Dynamic Mortality Forecasting Using Mixed-Frequency State-Space Models

Runze Li Rui Zhou David Pitt

University of Melbourne

Longevity 20 Conference, Singapore
September 8-9, 2025

Outline

Leveraging Mixed-Frequency Data

Annual mortality rates

- Stable, but only available with long delays

Monthly death counts

- Timely and informative, but noisy if used on their own

Advantages of a Mixed-Frequency Approach

- Capture short-term monthly dynamics while staying consistent with long-term annual trends
- Produce early forecasts that still provide reliable signals of annual mortality rates

State-Space Models for Annual Mortality

- Pedroza (2006), Fung, Peters, and Shevchenko (2015): state-space Lee-Carter models
- Andersson and Lindholm (2021): continuous-time state-space model

Advantage:

Well-established framework supporting efficient estimation

Limitation:

Only suitable for single-frequency mortality data

High-Frequency Data in Existing Mortality studies

- Recent studies on climate and pandemic effects highlight the value of high-frequency mortality data in capturing short-term mortality fluctuations
  - Robben, Antonio, and Kleinow (2025) use weekly death data from Eurostat to examine the relationship between environmental factors and weekly mortality in European regions.
  - With French weekly mortality data, Robben, Barigou, and Kleinow (2025) develop a regime-switching framework which explains short-term mortality fluctuations caused by temperature and epidemic shocks.

Current Study on Mixed-Frequency Mortality

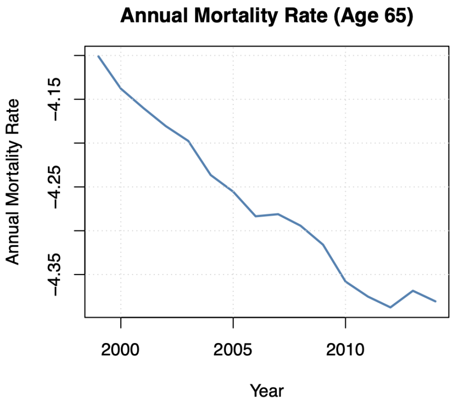
- Li, Zhou, and Pitt (2025): Develop Mixed Data Sampling (MIDAS) models for mortality forecasting
- Annual mortality rates enter the annual autoregressive component; monthly death counts enter the MIDAS component
- Forecasts of annual mortality rates are produced each month before year-end
- **Limitations:**
  - New month $\Rightarrow$ refit MIDAS: As the year advances (Jan$\rightarrow$Feb$\rightarrow$...), monthly inputs shift; re-estimate for the same target year.
  - New forecast horizon $\Rightarrow$ refit MIDAS: 1-year-ahead vs. 2-year-ahead, etc., require separate fits.
  - No monthly mortality projections

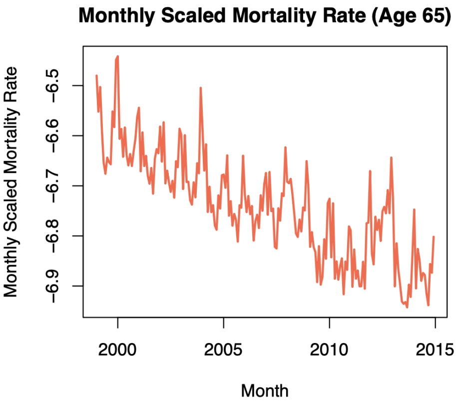
Objectives for Enhanced Mortality Modeling

1. **Utilize intra-year information:** Develop a mixed-frequency state-space model that combines annual and monthly data
2. **Enable real-time forecasting:** Forecast dynamically as new monthly data become available
3. **Improve algorithm parsimony:** Reduce the number of reparameterized models needed for forecasting
4. **Deliver multi-frequency forecasts:** Provide mortality projections at both annual and monthly levels

Data Overview

- U.S. mortality data from HMD and CDC
- Fitting period: January, 1999 – December, 2014
- Age range: 20 – 90





Lee–Carter as a State-Space Model

- State-Space models have two layers, the observation equation and the state equation.
- Example: Lee–Carter Model for annual mortality rates
  1. **Observation equation:**

  $$\ln m_{x,t} = \alpha_x + \beta_x k_t + \varepsilon_{x,t}$$

  links observed mortality rates $m_{x,t}$ to latent factors (states) k_t

  2. **State equation:**

  $$k_t = k_{t-1} + d + \eta_t$$

  describes how k_t evolves over time, i.e., with a random walk with drift

Extension to Mixed-Frequency Framework

Mixed-Frequency State-Space Models (MF-SSM) have been applied successfully in economics:

