

Beyond Annual Data: Mortality Forecasting with Mixed Frequency Data

Runze Li¹ Rui Zhou¹ David Pitt¹

¹ University of Melbourne

Longevity 19 Conference
Amsterdam, September 16-17, 2024

Outline

- 1 Introduction
 - Motivation and our objectives
 - MIDAS: applications in economic analysis
 - Data overview
- 2 Mortality rates and MIDAS models
 - AR-MIDAS nowcast model
 - AR-MIDAS forecast model
 - Extension to VAR-MIDAS
- 3 Empirical Results
 - Fitting the MIDAS models
 - Nowcasting and forecasting with MIDAS models
- 4 Conclusion

Motivation

- **Aim:** Forecast annual mortality rates using more granular data
- **Data availability:** Monthly and weekly death records are available.
- **Data usefulness:** More deaths occur in winter, with winter mortality dominating year-to-year mortality variation, providing insights for updating annual mortality forecasts.
- **The research gap:** Higher frequency mortality data is rarely used in mortality forecasting as existing models do not accommodate mixed frequency data records.

High-frequency data in existing mortality studies

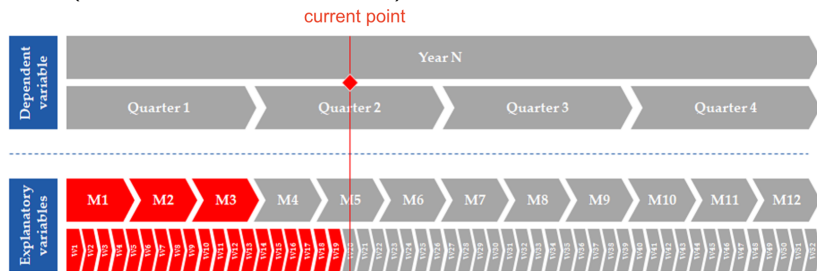
- Recent life table studies have demonstrated the added value of using higher frequency data in refining mortality rate estimates.
 - Boumezoued (2021), who focuses on period-cohort interaction, incorporates monthly birth counts to revise exposure to risk figures.
 - Pavía and Lledó (2021) study the seasonal-aging effect on mortality, using demographic microdata to construct sub-annual life table and derive life table indexes.
- We explore the use of high-frequency death counts in mortality forecasting.

Our objectives

- We propose a new approach that utilises most recent monthly (or weekly) death data to provide timely updates of annual mortality forecasts—No need to wait until year end.
- We assess the effectiveness of utilising high frequency death data in both short-term (current year) nowcasts and medium-term (next year and beyond) forecasts of annual mortality rates.

MIDAS technique

- Mixed data sampling (MIDAS) is a technique used to directly relate variables observed at different frequencies without aggregating the higher-frequency data.
- Example: Forecast quarterly GDP using monthly and weekly market indices (Jardet and Meunier, 2020)

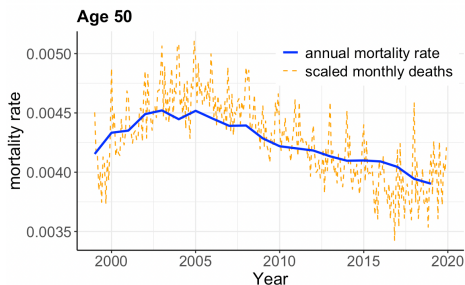
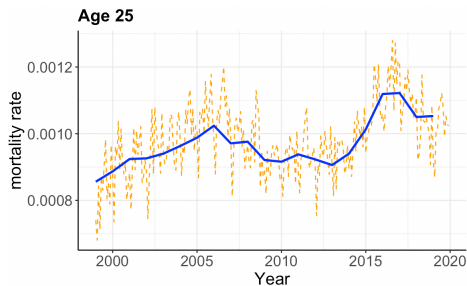


- Nowcast:** predict Q2 (current quarter) GDP using weekly and monthly predictors from M1-M3 and W1-W19.
- 1-step ahead forecast:** predict Q3 (next quarter) GDP using higher-frequency predictors from M1-M3 and W1-W19.

