A Neural Network Extension of the Mitchell Mortality Model

Chou-Wen Wang

Professor, Department of Finance

National Sun Yat-sen University

September 12 and 13, 2022

(with Hung-Tsung Hsiao and I-Chien Liu)

Outline

- Introduction
- Mortality Model under Neutral Network
- Mortality Prediction
- Conclusion

Introduction

- As the populations of the world's leading economies age, longevity risk has threatened the stability of the global financial system and become one of important , which brings unexpected challenges to individuals, corporations, and governments.

- Meanwhile, the rapid advances in Artificial Intelligence (AI) are also changing the environment in which we live. Advanced machine learning techniques can stimulate better solutions for longevity risk management through providing more sophisticated and accurate mortality models (Richman, 2018).

Contribution

- This paper proposes a neural network approach for fitting the Mitchell model with cohort effect to improve mortality modeling and forecasting.

- As well as we knows, this paper first incorporate the cohort effect in the neural network architecture.

- We develop
  - some neural networks that replicate the structure of the individual Mitchell model and
  - allow their joint fitting by analyzing the mortality data of all the considered populations simultaneously.

Contribution

- To this purpose, we embed the individual Mitchell models with cohort effects into a neural network in which the Mitchell parameters are jointly estimated by processing the mortality data of all populations simultaneously.

- The main advantage of our method is that the required parameters of Mitchell neutral model for mortality prediction **can be directly obtained (known when the training of neural network is finished)** without any further estimation by using time-series models.

Introduction

- Mortality rates have changed over time, techniques to predict future mortality rates become important for both demography and actuarial science.

- Starting from the Lee and Carter (1992), various mortality models have been developed

(see, e.g., Lee and Carter, 1992; Cairns et al., 2006; Renshaw and Haberman, 2006; Cairns et al., 2009; Mitchell et al., 2013). For recent development of mortality modeling and forecasting, see Blake et al. (2018) and the references therein.

Introduction

- Two well known examples of these techniques are the Lee-Carter (LC) and the Cairns-Blake-Dowd (CBD) models, which forecast mortality rates in two steps: (Richman and Wüthrich, 2020)
  - First, a low dimensional summary of past mortality rates is constructed by fitting statistical models to historical mortality data
  - Second, future mortality rates are projected by extrapolating the summarized mortality rates into the future using time series models.

Introduction

- The LC and CBD models are originally used to single populations. If we forecast for multiple populations, the models were fit to each population separately.

- However, as pointed out by Richman and Wüthrich (2020), it seems reasonable to expect that a large-scale **multi-population mortality forecasting** model would produce **more robust forecasts** of future mortality rates than those produced by single-population models.

Introduction

- In recent years, there are also much literature on employing machine learning in mortality modeling.
- Hainaut (2018) uses autoencoder networks to estimate latent factors for mortality.
- Based on neural network approach, Richman and Wüthrich(2020) seek to offer an alternative multi-population mortality forecasting model by extending the LC model to multiple populations.

Introduction

- Dong et al. (2020) use tensor decomposition to construct multi-population mortality models.

- Perla et al. (2021) use deep learning algorithms, such as recurrent neural network (RNN) and 1-dimensional CNN that are suitable for sequential data, to perform time-series mortality forecasting.

- Wang et al. (2021) provides a neighboring prediction model for mortality, which can capture the neighborhood effect and achieves more accurate out-of-sample forecasting.

Introduction

- Salvatore (2022) proposes a neural network approach for calibrating the LC models of multiple populations. The parameters of the LC models are jointly estimated through a neural network that simultaneously processes the mortality data of all populations.

- As shown by Salvatore (2022) , the best performance is the LC_LCN_Poisson model which employs locally connected layers to extract features from the mortality rate curves. (Local connectivity since each unit of the layer is connected only with a local area of the input)

=>As a result, we use LC_LCN_Poisson model as the benchmark for our mortality prediction.