- GDP modelling and forecasting (Mariano and Murasawa (2003))

Our MF-SSM for mortality modelling:

1. **Common monthly latent process**: represents the true underlying mortality dynamics shared by both monthly and annual data
2. **Two observation equations:**
   - **Monthly equation:** captures month-to-month dynamics $\Rightarrow$ links monthly mortality directly to the latent factors
   - **Annual equation:** captures long-term trends $\Rightarrow$ links annual mortality to the aggregate of monthly latent factors, smoothing short-term fluctuations

The Notations

- $P_{x,t}$: initial population size at the start of year t
- $d_{x,t+\frac{h}{12}}$: monthly death count in month h of year t
- $\mathbf{z}_{x,t+\frac{h}{12}}$: scaled monthly death rate in month h of year t

$$\mathbf{z}_{x,t+\frac{h}{12}} = \frac{d_{x,t+\frac{h}{12}}}{P_{x,t}}$$

- $m_{x,t}$: annual mortality rate, available at the end of year t
- $k_{t+\frac{h}{12}}$: monthly latent mortality index at month h of year t

Equations of MF-SSM

$$\text{Annual equation:} \quad \ln m_{x,t} = a_{1,x} + b_{1,x} \frac{\sum_{i=1}^{12} k_{t+\frac{i}{12}}}{12} + \epsilon_{x,t}$$

$$\text{Monthly equation:} \quad \ln z_{x,t+\frac{h}{12}} = a_{2,x} + b_{2,x}\, k_{t+\frac{h}{12}} + \eta_{x,t+\frac{h}{12}}$$

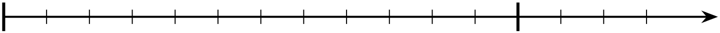
$$\text{State equation:} \quad k_{x,t+\frac{h}{12}} \sim \text{SARIMA}$$

- Projection at two frequencies $\Rightarrow$ produce forecasts of annual mortality rates from
  1. the annual equation
  2. the monthly equation with aggregation
- $k_{t+\frac{h}{12}}$ drives both process $\Rightarrow$ Question: Which frequency of data has greater influence on $k_{t+\frac{h}{12}}$?

Model Estimation

Estimation method: Kalman filter smoothing with EM

Annual data	$m_{x,t}$	NA	$\cdots$	NA	$\cdots$	$m_{x,t+1}$	NA	$\cdots$
Monthly data	$z_{x,t}$	$z_{x,t+\frac{1}{12}}$	$\cdots$	$z_{x,t+\frac{h}{12}}$	$\cdots$	$z_{x,t+\frac{12}{12}}$		$\cdots$



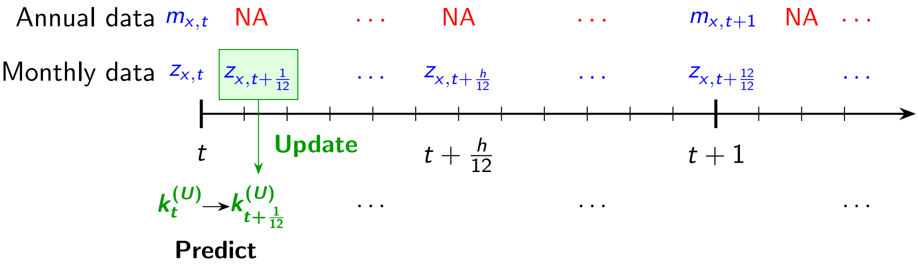
t $\quad t+\frac{h}{12}$ $\quad t+1$

$\boldsymbol{k_t^{(U)}} \rightarrow k_{t+\frac{1}{12}}^{(P)}$ $\cdots$ $\cdots$ $\cdots$