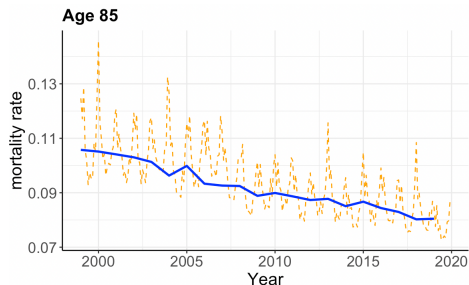
Applying MIDAS to mortality forecasting

- Similar to MIDAS forecasting in economics
 - Higher frequency data: monthly death counts
 - Lower frequency data: annual mortality rates
- Mortality data varies along two dimensions: age and time.
- Our study uses U.S. mortality data to illustrate the extension of the MIDAS technique to mortality modelling and forecasting.

U.S. mortality data

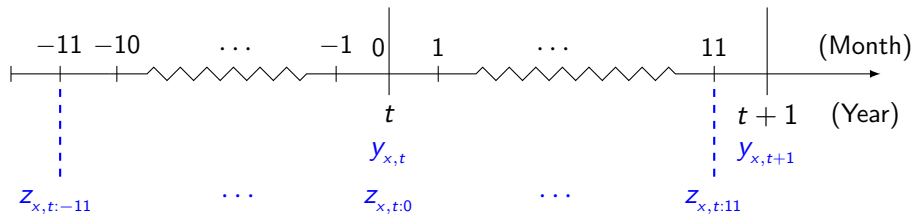


- Annual mortality rate from HMD, 1999-2021
- Monthly death counts from CDC WONDER, 1999-2019.
 - Co-movement of annual and monthly data
 - Seasonality in monthly data, especially for higher ages



Some notation

- Annual mortality improvement, $y_{x,t}$
 - log annual mortality rate at time t (rate for the year ending at t), differenced at yearly lag 1.
- Monthly excess death, $z_{x,t:k}$
 - log monthly death counts at time $t + \frac{k}{12}$ (count for the month ending at $t + \frac{k}{12}$), differenced at monthly lag 12.
 - Note that $t : k = t + \frac{k}{12}$. Positive k indicates $|k|$ months after time t , negative k indicates $|k|$ months before time t .



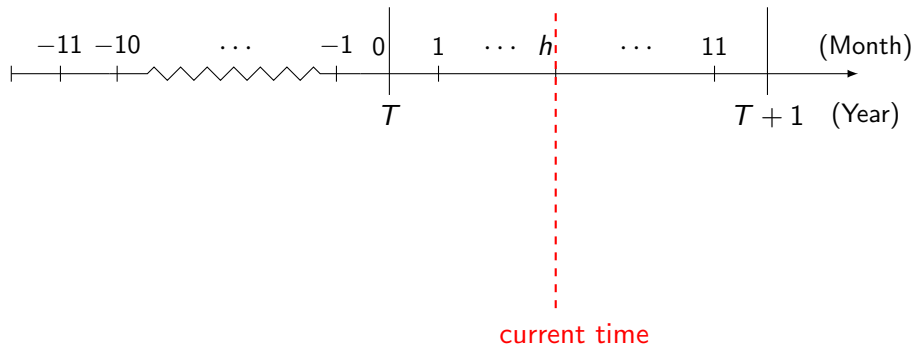
From AR to AR-MIDAS

AR: $y_{x,t+1} = y_{x,t} + \text{Error}$

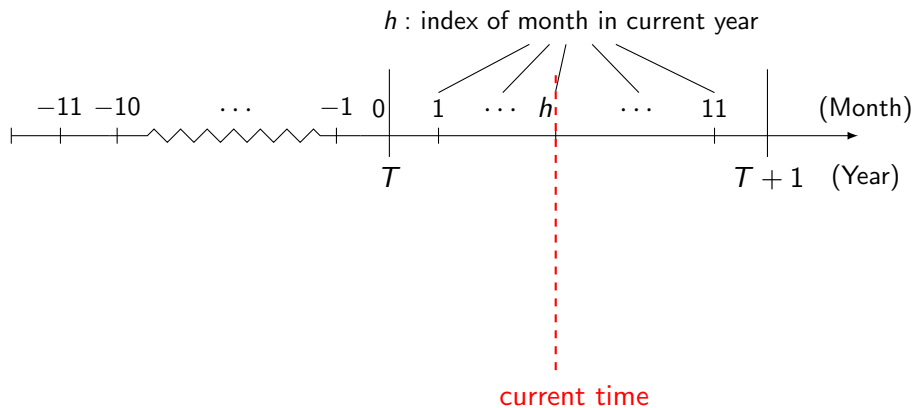
AR-MIDAS: $y_{x,t+1} = y_{x,t} + z_{x,t:k} + \text{Error}$

