Signatures

While the signature is infinite-dimensional in theory, practical usage requires truncation to a fixed level m, denoted as:

$$S^m(X) = (1, S^{(1)}(X), \ldots, S^{(d)}(X), S^{(1,1)}(X), \ldots, S^{(i_1,\ldots,i_m)}(X)). \qquad (2)$$

This is called the *truncated signature of order m*. As higher order iterated integrals often become progressively smaller or less informative, the bulk of the geometric information about the path is captured by lower order terms, making the truncated signature an efficient local approximation.

2.2 Randomized Signatures

The randomized signature is a dimensionality reduced version of the classical path signature. While the truncated signature encodes a path through a selected series of iterated integrals to capture its geometric features, one of its main drawbacks is the number of signature coefficients grows exponentially with the truncation order. In a high-dimensional settings or situations where rich higher-order information is critical, this exponential increase becomes computationally infeasible. Conversely, opting for a lower truncation order may overlook important nonlinearity and higher order interactions in the data. Randomized signatures alleviate this computational burden by applying a random projection to the high-dimensional signature space, thereby obtaining a compact, yet highly expressive, feature representation.

At the core of randomized signatures lies the Johnson–Lindenstrauss (JL) lemma, a powerful result in high-dimensional geometry. The JL lemma ensures that high-dimensional data can be projected into a lower-dimensional space with minimal distortion. More precisely, we have the following result:

Lemma 1 (Johnson–Lindenstrauss Lemma [16]) *For every $0 < \epsilon < 1$ and every set Q consisting of N points in $\mathbb{R}^d$, there is a linear map $f : \mathbb{R}^d \to \mathbb{R}^k$ with $k > \frac{4 \log N}{3\epsilon^2 - 2\epsilon^3}$ such that*

$$(1 - \epsilon)\|u - v\|^2 \leq \|f(u) - f(v)\|^2 \leq (1 + \epsilon)\|u - v\|^2$$

for all $u, v \in Q$.

This property is crucial when reducing the dimensionality of the signature. Even after random projection, the essential geometric structure of the path is maintained. The definition of the randomized signature is as follows:

Definition 2 (Randomized Signature [17]) *Given $k \in \mathbb{N}$ and random matrices $A_1, \ldots, A_d \in \mathbb{R}^{k \times k}$, random shifts $b_1, \ldots, b_d \in \mathbb{R}^k$, random starting point $z \in \mathbb{R}^k$, and any fixed activation function σ, the randomized signature of X at $t \in [0, T]$ is the solution of the differential equation:*

$$dZ_t = \sum_{i=1}^{d} \sigma(A_i Z_t + b_i) \, dX_t^i, \quad Z_0 = z \in \mathbb{R}^k. \tag{3}$$

The random projection defined by these matrices is justified by the JL lemma, ensuring that the lower-dimensional embedding Z_t retains the approximation properties of the full signature.

The randomized signature maintains the universal approximation property of the full signature [18]. Recent theoretical results have demonstrated that, under appropriate conditions, one can approximate the evolution of controlled differential equations arbitrarily well by regressing on features generated via randomized signatures. Error bounds derived from the JL lemma provide explicit guarantees: if the projection dimension k is chosen sufficiently large, then the distortion induced by the random projection is kept within a pre-specified tolerance [16]. This balance between reduced dimensionality and controlled approximation error is key to applying signature methods in high-dimensional settings.

The computational benefits of randomized signatures are twofold. First, by reducing the number of coefficients that need to be computed, they mitigate the exponential growth inherent in truncated signatures. Second, this reduction results in lower model complexity, which in turn accelerates training.These properties make randomized signatures especially attractive in machine learning tasks such as time series classification, system identification, and forecasting.

See [17, 18, 19, 20] for a rigorous account on randomized signature.

3 Methodology

3.1 The Hyndman-Ullah with Randomized Signatures Model

The proposed Hyndman-Ullah with randomized signatures (HUrs) model extends the original HU framework by replacing the computation of truncated signatures in the HUts model with randomized signatures. This modification leverages the JL lemma to efficiently approximate higher-order interactions while mitigating the exponential growth of signature coefficients.

3.1.1 Randomized Signature Algorithm

Algorithm 1 details the procedure to generate the randomized signature of a path $X \in \mathbb{R}^d$.