Predict

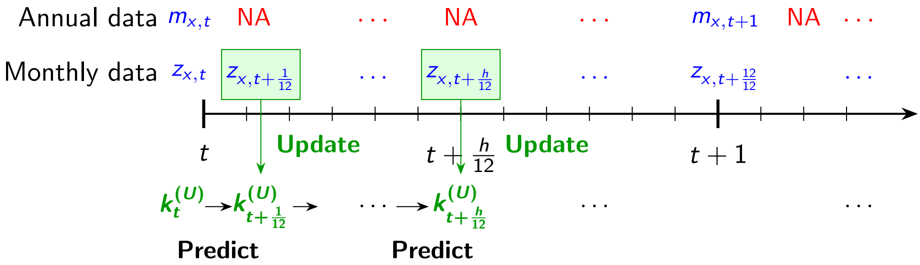
Model Estimation

Estimation method: Kalman filter smoothing with EM



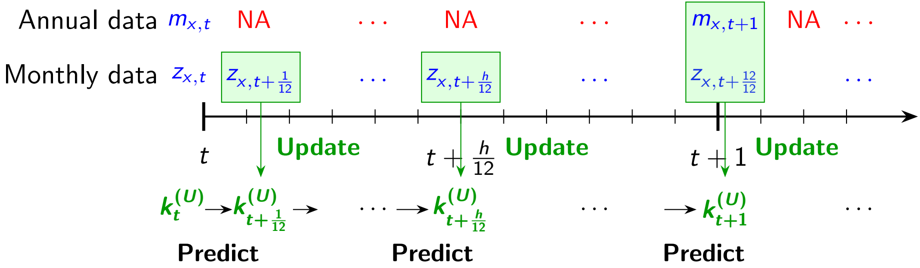
Model Estimation

Estimation method: Kalman filter smoothing with EM



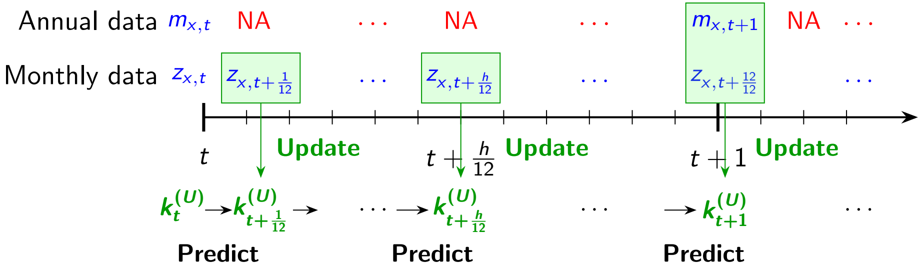
Model Estimation

Estimation method: Kalman filter smoothing with EM



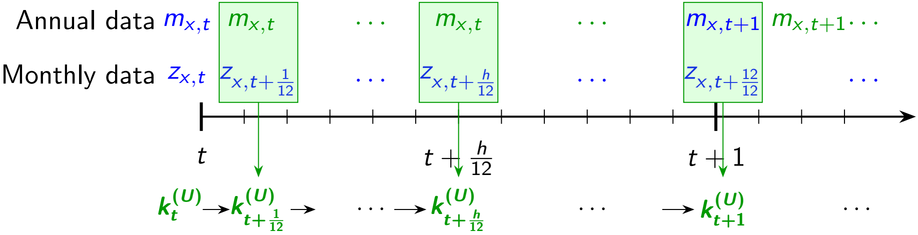
Model Estimation

Estimation method: Kalman filter smoothing with EM



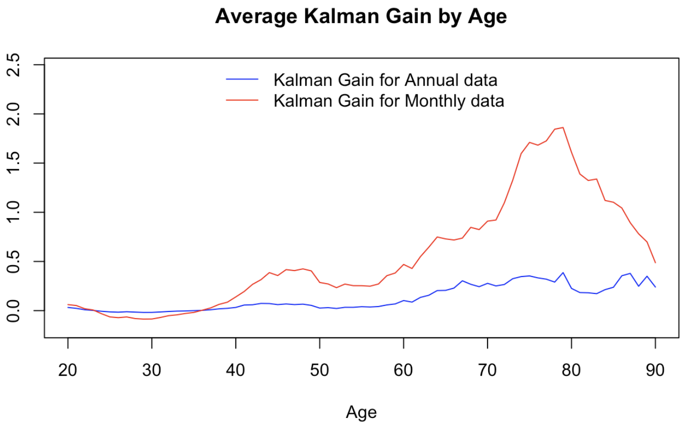
- Months 1–11: updates of $k_{t+\frac{h}{12}}$ solely affected by monthly data
- Month 12: annual data causes a sudden correction
- There is larger jump in the evolution of $k_{t+\frac{h}{12}}$ at every year-end