The Lee-Carter Model

$$ln\left(m_{x,t}\right) = \alpha_x + \beta_x k_t + e_{x,t}$$

- $m_{x,t}$ is the central death rate for age *x* in calendar year *t*, defined as the span from time to time .
- α_x describes the average pattern of mortality for the age group
- β_x represents the age-specific patterns of mortality change
- k_t explains the time trend of the general mortality level
- $e_{x,t}$ represents the deviation of the model from the observed log-central death rates, which should be a normal distribution with zero mean and a relatively small variance (Lee, 2000).

The Lee-Carter Model

- Two constraint conditions:

  $$\sum_t k_t = 0 \text{ and } \sum_x \beta_x = 1$$

- $\hat{\alpha}_x$ is simply the average value over time of $ln(m_{x,t})$
- $\hat{k}_t$ is the sum over various ages of $ln(m_{x,t}) - \hat{\alpha}_x$
- Using $ln(m_{x,t}) - \hat{\alpha}_x$ as the dependent variable and $\hat{k}_t$ as the explanatory variable, we can obtain $\hat{\beta}_x$ by using a simple regression model without an intercept parameter.

The Lee-Carter Model

$$k_t^* - k_{t-1}^* = \gamma + \varepsilon_t$$

where γ is a drift term, and ε_t is a sequence of independent and identically Gaussian random variables with mean 0 and variance σ^2.

Renshaw and Haberman (2006) Model

$$ln\left(m_{x,t}\right) = \alpha_x + \beta_x k_t + \eta_x \gamma_{t-x} + e_{x,t}$$

$$k_t - k_{t-1} = \mu + \varepsilon_t \quad \Leftrightarrow \quad \text{ARIMA(0,1,0)}$$

$$\Delta\gamma_{t-x} = \mu_\gamma + \alpha_\gamma\left(\Delta\gamma_{t-x-1} - \mu_\gamma\right) + \sigma_\gamma z_{t-x} \quad \Leftrightarrow \quad \text{ARIMA(1,1,0)}$$

Mitchell et al. (2013)

- Mitchell et al. (2013) propose an easy-to-implement method to model the mortality rates. They adopt the Lee-Carter model (Lee and Carter, 1992) on the log-difference of mortality rates and show that this improves the forecast accuracy.

$$\log m_{x,t+1} = \log m_{x,t} + a_x + b_x k_t.$$

Model Setup

- In this paper, we consider the model of Mitchell et al. (2013) with cohort effect as follows:

$$\log m_{x,t+1} = \log m_{x,t} + a_x + b_x k_t + \eta_x \gamma_{t-x}.$$

where $a_x, b_x, k_t, \eta_x, \gamma_{t-x}$ are estimated by the neural network with the mortality data of all populations.

Model Setup

- Following Salvatore (2022), we model a_x, b_x and η_x by the Fully-Connected Networks (FCN) layer and embedding layer as a learned representation of the gender and region labels.

Programming Code for a_x: (Similar coding for b_x)

```
Country <- layer_input(shape = c(1), dtype = 'int32', name = 'Country'  )
Country_embed = Country%>% layer_embedding(input_dim = idx_region,
output_dim = q_e)%>%layer_flatten(name= 'Country_embed'  )

Sex <- layer_input(shape = c(1), dtype = 'int32', name = 'Sex')
Sex_embed = Sex%>%layer_embedding(input_dim = 2, output_dim =
q_e)%>%layer_flatten(name= 'Sex_embed')

ax = Country_embed %>% list(Sex_embed)
%>%layer_concatenate()%>%layer_dropout(0.5)%>% layer_dense (unit = nyear,
name = "ax")
```

Model Setup

- Following Salvatore (2022), we model k_t by using LCN layer to produces as output a single value. However, different from Salvatore (2022), we use the **log of mortality ratio ($m_{x,t}$/ $m_{x,t-1}$)** to extract the factor k_t.

- **k_t is projected using the mean of k_t($\sum_t k_t = 0$) in training set.**

Programming Code for k_t:

```
kt = rates%>%layer_reshape(c(nyear,1))%>%
           layer_locally_connected_1d(filter =1, activation = act,
           kernel_size =nyear/q_z, strides = nyear/q_z)%>%
           layer_flatten()%>%layer_dropout(0.05)%>%
           layer_dense(units = 1, activation = 'linear')%>%
           layer_reshape(c(1,1) , name = "kt")
```

Model Setup

- We model γ_{t-x} by using LCN layer.