Algorithm 1 Generate randomized signature [17]

Require: $X \in \mathbb{R}^d$ sampled at $0 = t_0 < t_1 < \cdots < t_N = T$, randomized signature dimension k, activation function σ.
1: Initialize $Z_{t_0} \in \mathbb{R}^k$, and for each $i \in \{1, \ldots, d\}$, set $A_i \in \mathbb{R}^{k \times k}$ and $b_i \in \mathbb{R}^k$ with i.i.d. standard normal entries.
2: **for** $n = 1$ to N **do**
3:    Compute:

$$Z_{t_n} = Z_{t_{n-1}} + \sum_{i=1}^{d} \sigma(A_i Z_{t_{n-1}} + b_i)\left(X_{t_n}^i - X_{t_{n-1}}^i\right)$$

4: **end for**

3.1.2 HUrs Model Procedure

The steps to forecast age-specific mortality rates $y_t(x)$ with the HUrs model are as follows:

1. **Smoothing:** Smooth the observed mortality rates using constrained weighted penalized regression splines to estimate the underlying mortality curves $f_t(x)$:

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

   where $\sigma_t(x)$ is the age-dependent error variance and ϵ_t is a standard normal error term.
2. **Mean adjustment:** Estimate the mean function $\mu(x)$ and compute the mean-adjusted rates:

$$f_t^*(x) = \hat{f}_t(x) - \hat{\mu}(x).$$

3. **Signature embedding:** For each age x_i, embed the time series $\{f_t^*(x_i)\}_{t=1}^{n}$ using basepoint augmentation, time augmentation, and the lead-lag transformation [14]. Then calculate its randomized signature as per Algorithm 1.
4. **Principal component regression:** Apply principal component regression with the computed randomized signatures as predictors and $f_t^*(x)$ as the response. This decomposes the mortality curve into

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

Table 1 Mortality data by country

Country	Commencing Year	Age Range
Belgium	1920–2015	0–100+
Bulgaria	1947–2015	0–100+
Japan	1947–2015	0–100+
United States	1933–2015	0–100+

where $\beta_{t,k}$ are the time-dependent coefficients, $Z_k(x)$ is the kth principal component obtained via interpolation, and $e_t(x)$ is the model error.

5. **Forecasting:** Fit univariate ARIMA models to the coefficients $\beta_{t,k}$ to forecast $\beta_{t+h,k}$ for a forecasting horizon h. Reconstruct the forecasted mortality rates $y_{t+h}(x)$ using the forecasted coefficients.

The Hyndman-Ullah with truncated signatures (HUts) model follows the same procedure, except that in step 3 truncated signatures are computed in lieu of randomized signatures.

4 Finite sample properties

In this section, we assess the finite sample performance of the HUrs model by applying it to empirical mortality data. The evaluation is based on forecasting accuracy across multiple countries and forecasting horizons, enabling a comprehensive comparison with alternative models. First, we describe the mortality datasets used in our analysis, which span different countries, time periods, and age ranges. Following the data description, we present a detailed analysis of point forecasts, measured by the mean squared error (MSE), to demonstrate the performance of the proposed model.

4.1 Data description

Mortality data from four countries are obtained from the Human Mortality Database [21]. Age groups were created, 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. These are tabulated in Table 1. Figure 1 illustrates the log mortality rates of Japan.
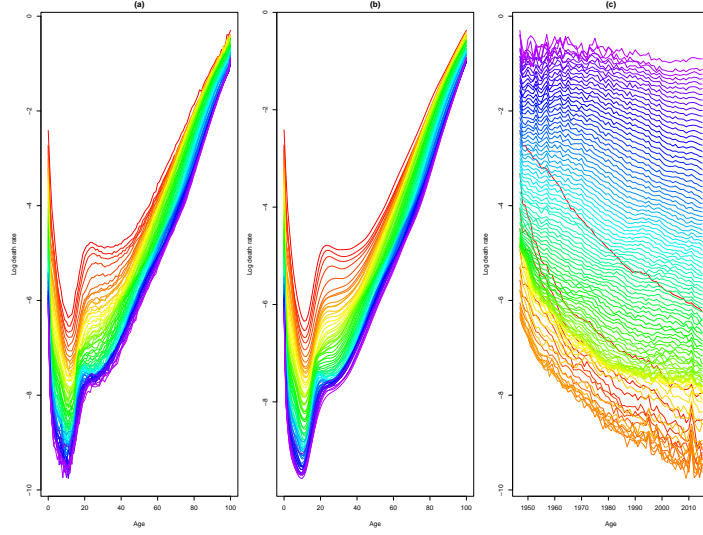
Fig. 1 Japan (1947 to 2015); (a) Observed log mortality rates; (b) Smoothed log mortality rates; (c) Time series for ages.

