Supplementary Material for "Forecasting mortality rates with functional signatures"

Yap Zhong Jing[1]   Dharini Pathmanathan[1,2,3]   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]UMR 8524—Laboratoire Paul Painlevé, Inria-MODAL, University of Lille, CNRS, Lille, 59000, France
[5]CNRS—Université de Montréal, CRM—CNRS, Montréal, Canada

Appendix A   Normality assumptions of errors

To obtain prediction intervals using distributional forecasts, all sources of error must satisfy the normality assumptions. If any source of error does not follow a normal distribution, bootstrapping methods should be used instead. While it is unrealistic to expect perfect normality, we prefer that most of the residuals resemble a normal distribution; otherwise, bootstrapped methods are preferred. We have plotted the residual histograms and QQ plots of ages 0, 10, 25, 50, 75, and 100 for all four models fitted with French (1899-2014) mortality data . Only the residuals of the HU model, seen in Figure 2, seem to resemble somewhat of a normal distribution. The residuals from the three remaining models, seen in Figures 1, 3, and 4, deviate significantly from a normal distribution and would benefit from the bootstrapping procedure when constructing prediction intervals.
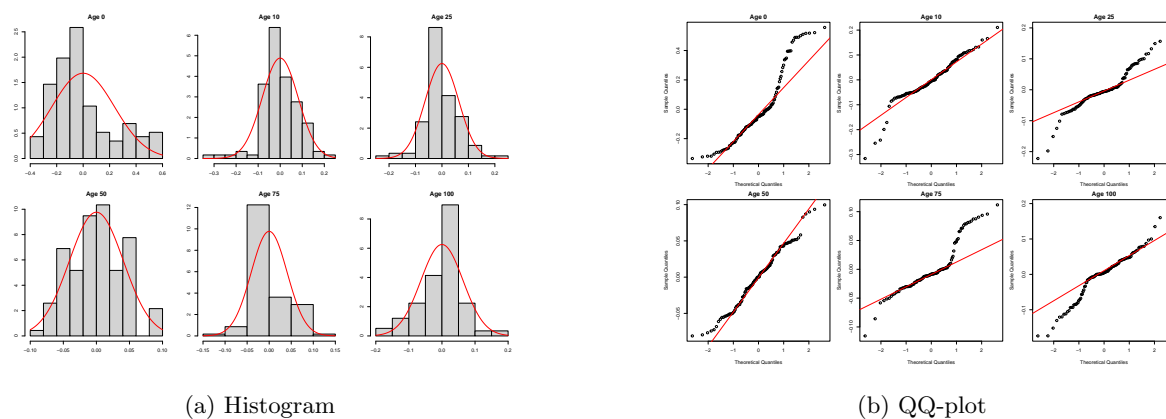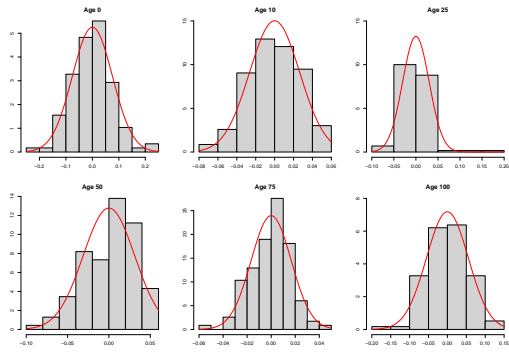

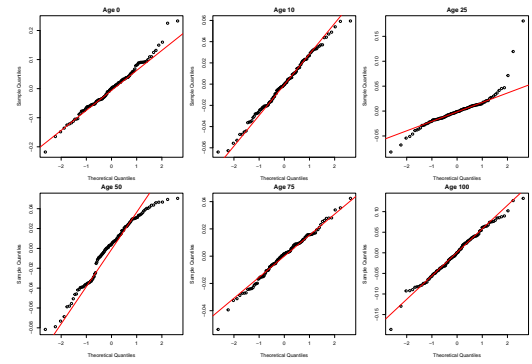
(a) Histogram

(b) QQ-plot

Figure 1: HUts model residuals
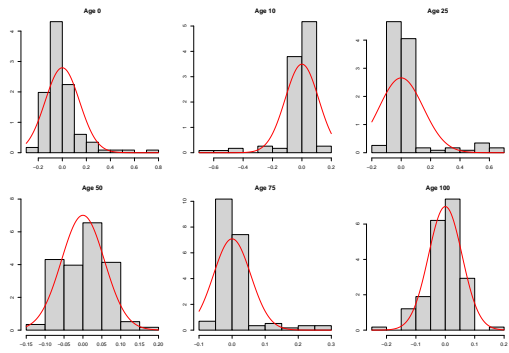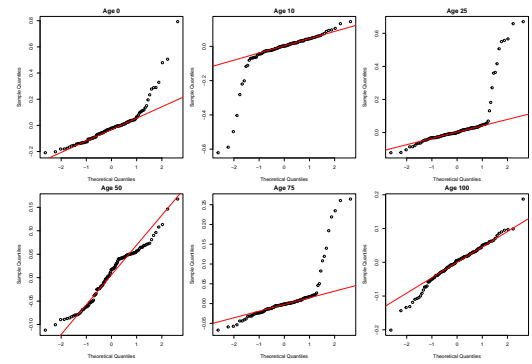
(a) Histogram