- AR: the impact of annual rate from previous year only
- AR-MIDAS: also includes the impact of current and previous years' monthly deaths

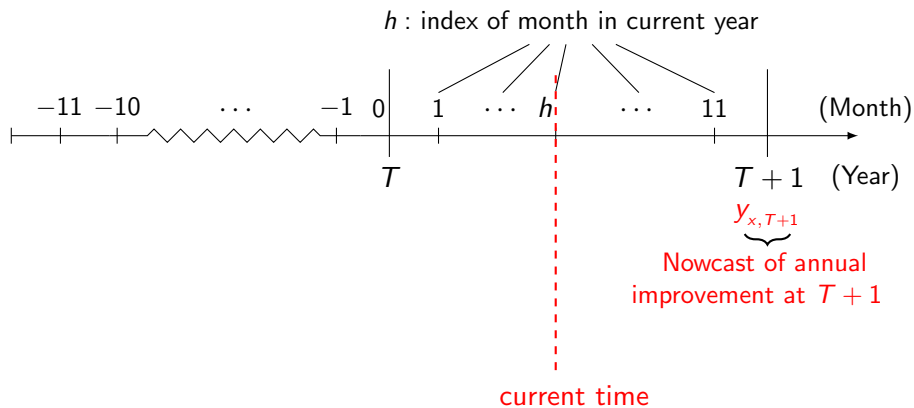
AR-MIDAS: Single age (group) nowcast model



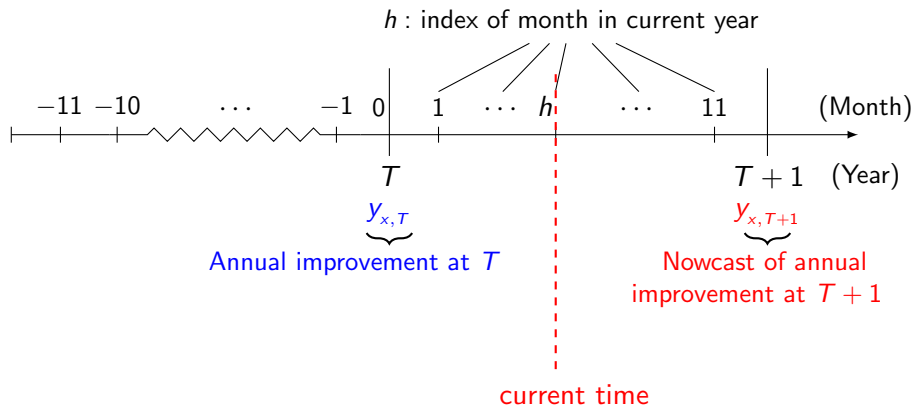
AR-MIDAS: Single age (group) nowcast model



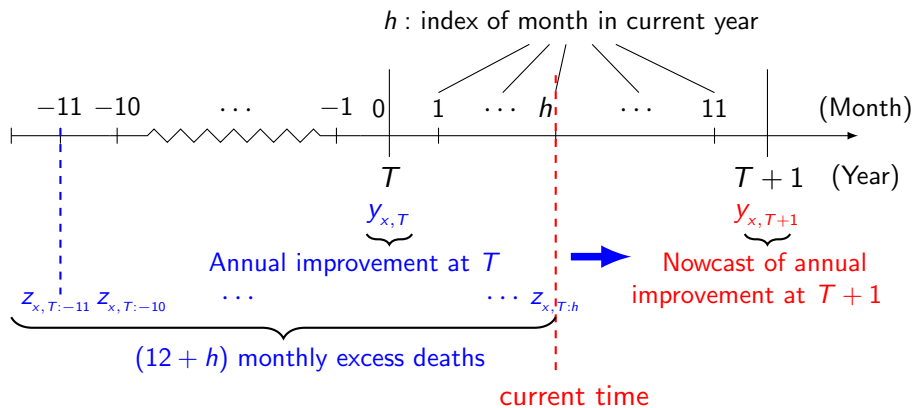
AR-MIDAS: Single age (group) nowcast model



