Joint Calibration of Multiple Renshaw-Haberman Models using Neural Networks

Francesca Perla[1] Salvatore Scognamiglio[1]

[1] Department of Management and Quantitative Studies, University of Naples, Parthenope

Longevity 19 risk conference

Introduction

Single-Population Models: Lee and Carter (1992), Renshaw and Haberman (2006), Cairns, Blake, and Dowd (2006)

- Can be applied to a single population.
- Fitted on population-specific subset of data.

Introduction

Single-Population Models: Lee and Carter (1992), Renshaw and Haberman (2006), Cairns, Blake, and Dowd (2006)

- Can be applied to a single population.
- Fitted on population-specific subset of data.

Multi-Population Models: Li and Lee (2005), Kleinow (2015)

- Generally applied to small subsets of data;
- Usually intended for forecasting the mortality of similar populations;
- Difficult to fit (complex optimization schemes/less commonly used statistical techniques).

Introduction

Single-Population Models: Lee and Carter (1992), Renshaw and Haberman (2006), Cairns, Blake, and Dowd (2006)

- Can be applied to a single population.
- Fitted on population-specific subset of data.

Multi-Population Models: Li and Lee (2005), Kleinow (2015)

- Generally applied to small subsets of data;
- Usually intended for forecasting the mortality of similar populations;
- Difficult to fit (complex optimization schemes/less commonly used statistical techniques).

Jointly calibrated Single-Population Models:

- Allows for a more efficient use of the available mortality data.
- Provides robust calibration of the single-population mortality model parameters.
- Can be applied also to many populations.

The Renshaw-Haberman Model

Let $\mathcal{X} = \{x_0, x_1, \ldots, x_\omega\}$ be the set of the ages and $\mathcal{T} = \{t_0, t_1, \ldots, t_n\}$ the set of calendar years considered, and $\mathcal{I} = \{\text{pop}_1, \text{pop}_2, \ldots, \text{pop}_{n_p}\}$ the set of populations.

The Renshaw and Haberman model (RH) defines the death count in age x at time t in the population i denoted as $D^{(i)}_{x,t}$ as a Poisson random variable:

$$D^{(i)}_{x,t} \sim Poisson(E^{(i)}_{x,t}\eta^{(i)}_{x,t})$$

where $E^{(i)}_{x,t}$ is the number of exposure-to-risk and:

$$\log \eta^{(i)}_{x,t} = a^{(i)}_x + b^{(i)}_x k^{(i)}_t + c^{(i)}_x \gamma^{(i)}_{t-x}.$$

Identifiability of the model can be ensured using the following set of parameter constraints:

$$\sum_x b^{(i)}_x = 1 \quad \sum_t k^{(i)}_t = 0 \quad \sum_x c^{(i)}_x = 1 \quad \sum_{l=t_0-x_\omega}^{t_n-x_0} \gamma^{(i)}_l = 0.$$

Model Fitting

Parameters can be estimated via Maximum Likelihood Estimator.

We have to solve the sequence of problems:

$$\underset{\boldsymbol{a}^{(i)},\boldsymbol{b}^{(i)},\boldsymbol{c}^{(i)},\boldsymbol{k}^{(i)},\gamma^{(i)}}{\arg\max} \sum_{x\in\mathcal{X}} \sum_{t\in\mathcal{T}} \left(D_{x,t}^{(i)} \log \eta_{x,t}^{(i)} - E_{x,t}^{(i)} \eta_{x,t}^{(i)} \right) + H_i, \quad \forall i \in \mathcal{I}$$

where $H_i \in \mathbb{R}$ is a population specific constant.

Model Fitting

Parameters can be estimated via Maximum Likelihood Estimator.

We have to solve the sequence of problems:

$$\underset{\boldsymbol{a}^{(i)},\boldsymbol{b}^{(i)},\boldsymbol{c}^{(i)},\boldsymbol{k}^{(i)},\gamma^{(i)}}{\arg\max} \sum_{x\in\mathcal{X}} \sum_{t\in\mathcal{T}} \left(D_{x,t}^{(i)} \log \eta_{x,t}^{(i)} - E_{x,t}^{(i)} \eta_{x,t}^{(i)} \right) + H_i, \quad \forall i \in \mathcal{I}$$

where $H_i \in \mathbb{R}$ is a population specific constant.

In this original version, the model is fitted by using the population-specific data that is a subset of the available mortality data.