(b) QQ-plot
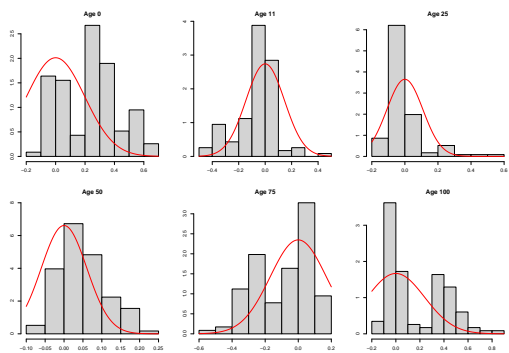
Figure 2: HU model residuals
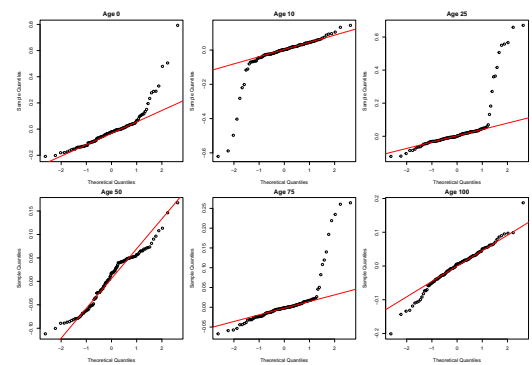

(a) Histogram

(b) QQ-plot

Figure 3: HUrob model residuals

(a) Histogram

(b) QQ-plot

Figure 4: wHU model residuals

Appendix B   Selection of truncation order

As the number of signature coefficients increases exponentially with the truncation order, we employ a naive search based on the MSE and MAE of the model's forecasting performance. In signature regression, it is generally advisable to keep the truncation order, m, below 5. A truncation order of 5 already includes 182 signature coefficients, significantly impacting the computational effort required. This approach is demonstrated using French mortality data. As depicted in Figure 5, the MSE and MAE values of the HUts model with $m = 3, 4, 5$ are larger than when $m = 2$. Since our path dimension is only 3, a high order of signature coefficients may not be necessary to accurately represent the path.
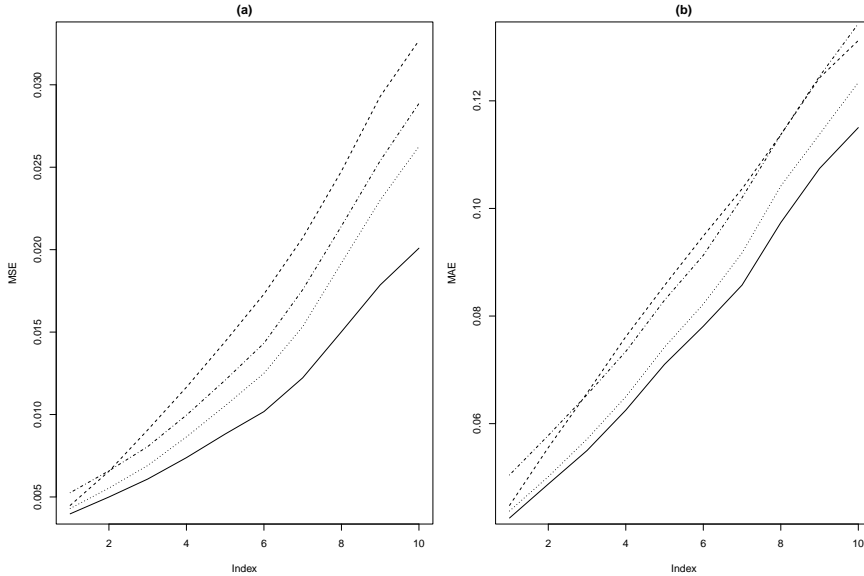

Figure 5: (a) MSE and (b) MAE plots for French mortality of the HUts model with $m = 2$ (solid line), $m = 3$ (dashed line), $m = 4$ (dotted line), and $m = 5$ (dash-dotted line)

Appendix C   Data availability and codes

The codes are available upon request.

Appendix D   Mean squared error and mean absolute error tables

Table 1: MSE of one-step-ahead point forecasts of log mortality rates by method and country

	HUts	HU	HUrob	wHU
Australia	0.00982	0.01027	0.01696	**0.00952**
Belgium	0.02183	0.02144	0.02419	**0.02047**
Bulgaria	**0.01942**	0.02131	0.02158	0.01968
Denmark	0.03636	0.03810	0.04098	**0.03590**
Finland	0.03856	0.04271	0.04435	**0.03705**
France	**0.00398**	0.00642	0.00710	0.00399
Ireland	**0.05062**	0.05174	0.05735	0.05161
Italy	0.00711	0.00920	0.01600	**0.00680**
Japan	**0.00674**	0.00747	0.00938	0.00687
Netherlands	0.01242	0.01344	0.01569	**0.01143**
Norway	0.04397	0.04303	0.04800	**0.04264**
United States	0.00133	0.00207	0.00535	**0.00129**
Mean	0.02101	0.02227	0.02558	**0.02061**

Table 2: MSE of five-step-ahead point forecasts of log mortality rates by method and country

	HUts	HU	HUrob	wHU
Australia	0.01835	0.02245	0.02890	**0.01642**
Belgium	0.02627	0.02863	0.03211	**0.02620**
Bulgaria	0.04595	0.05212	**0.04574**	0.05478
Denmark	**0.04891**	0.05484	0.05764	0.05007
Finland	0.04698	0.05988	0.05328	**0.04538**
France	**0.00882**	0.02290	0.01489	0.00978
Ireland	**0.07035**	0.07198	0.07707	0.07785
Italy	0.01625	0.03153	0.02982	**0.01546**
Japan	0.01045	0.01288	0.01361	**0.01020**
Netherlands	**0.01785**	0.02171	0.02374	0.01859
Norway	0.05056	0.05309	0.05856	**0.05053**
United States	**0.00746**	0.00900	0.01252	0.00896
Mean	**0.03068**	0.03675	0.03732	0.03202