- Different from the traditional estimating procedure,
  - the γ_{t-x} under the neural network framework are related to cohort year and **are changed over time, which means that, for the same cohort year, different combination of t(time) and x (age) has different γ_{t-x}.**
  - **In addition, because *cohort years* are predictable (known) , *the future values of γ_{t-x}* are also *predictable (known)* without further estimating by time-series approach.**

Mortality Prediction

- The data source is the Human Mortality Database (HMD), which provides mortality data for male and female populations of a large set of countries.

- Following the experiments' scheme in Perla et al. (2021), Richman and Wüthrich(2020) and Salvatore (2022), we only consider mortality data from 1950 onwards, and we set 1999 as final observation.

- The mortality rates of calendar years in **$\{t \geq 2000\}$** is **testing set**, using a model fitted on mortality data of calendar years in **$\{1950 \leq t < 2000\}$ (training set)**

HMD Data

Country	Gender	Start Year	End Year
Australia	Female	1921	2019
	Male	1921	2019
Austria	Female	1947	2019
	Male	1947	2019
Bulgaria	Female	1947	2020
	Male	1947	2020
Canada	Female	1921	2019
	Male	1921	2019
Czechia	Female	1950	2019
	Male	1950	2019
Denmark	Female	1835	2021
	Male	1835	2021
Finland	Female	1878	2021
	Male	1878	2021
France	Female	1816	2019
	Male	1816	2019
Hungary	Female	1950	2020
	Male	1950	2020
Iceland	Female	1838	2020
	Male	1838	2020
Ireland	Female	1950	2020
	Male	1950	2020

Italy	Female	1872	2019
	Male	1872	2019
Japan	Female	1947	2020
	Male	1947	2020
Netherlands	Female	1850	2019
	Male	1850	2019
New Zealand	Female	1948	2021
	Male	1948	2021
Norway	Female	1846	2020
	Male	1846	2020
Portugal	Female	1940	2021
	Male	1940	2021
Slovakia	Female	1950	2019
	Male	1950	2019
Spain	Female	1908	2020
	Male	1908	2020
Sweden	Female	1751	2021
	Male	1751	2021
Switzerland	Female	1876	2021
	Male	1876	2021
U.K.	Female	1922	2020
	Male	1922	2020
U.S.A.	Female	1933	2020
	Male	1933	2020

Numerical Results

Country	Gender	LC	CBD	Salvatore (2022)		Our Model	
				MAPE_Train	MAPE_Test	MAPE_Train	MAPE_Test
Australia	Female	11.37%	11.68%	8.28%	9.54%	4.28%	**6.55%**
	Male	20.32%	18.64%	8.33%	**13.09%**	3.76%	13.91%
Austria	Female	11.22%	11.75%	6.71%	18.35%	3.98%	**8.79%**
	Male	16.98%	14.76%	6.82%	**12.01%**	3.84%	14.48%
Bulgaria	Female	18.33%	14.31%	9.20%	21.69%	5.77%	**13.80%**
	Male	14.83%	**10.79%**	8.77%	13.07%	5.58%	10.91%
Canada	Female	8.29%	9.25%	9.57%	7.92%	3.30%	**5.70%**
	Male	21.58%	20.56%	9.50%	17.10%	2.95%	**16.76%**
Czechia	Female	15.34%	13.91%	6.29%	15.01%	3.87%	**13.13%**
	Male	17.66%	**15.74%**	6.60%	18.07%	3.86%	16.21%
Denmark	Female	18.32%	17.93%	11.44%	**16.07%**	5.21%	17.83%
	Male	22.34%	20.40%	7.47%	**18.13%**	4.78%	23.12%
Finland	Female	17.59%	16.68%	9.14%	15.03%	5.67%	**11.59%**
	Male	22.67%	20.70%	8.73%	19.29%	5.54%	**14.20%**
France	Female	16.68%	16.67%	6.69%	15.69%	1.92%	**6.32%**
	Male	20.64%	19.99%	7.18%	13.88%	1.82%	**10.78%**
Hungary	Female	16.01%	15.08%	7.36%	20.10%	4.29%	**13.61%**
	Male	26.61%	**17.49%**	8.03%	18.35%	4.30%	18.87%
Iceland	Female	**21.12%**	21.15%	26.18%	24.82%	22.48%	49.58%
	Male	**21.84%**	21.85%	21.58%	106.48%	19.86%	47.71%
Ireland	Female	25.32%	**23.57%**	11.38%	24.24%	7.35%	26.28%
	Male	33.40%	**32.19%**	9.79%	48.49%	6.27%	42.65%