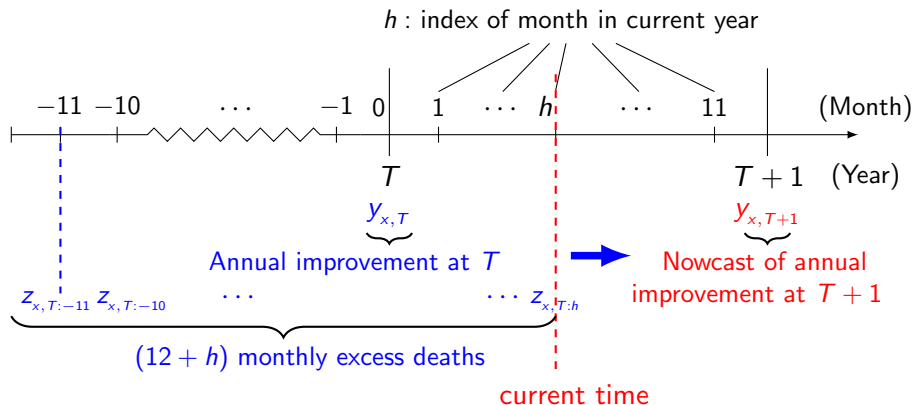
AR-MIDAS: Single age (group) nowcast model



AR-MIDAS: Single age (group) nowcast model



AR-MIDAS: Single age (group) nowcast model



$$y_{x,h,T+1} = \underbrace{\mu_x + \alpha y_{x,T}}_{\text{AR(1)}} + \underbrace{\mu_h + \alpha_h \sum_{k=-11}^h w_h(k; \lambda) z_{x,T:k}}_{\text{MIDAS}} + \text{Error}$$

- **AR(1)**: uses annual mortality rate data

- **MIDAS**: uses monthly death count information

AR-MIDAS: weight function $w_h(k; \lambda)$

$$y_{x,h,T+1} = \cdots + \alpha_h \sum_{k=-11}^h w_h(k; \lambda) z_{x,T:k} + \cdots$$

Exponential Almon Lag weight

(with two shape parameters, λ_1, λ_2)

$$w(k; \lambda) = \frac{\exp(\lambda_1(h+1-k) + \lambda_2(h+1-k)^2)}{\sum_{s=-11}^h \exp(\lambda_1(h+1-s) + \lambda_2(h+1-s)^2)}$$

- Sum to unity
- Parsimony
- Flexibility

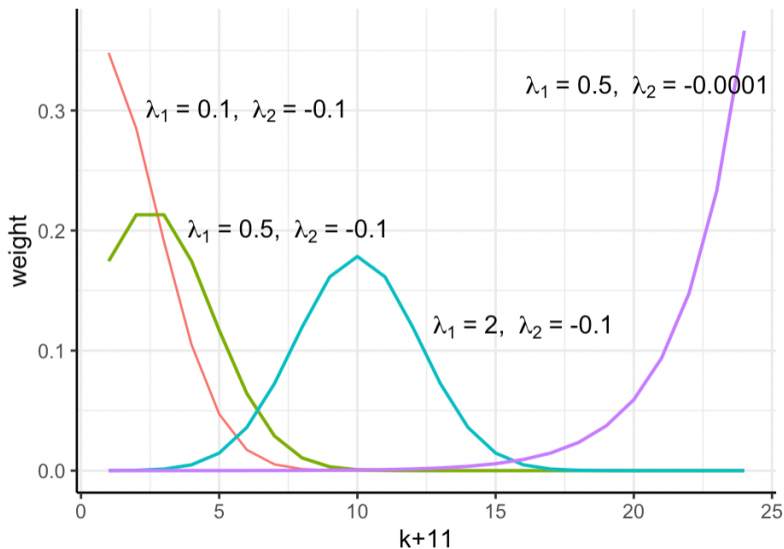
Equal weight

($\lambda_1 = \lambda_2 = 0$)