Fill the missing values



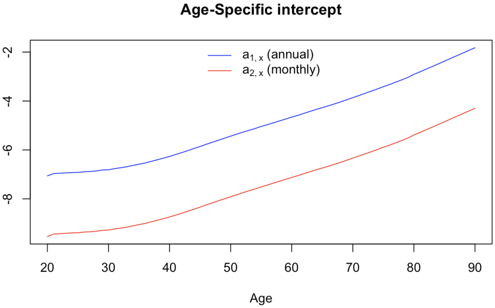
- Missing annual values are filled using the last observed level
- Acts as an anchor for the monthly latent factors $k_{t+\frac{h}{12}}$

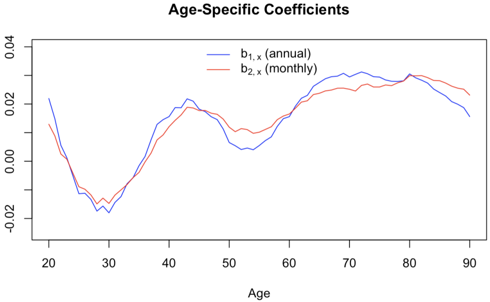
Empirical results: Fitted Kalman Gain



- **Trade-off between model and data:** Kalman Gain is the contribution weight of observed data in the evolution of monthly mortality index
- **Monthly vs Annual:** greater impact from monthly data
- **Old vs Young:** greater contribution from mortality patterns of old ages

Annual vs Monthly Age Parameters





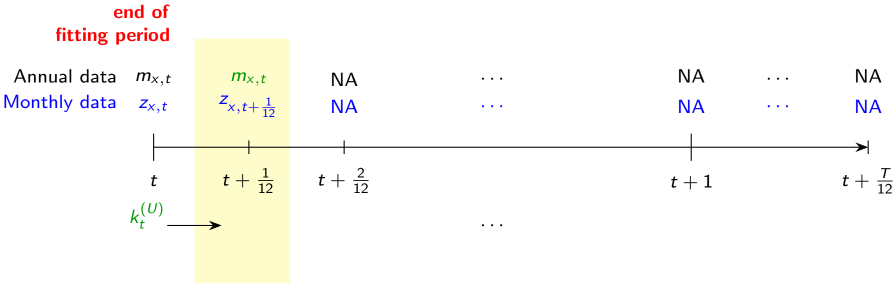
- $a_{2,x}$ **is smaller than** $a_{1,x} \Rightarrow$ it represents monthly mean crude death rate
- Overall, most ages exhibit similar sensitivity towards annual and monthly mortality trends
- Both (a_1, b_1) and (a_2, b_2) are necessary since 12 monthly inputs do not sum to the annual input $\sum_{i=0}^{11} \ln z_{x,t-\frac{i}{12}} \neq \ln m_{x,t}$