4.2 Point forecast

The mortality rates of four countries were forecasted using the HUrs model with $k = 100$ and a linear activation function $\sigma = \frac{1}{d\sqrt{k}}$. Treatment of the path embeddings will result with a path dimension, d of 3 as in [14]. Consequently, it is intuitive to compare the forecast accuracy of the proposed method with the HUts model and the weighted HU (wHU) model, which were the best performing models in [14].

The models were fitted using an expanding window approach, incorporating data from each country's beginning years as outlined in Table 1. The initial forecasting period covers the last two decades, ending at 2015. We fit the data with both models and produce forecasts up to ten years ahead. We evaluate the accuracy of these forecasts. Then, we extend the training period by a year and recalculate the forecasts. This process continues until 2014.

To assess the accuracy of the HUrs model's point forecast, mean squared errors (MSE) are computed and compared to those of the HU model variants. MSE quantifies the average squared difference between predicted and observed values, indicating the overall accuracy and precision of predictions. Lower MSE values indicate better forecasting performance. The MSE is calculated using the following equation:

$$MSE(h) = \frac{1}{pq} \sum_{t=1}^{q} \sum_{i=1}^{p} (y_t(x_i) - \hat{y}_{t|t-h}(x_i))^2, \tag{4}$$

Table 2 MSE: HUrs, HUts, and wHU across different countries for $h = 1$

Country	HUrs	HUts	wHU
Belgium	0.02371	0.02183	**0.02047**
Bulgaria	**0.01930**	0.01942	0.01968
Japan	0.00680	**0.00674**	0.00687
United States	0.00162	0.00133	**0.00129**

Table 3 MSE: HUrs, HUts, and wHU across different countries for $h = 5$

Country	HUrs	HUts	wHU
Belgium	0.04172	0.02627	**0.02620**
Bulgaria	**0.04442**	0.04595	0.05478
Japan	**0.01000**	0.01045	0.01020
United States	0.01023	**0.00746**	0.00896

Table 4 MSE: HUrs, HUts, and wHU across different countries for $h = 10$

Country	HUrs	HUts	wHU
Belgium	0.09131	**0.0364**	0.0404
Bulgaria	**0.08949**	0.0976	0.1280
Japan	**0.03289**	0.0348	0.0362
United States	0.03197	**0.0226**	0.0305

where h represents the forecasting horizon, p denotes the number of ages, q signifies the number of years considered, $y_t(x_i)$ denotes observed log mortality, and $\hat{y}_{t|t-h}(x_i)$ represents the predicted value. The results are tabulated in Tables 2, 3 and 4.

The result demonstrates that different models excel for different countries and forecasting horizons. Specifically, the HUrs model consistently outperforms others for Bulgaria across all three forecasting horizons. Additionally, in Japan, the HUrs model shows strong performance for both $h = 5, 10$. These findings highlight the importance of selecting appropriate models tailored to each country's unique characteristics and the specific forecasting periods.

5 Conclusion and discussions

In this study, we present a new extension of the Hyndman-Ullah (HU) model for forecasting mortality rates by incorporating randomized signatures. We assess the finite sample properties of the proposed model in comparison with two alternative versions of the HU model. The HUrs model demonstrates robust performance across the selected countries, particularly in Bulgaria and Japan. Furthermore, the modular design of the HUrs methodology indicates significant potential for optimization.

Prospective advancements might involve experimenting with diverse embeddings, fine-tuning the k values, and evaluating various activation functions to enhance their adaptability and performance across different scenarios.

Furthermore, beyond its application in mortality forecasting, the randomized signature method shows significant potential for tasks involving high-dimensional data. In such applications, it can greatly reduce computational costs and processing time, making it a valuable tool for efficiently handling complex datasets.

Competing Interests The authors have no conflicts of interest to declare that are relevant to the content of this chapter.