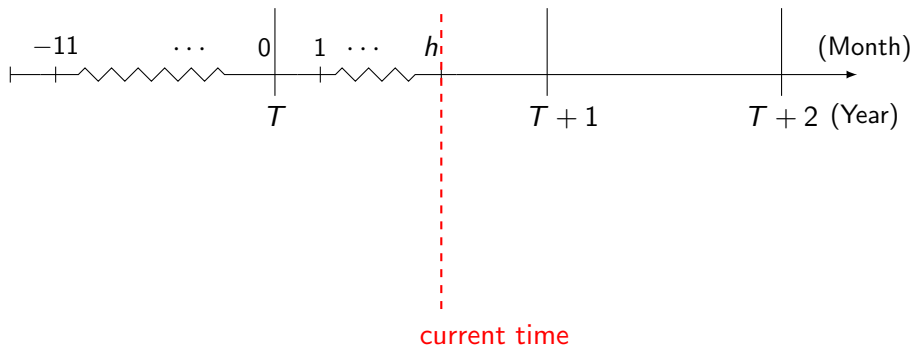
$$w(k; \lambda) = \frac{1}{12+h}$$

- Sum to unity
- Stability

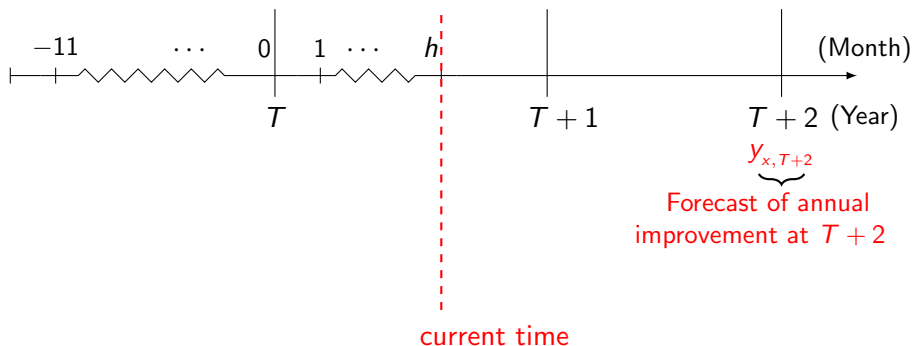
AR-MIDAS: Flexibility of Exponential Almon function



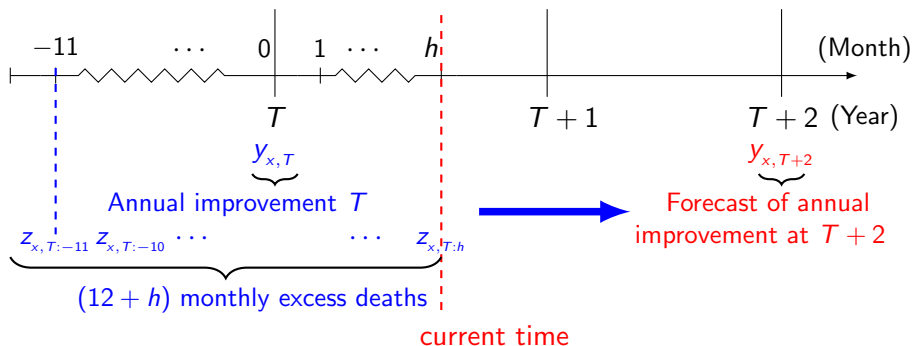
AR-MIDAS: Single age (group) forecast model



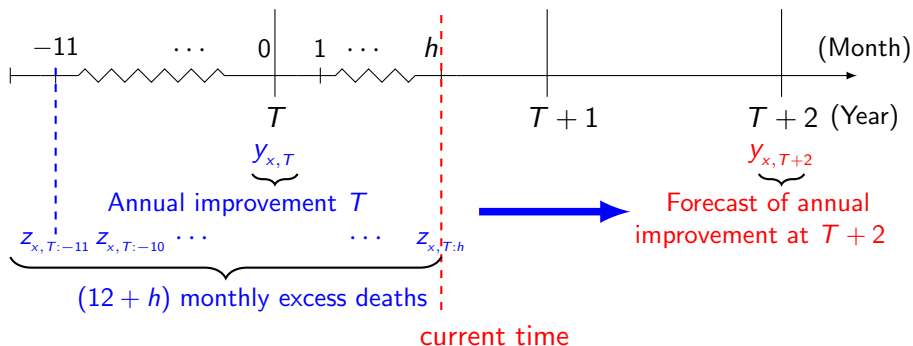
AR-MIDAS: Single age (group) forecast model



AR-MIDAS: Single age (group) forecast model



AR-MIDAS: Single age (group) forecast model



$$y_{x,h,T+2} = \boxed{\mu_x + \alpha \mu_x + \alpha^2 y_{x,T}} + \boxed{\mu_h + \alpha_h \sum_{k=-11}^h w_h(k; \lambda) z_{x,T:k}} + \text{Error}$$

- **Annual component:**
two-step ahead forecast of annual rate (T to $T+2$)