Table 3: MSE of ten-step-ahead point forecasts of log mortality rates by method and country

	HUts	HU	HUrob	wHU
Australia	0.0418	0.0511	0.0509	**0.0351**
Belgium	**0.0364**	0.0380	0.0450	0.0404
Bulgaria	0.0976	0.0931	**0.0811**	0.1280
Denmark	**0.0764**	0.0825	0.0927	0.0825
Finland	0.0653	0.0894	0.0773	**0.0617**
France	**0.0201**	0.0526	0.0278	0.0242
Ireland	**0.1111**	0.1297	0.1177	0.1413
Italy	**0.0403**	0.0603	0.0594	0.0433
Japan	0.0348	0.0399	**0.0289**	0.0362
Netherlands	**0.0326**	0.0391	0.0406	0.0377
Norway	**0.0695**	0.0775	0.0802	0.0718
United States	0.0226	**0.0196**	0.0342	0.0305
Mean	**0.0540**	0.0644	0.0613	0.0611

The mean absolute error expresses the average difference between predicted and observed values and is computed using:

$$MAE(h) = \frac{1}{pq} \sum_{t=1}^{q} \sum_{i=1}^{p} |y_t(x_i) - \hat{y}_{t|t-h}(x_i)|.$$

Table 4: MAE of one-step-ahead point forecasts of log mortality rates by method and country

	HUts	HU	HUrob	wHU
Australia	0.06111	0.06614	0.09498	**0.06036**
Belgium	0.08404	0.09092	0.10116	**0.08193**
Bulgaria	**0.09106**	0.10052	0.09727	0.09138
Denmark	0.11169	0.12029	0.12909	**0.11127**
Finland	0.11279	0.12608	0.13552	**0.10976**
France	**0.04246**	0.06100	0.06336	0.04256
Ireland	**0.12512**	0.13419	0.14838	0.12719
Italy	0.05330	0.06842	0.08936	**0.05210**
Japan	0.05061	0.05701	0.06955	**0.04975**
Netherlands	0.06802	0.07839	0.08692	**0.06657**
Norway	0.11718	0.11890	0.13299	**0.11479**
United States	0.02648	0.03459	0.05380	**0.02628**
Mean	0.07866	0.08804	0.10020	**0.07783**

Table 5: MAE of five-step-ahead point forecasts of log mortality rates by method and country

	HUts	HU	HUrob	wHU
Australia	0.09131	0.11006	0.12925	**0.08682**
Belgium	**0.10163**	0.11696	0.12482	0.10307
Bulgaria	0.15511	0.17190	**0.15103**	0.15636
Denmark	**0.14989**	0.16413	0.17112	0.15113
Finland	0.14040	0.17654	0.15844	**0.13800**
France	**0.07102**	0.12234	0.09527	0.07427
Ireland	**0.16031**	0.18212	0.18869	0.17752
Italy	0.09528	0.14200	0.13301	**0.08831**
Japan	0.07659	0.08745	0.08784	**0.07277**
Netherlands	**0.09511**	0.11317	0.11668	0.09702
Norway	0.13801	0.15506	0.16166	**0.13688**
United States	**0.06428**	0.07333	0.08508	0.06843
Mean	**0.11158**	0.13459	0.13357	0.11255

Table 6: MAE of ten-step-ahead point forecasts of log mortality rates by method and country

	HUts	HU	HUrob	wHU
Australia	0.1483	0.1705	0.1749	**0.1383**
Belgium	**0.1323**	0.1421	0.1558	0.1434
Bulgaria	0.2490	0.2389	**0.2093**	0.2341
Denmark	**0.2097**	0.2182	0.2352	0.2151
Finland	0.1804	0.2334	0.2105	**0.1754**
France	**0.1150**	0.1905	0.1360	0.1204
Ireland	**0.2164**	0.2663	0.2510	0.2641
Italy	0.1623	0.2056	0.2019	**0.1550**
Japan	0.1410	0.1536	**0.1289**	0.1435
Netherlands	**0.1448**	0.1601	0.1637	0.1564
Norway	**0.1809**	0.2089	0.2087	0.1837
United States	**0.1121**	0.1122	0.1380	0.1250
Mean	**0.1660**	0.1917	0.1845	0.1712

Appendix E   Remaining MSE and MAE plots

In this section, we categorized the MSE and MAE plots of the remaining 10 countries into two groups: those favorable to the HUts model and those where the HUts model is comparable to the other models.