Forecasting at Month 1

At Month 1 of year $t + 1$

1. Predict next state

From Month 2 onwards

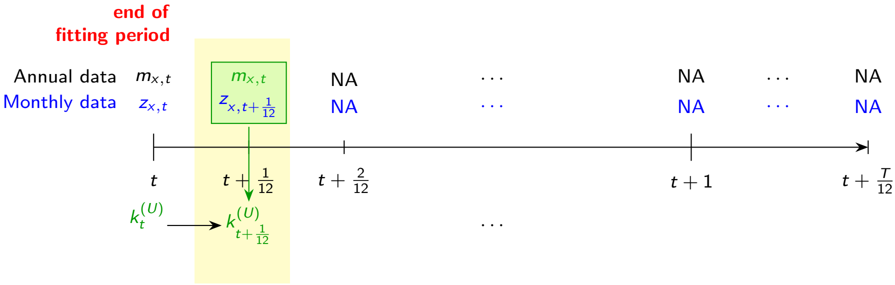


Forecasting at Month 1

At Month 1 of year $t+1$

1. Predict next state
2. Update using new monthly data

From Month 2 onwards



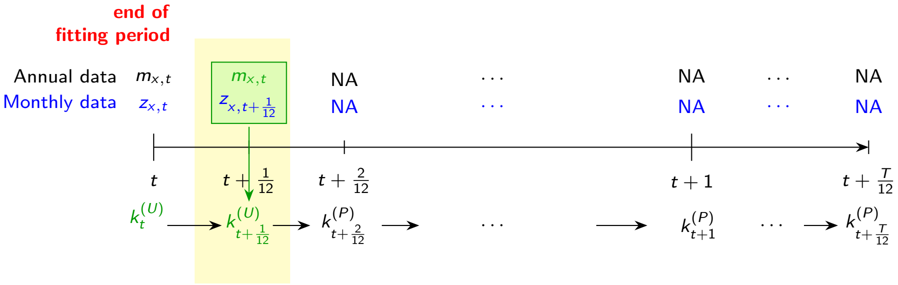
Forecasting at Month 1

At Month 1 of year $t+1$

1. Predict next state
2. Update using new monthly data

From Month 2 onwards

1. Predict $k^{(P)}_{t+\frac{i}{12}}$ for $i \in \{2, \dots, T\}$, starting from $k^{(U)}_{t+\frac{1}{12}}$

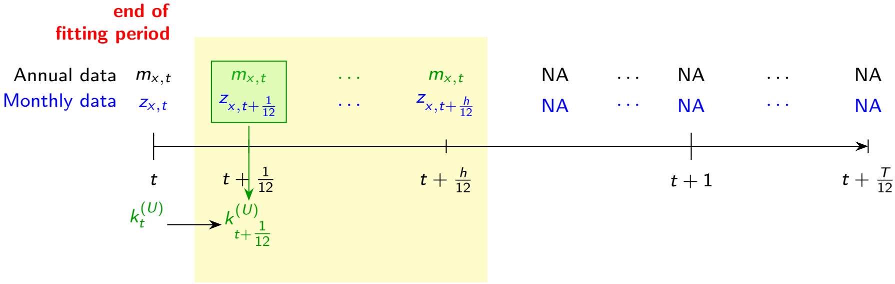


Forecasting at Month h

For Month 1-h of year $t+1$

1. Perform predict-update for $k^{(P)}_{t+\frac{1}{12}}$

For Month $h+1$ onwards

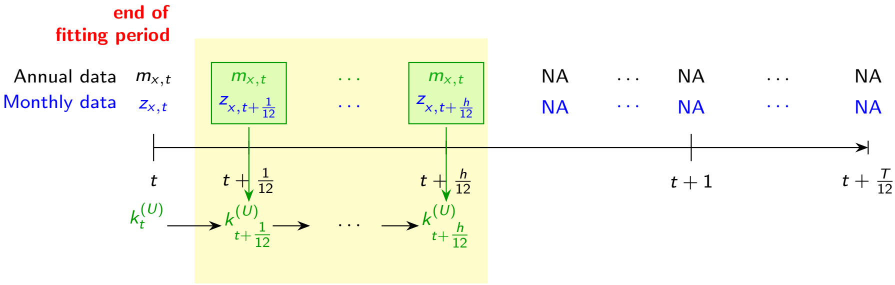


Forecasting at Month h

For Month 1-h of year $t+1$

1. Perform predict-update for $k^{(P)}_{t+\frac{1}{12}}$
2. Iterate until current month h

For Month $h+1$ onwards



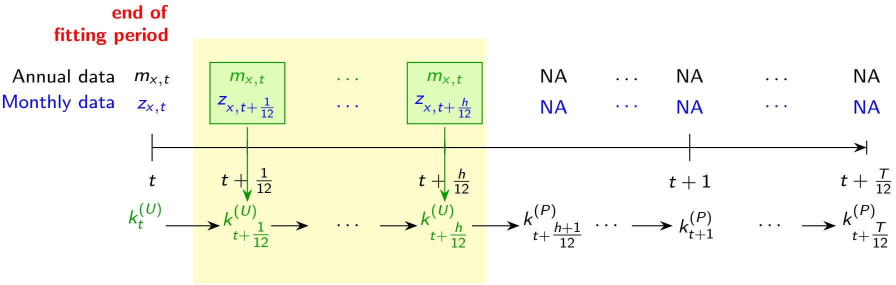
Forecasting at Month h

For Month 1-h of year $t+1$

1. Perform predict-update for $k^{(P)}_{t+\frac{1}{12}}$
2. Iterate until current month h

For Month $h+1$ onwards

1. Predict $k^{(P)}_{t+\frac{i}{12}}$ for $i \in \{h+1, ..., T\}$, starting from $k^{(P)}_{t+\frac{h}{12}}$



Design for Forecast Evaluation

Benchmark models

- Annual Lee-Carter model: $\ln m_{x,t} = a_x + b_x k_t + \epsilon_{x,t}$
  - k_t follows a random walk with drift
  - Forecasting for year $t+n$: forecast $\hat{k}_{t+1}, \ldots, \hat{k}_{t+n}$ and calculate $m_{x,t+n}$
- Monthly Lee-Carter model: $\ln z_{x,t+\frac{h}{12}} = a_x + b_x k_{t+\frac{h}{12}} + \epsilon_{x,t+\frac{h}{12}}$
  - $k_{t+\frac{h}{12}}$ follows the same SARIMA process as in the state-space model
  - Forecasting for year $t+n$: update $k_{t+\frac{h}{12}}$ for months with new information by OLS estimation; forecast $k_{t+\frac{h}{12}}$ by SARIMA for future months
  - Consistent with the updating process in state-space model