- **Monthly component:**
direct regression on monthly data

Incorporating age dependency

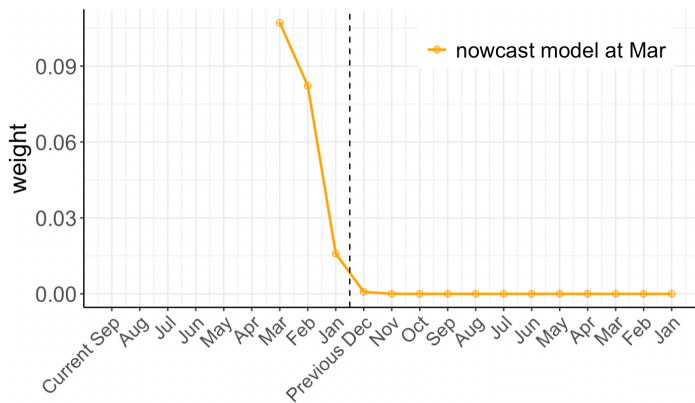
- Li and Shi (2021) uses VAR to capture age-dependency by linking annual mortality rates at age x with the two neighboring ages, $x - 1$ and $x + 1$
- VAR-MIDAS further incorporates the monthly deaths of the three ages

$$\begin{aligned}
 y_{x,h,T+1} = & \underbrace{\mu_x + \beta y_{x+1,h,T} + \alpha y_{x,h,T} + \gamma y_{x-1,h,T}}_{\text{Annual component}} \\
 & + \underbrace{\mu_h + \beta_h \sum_k w_h^{(x+1)} z_{x+1,T:k} + \alpha_h \sum_k w_h^{(x)} z_{x,T:k} + \gamma_h \sum_k w_h^{(x-1)} z_{x-1,T:k}}_{\text{Monthly component}} \\
 & + \text{Error}
 \end{aligned}$$

Estimating nowcast and forecast models

- Fitting period: 1999-2017
- Age range: 20 – 90
- Procedure:
 - (1) Estimate (AR(1) or VAR) using annual improvement rate
→ obtain residuals
 - (2) For each value of $h \in \{1, \dots, 11\}$, fit the MIDAS component to the residuals from (1) \Rightarrow obtain $\hat{\mu}_h, \hat{\alpha}_h, \hat{w}_h(k, \hat{\lambda})$
- We obtain 11 estimated models corresponding to $h \in \{1, \dots, 11\}$

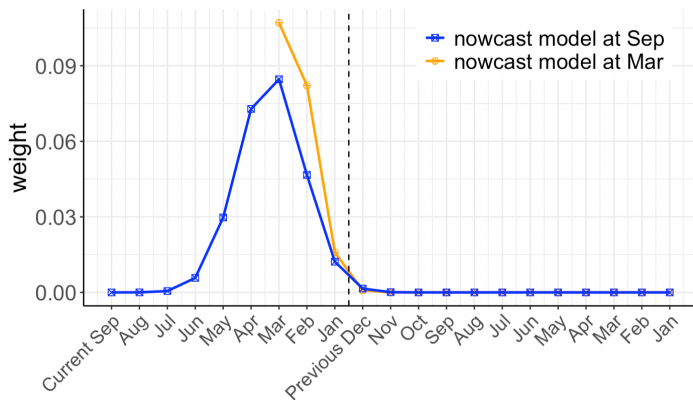
AR-MIDAS for Age 20-90, weight estimates



Parameter	Nowcast Mar
μ_h	-0.002***
α_h	0.206***
λ_1	1.799
λ_2	-0.688*

Significance levels: '***' $p < 0.01$, '**' $p < 0.05$, '*' $p < 0.1$

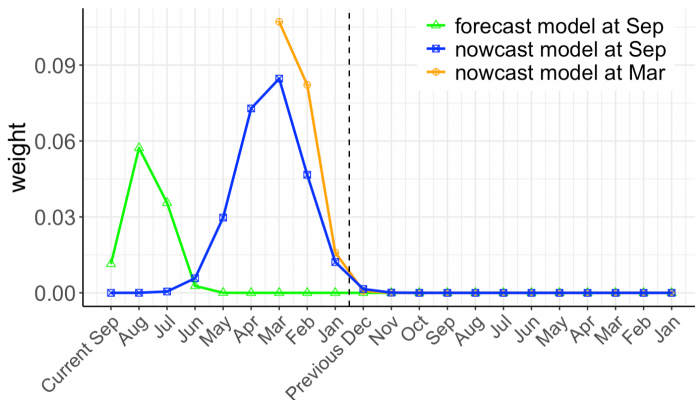
AR-MIDAS for Age 20-90, weight estimates