Appendix E.1   Favourable to the HUts model
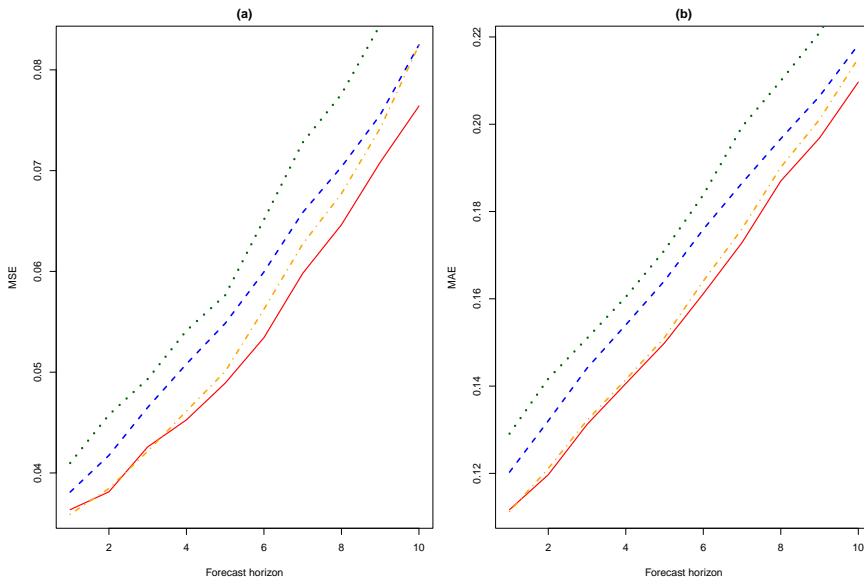
Figure 6: (a) MSE and (b) MAE plots for Danish mortality of the HUts (red solid line), HU (blue dashed line), HUrob (green dotted line), and the wHU (orange dash-dotted line) models
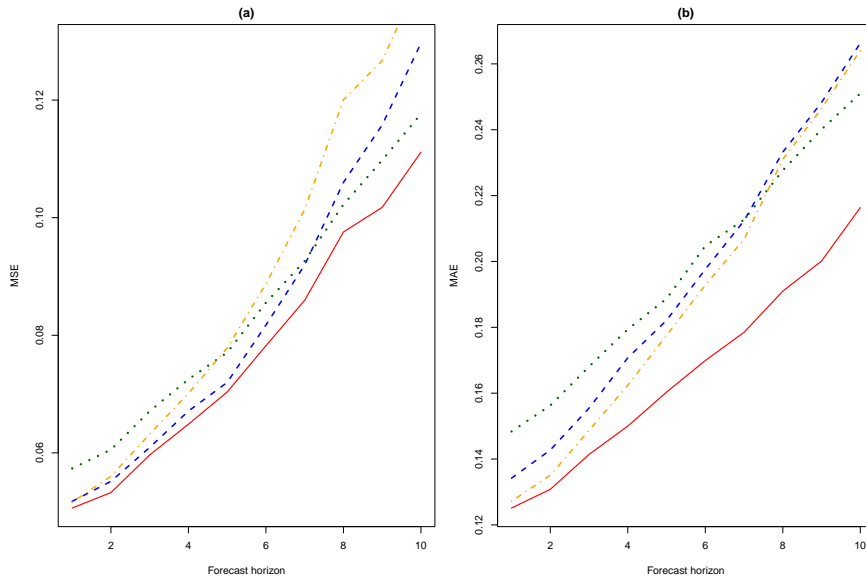
Figure 7: (a) MSE and (b) MAE plots for Irish mortality of the HUts (red solid line), HU (blue dashed line), HUrob (green dotted line), and the wHU (orange dash-dotted line) models
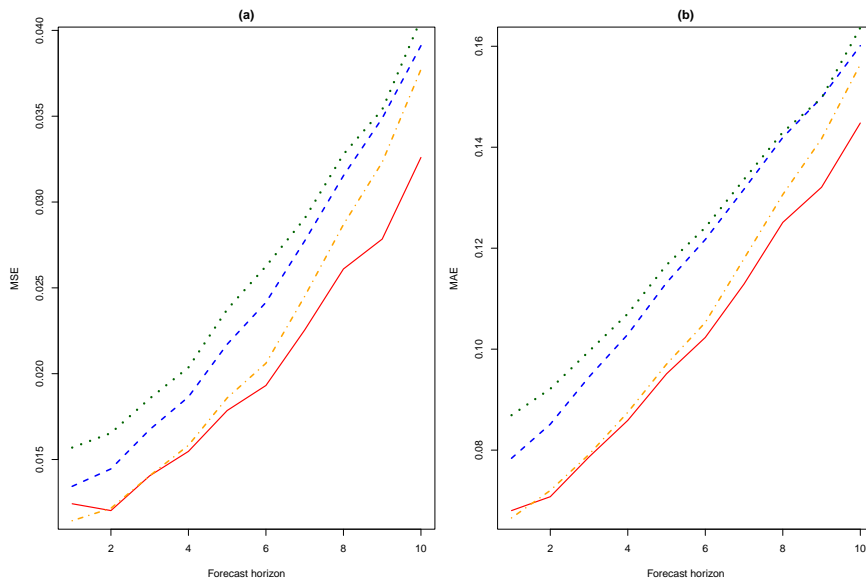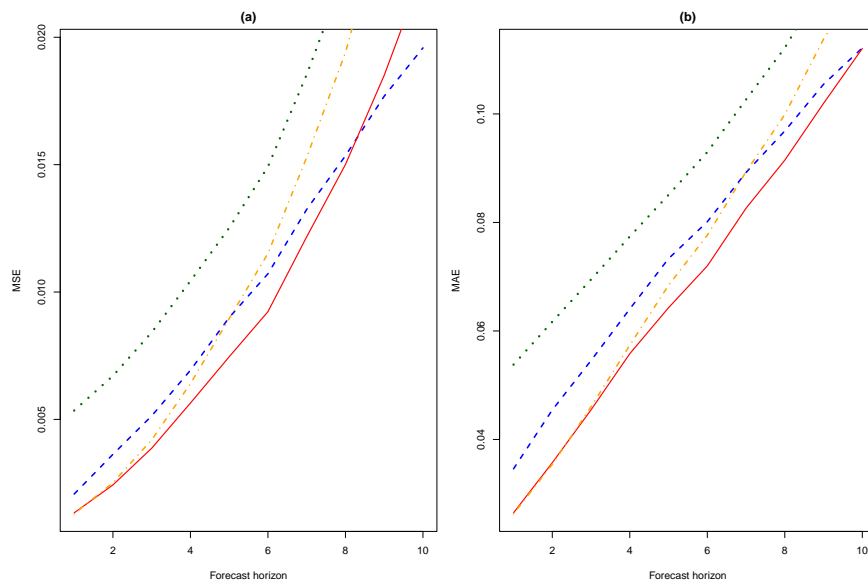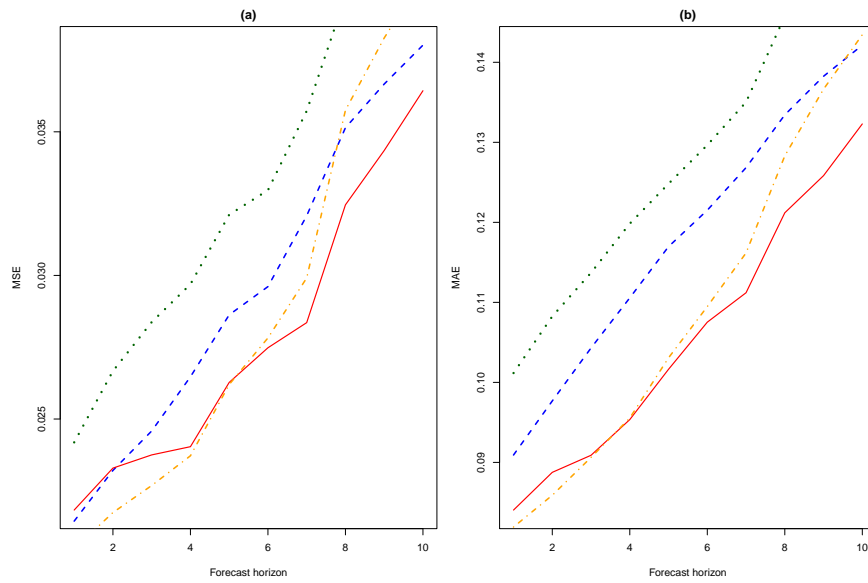