Abbr	Gender	LC	CBD	Salvatore (2022)		Our Model	
				MAPE_Train	MAPE_Test	MAPE_Train	MAPE_Test
Italy	Female	12.72%	12.20%	6.74%	11.15%	2.30%	**5.93%**
	Male	19.32%	18.18%	6.44%	18.70%	2.17%	**15.99%**
Japan	Female	12.90%	8.20%	8.52%	17.08%	2.19%	**6.06%**
	Male	7.29%	5.62%	6.45%	7.44%	2.05%	**5.53%**
Netherlands	Female	13.75%	13.56%	8.54%	12.69%	3.21%	**6.68%**
	Male	22.90%	**22.06%**	7.63%	19.57%	2.82%	24.05%
New Zealand	Female	**12.72%**	13.63%	11.37%	16.44%	8.40%	14.76%
	Male	22.58%	**20.97%**	9.71%	30.42%	7.24%	24.43%
Norway	Female	16.78%	16.18%	8.54%	22.21%	4.97%	**10.83%**
	Male	25.77%	24.69%	8.30%	**21.53%**	4.69%	29.07%
Portugal	Female	18.44%	17.37%	8.40%	24.91%	5.07%	**16.28%**
	Male	17.95%	16.06%	7.18%	15.54%	4.37%	**14.35%**
Slovakia	Female	13.66%	**13.21%**	7.81%	21.28%	6.38%	13.27%
	Male	25.34%	**16.31%**	8.71%	16.86%	6.16%	18.47%
Spain	Female	18.94%	14.04%	8.22%	22.90%	3.07%	**8.80%**
	Male	18.64%	15.61%	7.75%	15.40%	2.83%	**11.36%**
Sweden	Female	14.03%	12.88%	8.32%	8.71%	4.26%	**6.92%**
	Male	21.37%	19.46%	8.97%	23.84%	3.75%	**14.65%**
Switzerland	Female	11.38%	10.89%	8.72%	7.51%	5.45%	**7.94%**
	Male	19.12%	17.43%	7.42%	26.68%	4.65%	**12.54%**
U.K.	Female	14.67%	13.27%	6.32%	18.68%	2.17%	**10.57%**
	Male	22.95%	20.64%	7.67%	26.35%	2.06%	**19.50%**
U.S.A.	Female	5.54%	6.83%	11.93%	17.15%	1.77%	**5.22%**
	Male	14.17%	13.13%	10.28%	20.54%	1.57%	**11.63%**
Average		**17.86%**	**16.25%**	9.02%	**20.22%**	4.88%	**15.60%**

Numerical Results

- As shown by Table 2, we can see that, In terms of **overall average MAPE**,
  - our model is 15.60% .
  - The model of Salvatore (2022) is 20.22%
  - The LC model is 17.86%
  - The CBD model is 16.25%

- In terms of **average MAPE**, for different gender and region,
  - 76.09% of our model is better than the model of Salvatore (2022)
  - 80.43% of our model is better than the LC model
  - 69.57% of our model is better than the CBD model

Conclusion

- In this paper, following Richman and Wüthrich(2020) and Salvatore (2022), we provide a neural network version of Mitchell et al. (2013) model with cohort effect.

- In most cases, k_t and/or γ_{t-x} are needed re-estimated by time-series model for mortality prediction. The advantage of this paper is that the parameters of k_t and γ_{t-x} **do not need be** re-estimated by time-series model for mortality prediction.

Thank You