References

1. Lee, R.D., Carter, L.R.: Modelling and forecasting U.S. mortality. J. Am. Stat. Assoc. **87**(49), 659–671 (1992). doi:10.2307/2290201
2. Lee, R.D., Miller, T.: Evaluating the performance of the Lee–Carter method for forecasting mortality. Demography **38**, 537–549 (2001). doi:10.1353/dem.2001.0036
3. Booth, H., Maindonald, J., Smith, L.: Applying Lee–Carter under conditions of variable mortality decline. Popul. Stud. **56**(3), 325–336 (2002). doi:10.1080/00324720215935
4. Renshaw, A., Haberman, S.: Lee–Carter mortality forecasting: A parallel generalized linear modelling approach for England and Wales mortality projections. J. R. Stat. Soc. Ser. C (Appl. Stat.) **52**(1), 119–137 (2003). doi:10.1111/1467-9876.00393
5. Currie, I.D.: Smoothing constrained generalized linear models with an application to the Lee–Carter model. Stat. Model. **13**(1), 69–93 (2013). doi:10.1177/1471082X12471373
6. Li, N., Lee, R.: Coherent mortality forecasts for a group of populations: An extension of the Lee–Carter method. Demography **42**(3), 575–594 (2005). doi:10.1353/dem.2005.0021
7. Richman, R., Wüthrich, M.V.: A neural network extension of the Lee–Carter model to multiple populations. Ann. Actuar. Sci. **15**(2), 346–366 (2021). doi:10.1017/S1748499519000071
8. Schnürch, S., Korn, R.: Point and interval forecasts of death rates using neural networks. ASTIN Bull. **52**(1), 333–360 (2022). doi:10.1017/asb.2021.34
9. Marino, M., Levantesi, S., Nigri, A.: A neural approach to improve the Lee–Carter mortality density forecasts. N. Am. Actuar. J. **27**(1), 148–165 (2022). doi:10.1080/10920277.2022.2050260
10. Basellini, U., Camarda, C.G., Booth, H.: Thirty years on: A review of the Lee–Carter method for forecasting mortality. Int. J. Forecast. **39**(3), 1033–1049 (2023). doi:10.1016/j.ijforecast.2022.11.002
11. Hyndman, R.J., Ullah, M.S.: Robust forecasting of mortality and fertility rates: A functional data approach. Comput. Stat. Data Anal. **51**(10), 4942–4956 (2007). doi:10.1016/j.csda.2006.07.028
12. Ramsay, J.O., Silverman, B.W.: Functional Data Analysis, 2nd edn. Springer, New York (2005). doi:10.1007/b98888
13. Hyndman, R.J., Shang, H.L.: Forecasting functional time series. J. Korean Stat. Soc. **35**(3), 199–211 (2009). doi:10.1016/j.jkss.2009.06.002
14. Yap, Z.J., Pathmanathan, D., Dabo-Niang, S.: Forecasting mortality rates with functional signatures. ASTIN Bull. **55**(1), 97–120 (2025). doi:10.1017/asb.2024.38
15. Fermanian, A.: Functional linear regression with truncated signatures. J. Multivar. Anal. **192**, 105031 (2007). doi:10.1016/j.jmva.2022.105031
16. Debus, P.: Randomized signatures and applications in machine learning. Thesis, FernUniversität in Hagen (2024)
17. Compagnoni, E.M., Scampicchio, A., Biggio, L., Orvieto, A., Hofmann, T., Teichmann, J.: On the effectiveness of randomized signatures as reservoir for learning rough dynamics. In: 2023 International Joint Conference on Neural Networks (IJCNN), pp. 1–8 (2023). doi:10.1109/IJCNN54540.2023.10191624
18. Cuchiero, C., Gonon, L., Grigoryeva, L., Ortega, J.-P., Teichmann, J.: Discrete-time signatures and randomness in reservoir computing. IEEE Trans. Neural Netw. Learn. Syst. **33**, 6321–6330 (2020). doi:10.1109/TNNLS.2021.3076777
19. Cuchiero, C., Gazzani, G., Svaluto-Ferro, S.: Signature-based models: theory and calibration. SIAM J. Financ. Math. **14**(3), 910–957 (2023). doi:10.1137/22M1512338
20. Biagini, F., Gonon, L., Walter, N.: Universal randomised signatures for generative time series modelling. arXiv preprint arXiv:2406.10214 (2024). doi:10.48550/arXiv.2406.10214
21. Human Mortality Database (2024). www.mortality.org