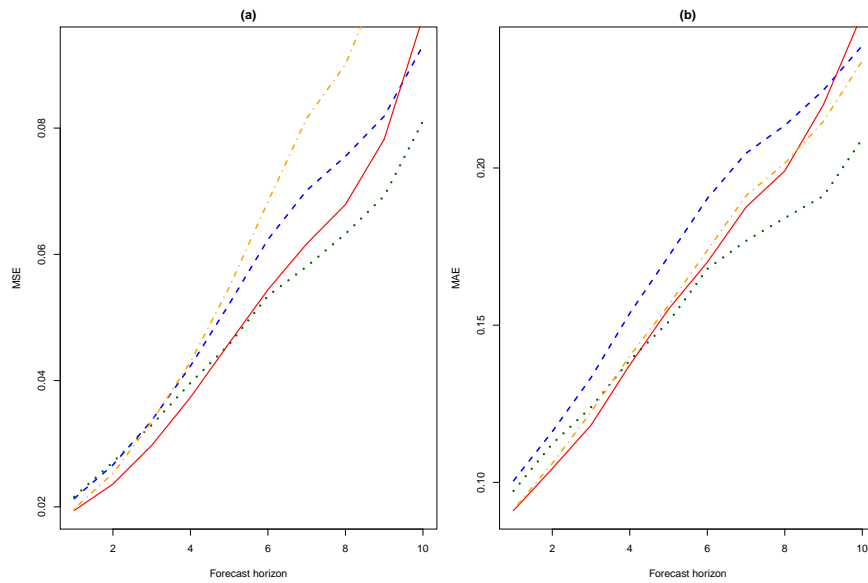
Figure 8: (a) MSE and (b) MAE plots for Dutch mortality of the HUts (red solid line), HU (blue dashed line), HUrob (green dotted line), and the wHU (orange dash-dotted line) models

Figure 9: (a) MSE and (b) MAE plots for American mortality of the HUts (red solid line), HU (blue dashed line), HUrob (green dotted line), and the wHU (orange dash-dotted line) models
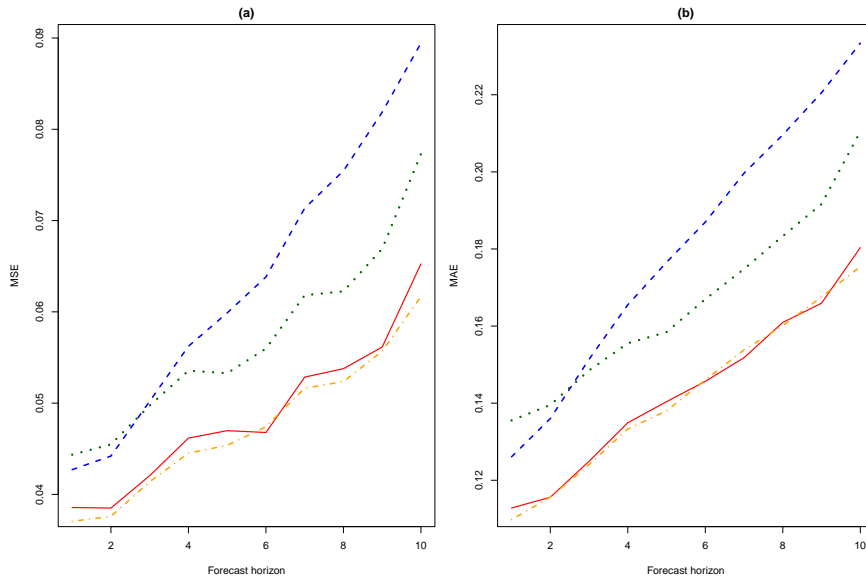
8