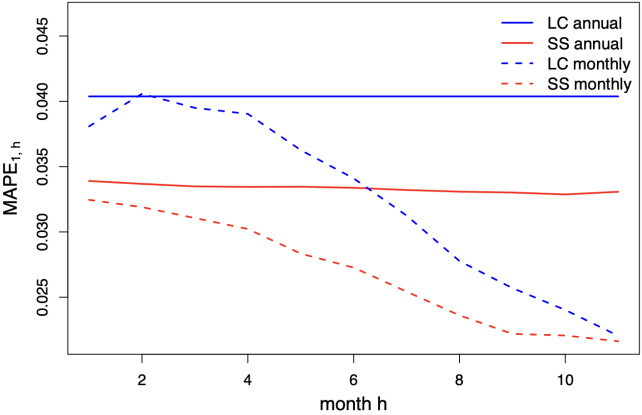
Forecast results are produced by models fitted on 1999-2014

1-year-ahead forecast for Year 2015

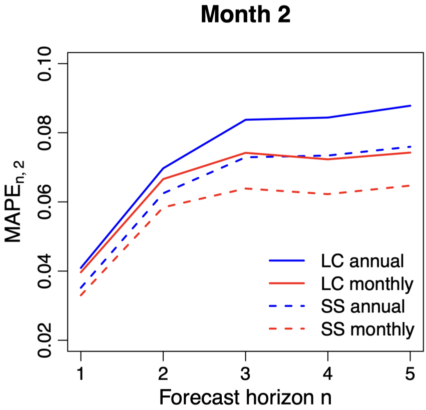
$MAPE_{n,h}$: mean average percentage errors for n-year-ahead forecast at month h, averaged across ages

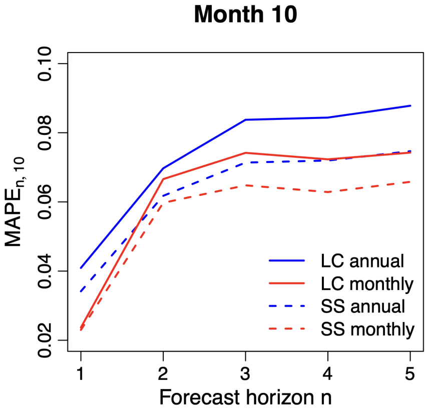
LC: Lee-Carter model

SS: State-Space model



1 to 5-year-ahead forecasts





- Higher forecast accuracy by monthly models: ability to capture intra-year mortality movement
- Better performance by state-space models: incorporation of both monthly and annual information

Conclusion

This study develops a mixed-frequency state-space framework for mortality modeling and forecasting:

- Integrating annual mortality rates and monthly death counts within a unified state-space framework
- Achieving real-time forecasts by incorporating newly available monthly data
- Improving forecast accuracy compared with single-frequency Lee-Carter models

Reference

Andersson, Patrik and Mathias Lindholm (2021). "Mortality forecasting using a Lexis-based state-space model". In: *Annals of Actuarial Science* 15.3, pp. 519–548. DOI: 10.1017/S1748499520000275.

Fung, Simon Man Chung, Gareth Peters, and Pavel Shevchenko (Aug. 2015). "A State-Space Estimation of the Lee-Carter Mortality Model and Implications for Annuity Pricing". In: *SSRN Electronic Journal.*

Li, Runze, Rui Zhou, and David Pitt (2025). "Beyond Annual Data: Mortality Forecasting with Mixed Frequency Data". In: *Department of Economics, University of Melbourne Working Paper*. Preprint.

Mariano, Roberto S. and Yasutomo Murasawa (2003). "A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series". In: *Journal of Applied Econometrics* 18.4, pp. 427–443. (Visited on 08/23/2025).

Pedroza, Claudia (Feb. 2006). "A Bayesian forecasting model: predicting U.S. male mortality". In: *Biostatistics* 7.4, pp. 530–550. ISSN: 1465-4644.

Robben, Jens, Katrien Antonio, and Torsten Kleinow (May 2025). "The