Parameter	Nowcast Mar	Nowcast Sep
μ_h	-0.002***	-0.003***
α_h	0.206***	0.254***
λ_1	1.799	4.999***
λ_2	-0.688*	-0.373***

Significance levels: '***' $p < 0.01$, '**' $p < 0.05$, '*' $p < 0.1$

AR-MIDAS for Age 20-90, weight estimates



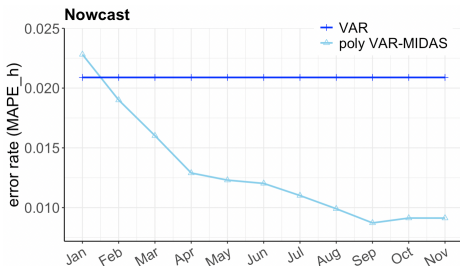
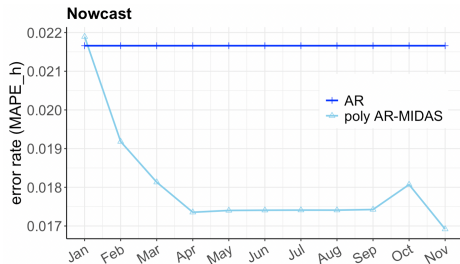
Parameter	Nowcast Mar	Nowcast Sep	Forecast Sep
μ_h	-0.002***	-0.003***	-0.002*
α_h	0.206***	0.254***	0.107***
λ_1	1.799	4.999***	4.744
λ_2	-0.688*	-0.373***	-1.044

Significance levels: '***' $p < 0.01$, '**' $p < 0.05$, '*' $p < 0.1$

Forecast performance evaluation

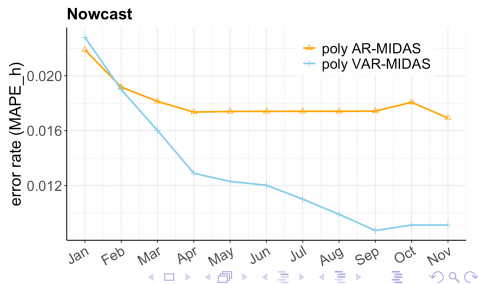
- Criterion: mean absolute percentage error (MAPE)
 - $MAPE_h$: average MAPE across all ages and years, fixing the value of h . One $MAPE_h$ is calculated for each of month 1 to 11.
 - Overall MAPE: average of $MAPE_h$ across all h
- Two types of predictions
 1. Single-period nowcast (for year 2018) and forecast (for year 2019)
 - models fitted to 1999-2017 data
 2. Rolling window nowcast and forecast:
 - models fitted to 1999-2010, 1999-2011, \dots , 1999-2017
 - nowcast for year 2011, 2012, \dots , 2018
 - forecast for year 2012, 2013, \dots , 2019

Single-period nowcasting—poly (Exponential Almon) weight

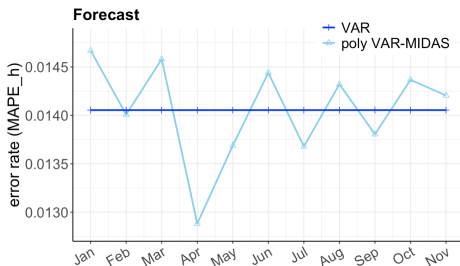
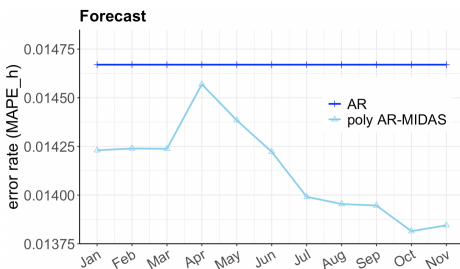


Observations and implications

- Incorporating MIDAS into AR or VAR reduces nowcasting error constantly for $h \geq 2$
- Nowcast is more accurate with age-dependency incorporated.