Appendix E.2   Comparable to the HUts model

Figure 10: (a) MSE and (b) MAE plots for Belgian mortality of the HUts (red solid line), HU (blue dashed line), HUrob (green dotted line), and the wHU (orange dash-dotted line) models
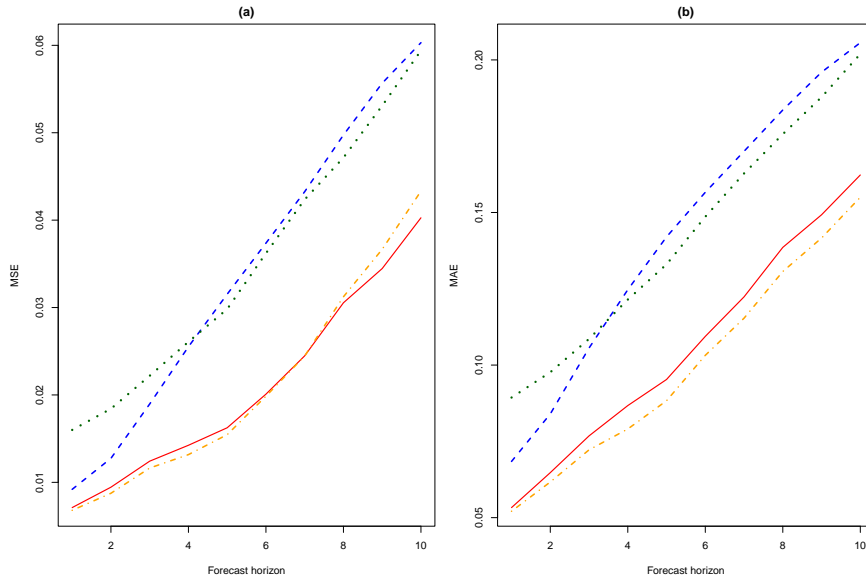

Figure 11: (a) MSE and (b) MAE plots for Bulgarian mortality of the HUts (red solid line), HU (blue dashed line), HUrob (green dotted line), and the wHU (orange dash-dotted line) models

Figure 12: (a) MSE and (b) MAE plots for Finnish mortality of the HUts (red solid line), HU (blue dashed line), HUrob (green dotted line), and the wHU (orange dash-dotted line) models

Figure 13: (a) MSE and (b) MAE plots for Italian mortality of the HUts (red solid line), HU (blue dashed line), HUrob (green dotted line), and the wHU (orange dash-dotted line) models
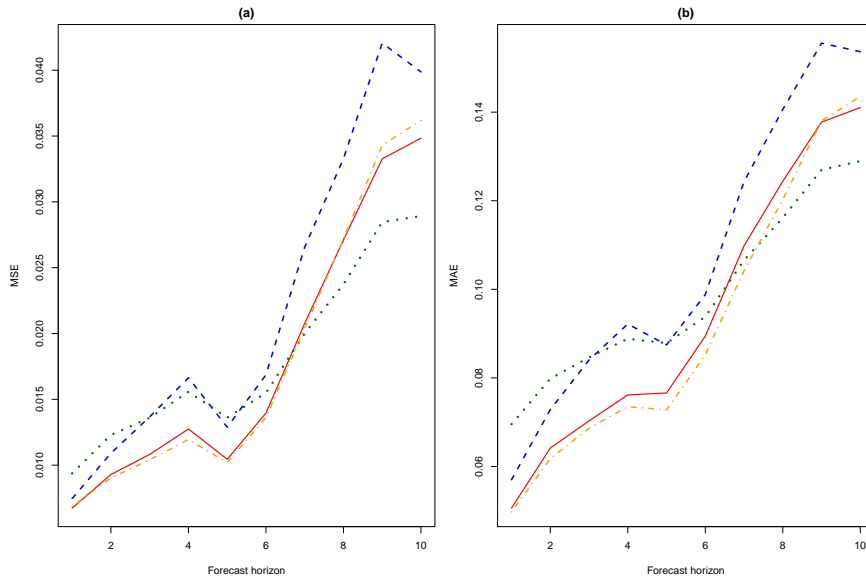
Figure 14: (a) MSE and (b) MAE plots for Japanese mortality of the HUts (red solid line), HU (blue dashed line), HUrob (green dotted line), and the wHU (orange dash-dotted line) models
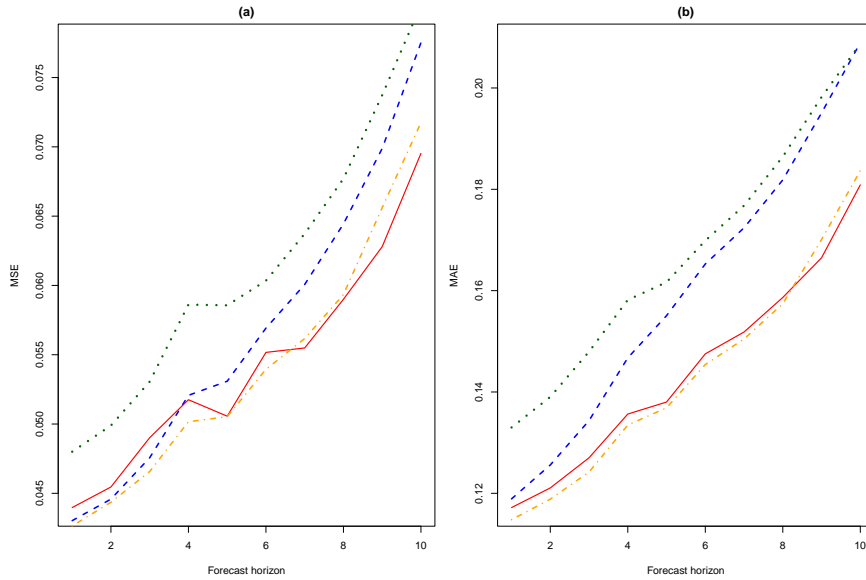

Figure 15: (a) MSE and (b) MAE plots for Norwegian mortality of the HUts (red solid line), HU (blue dashed line), HUrob (green dotted line), and the wHU (orange dash-dotted line) models

Appendix F Mean error plots

The mean error (ME) metric can be used to identify whether the models tend to overpredict or under-predict mortality rates at specific ages. It is computed with:

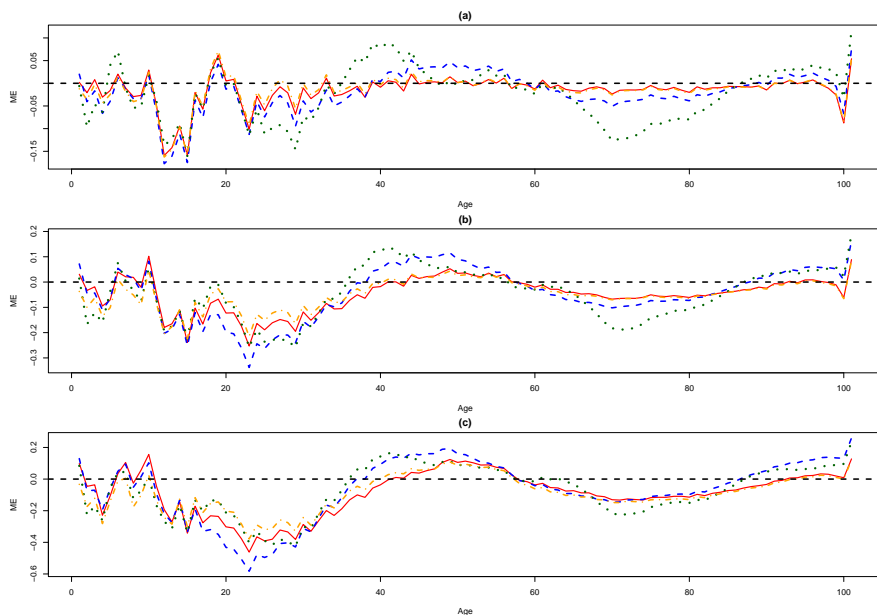$$ME(h, x_i) = \frac{1}{q} \sum_{t=1}^{q} (y_t(x_i) - \hat{y}_{t|t-h}(x_i)).$$

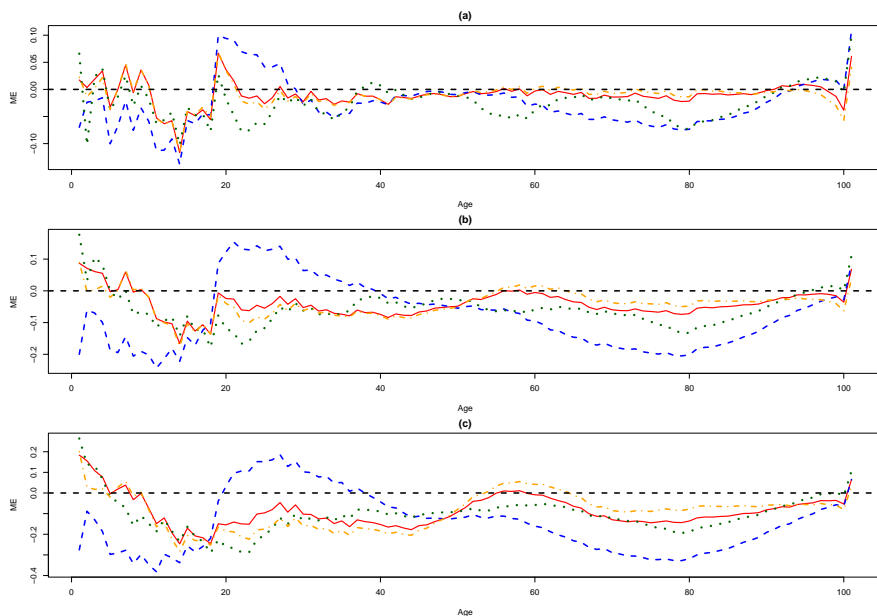Figure 16: Mean error plot of Australian mortality at (a) h=1; (b) h=5; (c) h=10

Figure 17: Mean error plot of French mortality at (a) h=1; (b) h=5; (c) h=10

Upon a closer inspection of the ME plots across different age groups of the French mortality data in Figure 17, the HUts model generally demonstrates minimal bias across ages. For one-step-ahead forecasts (Figure 17 (a)), the ME plot indicates that the HUts model's forecasts are mostly unbiased, with slight

dips around age 15 and peaks near ages 20 and 100. The wHU model shows a similar pattern, while the HU and HUrob models generally underpredict mortality across ages.