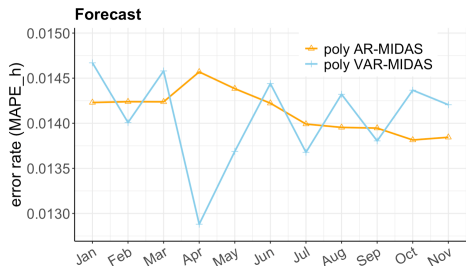


Single-period forecasting—poly weight

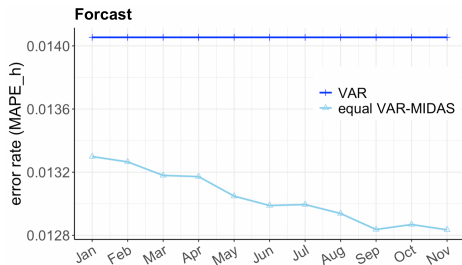
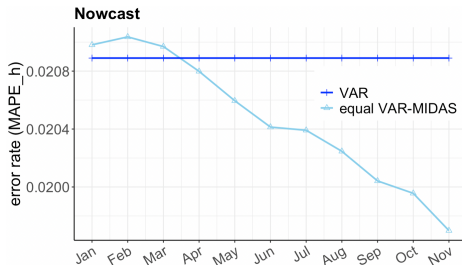


Observations and implications

- Adding MIDAS to AR improves forecasting accuracy.
- Adding MIDAS to VAR does not improve forecasting accuracy, possibly due to overfitting from the additional MIDAS shape parameters.



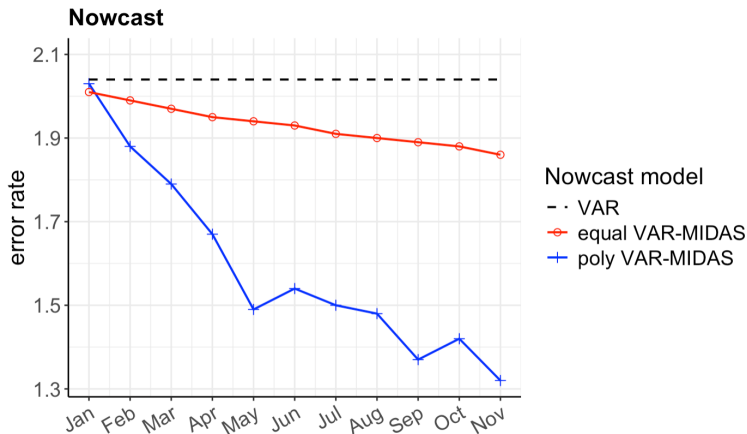
Single-period nowcasting and forecasting—equal weight



Observations and implications

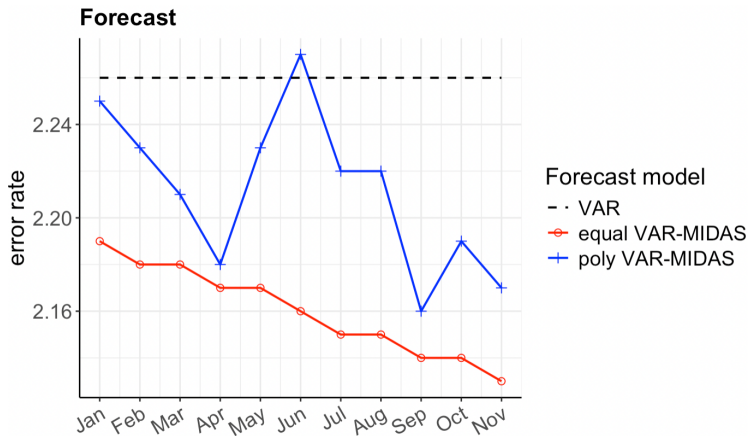
- Error rates decrease with h for both nowcast and forecast
- Equal weight MIDAS introduces no additional shape parameter, addressing overfitting issue.

Rolling window nowcasting



- MIDAS algorithm explains a significant proportion of residuals from VAR(1).
- Flexible weight function provides more accurate nowcast when two or more months of current year death records are available.

Rolling window forecasting



- Due to the use of relatively dated monthly data, flexible polynomial weight function may overfit.
- Equal weight function is preferred for forecasting.

Conclusion

This work develops and applies the MIDAS algorithm to the nowcasting and forecasting of mortality rates

- utilises monthly data to achieve timely prediction of annual mortality rates
- Demonstrates the usefulness of current year's monthly death data when producing mortality rate nowcasts and forecasts
- Identifies the significance of age dependence in improving the accuracy of mortality nowcasting and forecasting

Future research

Our future research will explore the application of MIDAS model during the pandemic.

- MIDAS model is sensitive to, and can detect year-to-year variation better than annual mortality model.
- We plan to examine and compare the performance of our proposed model with the annual model at the onset of a pandemic, as well as during and after the pandemic.