For five-step-ahead forecasts (Figure 17 (b)), the ME plot for the HUts model remains close to zero, indicating sustained minimal bias. The wHU model continues to exhibit similar trends, whereas the HU and HUrob models persist in underpredicting mortality. In the ten-step-ahead forecasts (Figure 17 (c)), the HUts model's ME plot shows a slight bias at extreme ages but maintains near-zero mean error across most ages, suggesting accurate and unbiased forecasts overall. The HU and HUrob models display larger biases, contributing to less reliable predictions.

Overall, these observations suggest that the HUts model produces mostly unbiased forecasts, especially as the forecast horizon increases. While the wHU model also performs well at shorter horizons, the HUts model becomes more favorable for longer-term predictions due to its minimal bias across ages.

Similarly, the ME plots for the Australian mortality data in Figure 16 highlight the effectiveness of the HUts model, particularly in short-term forecasting. For one-step-ahead forecasts (Figure 16 (a)), the ME plot reveals minimal bias across ages for the HUts model, mirroring the performance of the wHU model. In contrast, the HU and HUrob models exhibit a tendency to underpredict mortality.

At the five-step-ahead horizon (Figure 16 (b)), the ME plot for the HUts model indicates underprediction biases in the 20-40 age range, contributing to increased errors in this group. By the ten-step-ahead forecasts (Figure 16 (c)), this underprediction becomes more pronounced in the same age range, significantly affecting the model's overall accuracy. Despite this, the HUts model maintains reasonable performance across other age groups, though the wHU model generally achieves better results at longer horizons.

Appendix G   The Hyndman-Ullah with randomised signatures model

Randomised signatures are a computationally efficient alternative to the truncated signatures from rough path theory. Randomized signatures projects the high-dimensional signature space into a lower-dimensional space using random linear projections. These projections approximate the original signature's properties and expressiveness, potentially reducing the computational need. This is achieved by evolving a system of random differential equations driven by the input path, resulting in a feature representation that maintains the theoretical guarantees of the signature while being computationally tractable for downstream tasks such as classification or system identification. See Compagnoni et al. (2023), Cuchiero and Möller (2024), and Cuchiero et al. (2020, 2024) for a more detailed introduction on randomised signatures. The following is an algorithm to generate randomised signatures:

Algorithm 1 Generate randomized signature (Compagnoni et al., 2023)

Require: $X \in \mathbb{R}^d$ sampled at $0 = t_0 < t_1 < \cdots < t_N = T$, randomised signature dimension k, activation function σ.

1: Initialize $Z_{t_0} \in \mathbb{R}^k$, $A_i \in \mathbb{R}^{k \times k}$, $b_i \in \mathbb{R}^k$ to have i.i.d. standard normal entries for $i \in \{1, \ldots, d\}$.

2: **for** $n = 1$ to N **do**

3:    Compute:

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

4: **end for**

By replacing the usage of truncated signatures with randomised signature, we can model the morality rates using the Hyndman-Ullah with randomised signatures (HUrs) model. We set the number of signature dimension $k = 75$, and use the linear activation function. The MSE results are tabulated below.

As observed in Table 7, only Bulgaria appears to have experienced an improvement in forecasting accuracy. Nevertheless, the randomized signature requires numerous variables that require predetermination, necessitating further investigation to determine the optimal set of hyperparameters.

Table 7: MSE: HUts, HUrs, and wHU across different countries for $h = 1, 5, 10$

Country	h=1			h=5			h=10		
	HUts	HUrs	wHU	HUts	HUrs	wHU	HUts	HUrs	wHU
Australia	0.00982	0.01062	0.00952	0.01835	0.03038	0.01642	0.0418	0.08545	0.0351
Belgium	0.02183	0.02371	0.02047	0.02627	0.04172	0.02620	0.0364	0.09131	0.0404
Bulgaria	0.01942	0.01930	0.01968	0.04595	0.04442	0.05478	0.0976	0.08949	0.1280
Denmark	0.03636	0.03914	0.03590	0.04891	0.07024	0.05007	0.0764	0.14972	0.0825
Finland	0.03856	0.04569	0.03705	0.04698	0.07012	0.04538	0.0653	0.12988	0.0617
France	0.00398	0.00615	0.00399	0.00882	0.02614	0.00978	0.0201	0.08445	0.0242
Ireland	0.05062	0.05275	0.05161	0.07035	0.09055	0.07785	0.1111	0.17584	0.1413
Italy	0.00711	0.00995	0.00680	0.01625	0.03723	0.01546	0.0403	0.11560	0.0433
Japan	0.00674	0.00680	0.00687	0.01045	0.01000	0.01020	0.0348	0.03289	0.0362
Netherlands	0.01242	0.01435	0.01143	0.01785	0.03353	0.01859	0.0326	0.09922	0.0377
Norway	0.04397	0.04692	0.04264	0.05056	0.06853	0.05053	0.0695	0.13147	0.0718
United States	0.00133	0.00162	0.00129	0.00746	0.01023	0.00896	0.0226	0.03197	0.0305
Mean	0.02101	0.02308	0.02061	0.03068	0.04443	0.03202	0.0540	0.1014	0.0611

References

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

Cuchiero, C., Gazzani, G., Möller, J., & 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., Gonon, L., Grigoryeva, L., Ortega, J.-P., & 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., & Möller, J. (2024). Signature methods in stochastic portfolio theory. https://doi.org/10.48550/arXiv.2310.02322