Real Data Applications: United States

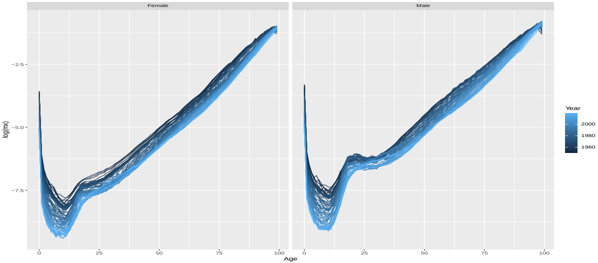


Figure: Log mortality rates for different ages in the USA from 1950 to 2018. Source: HMD.

Real Data Applications: United States

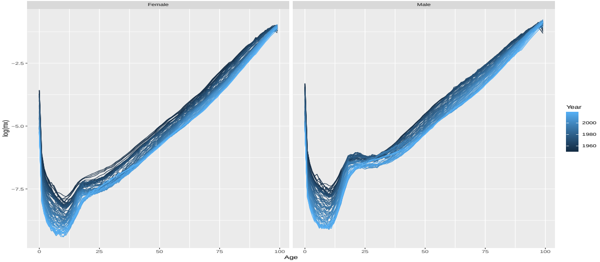


Figure: Log mortality rates for different ages in the USA from 1950 to 2018. Source: HMD.

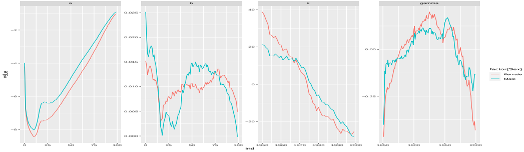


Figure: Estimates of the RH parameters for USA: fitting period 1950-1999.

Real Data Applications: Greece

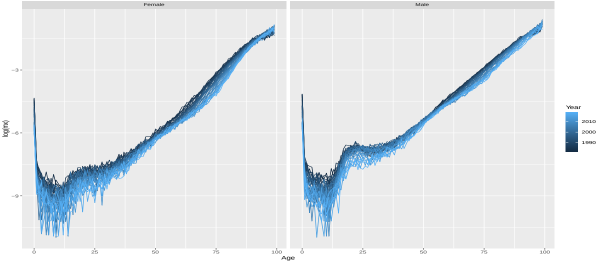


Figure: Log mortality rates for different ages in Greece from 1950 to 2018. Source: HMD.

Real Data Applications: Greece

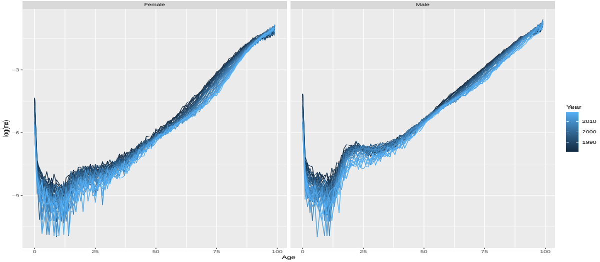


Figure: Log mortality rates for different ages in Greece from 1950 to 2018. Source: HMD.

Figure: Estimates of the RH parameters for Greece: fitting period 1950-1999.

Real Data Applications: Luxembourg

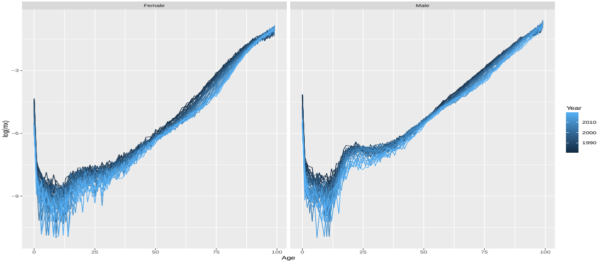


Figure: Log mortality rates for different ages in Luxembourg from 1950 to 2018. Source: HMD.

Real Data Applications: Luxembourg

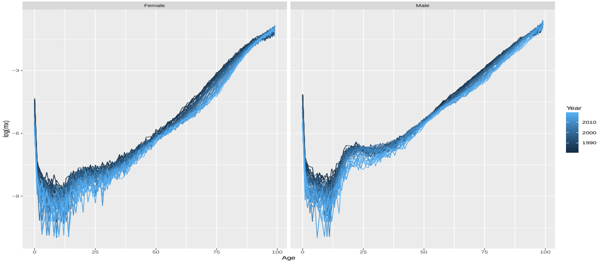


Figure: Log mortality rates for different ages in Luxembourg from 1950 to 2018. Source: HMD.

No solutions to the calibration problem were found.

Real Data Applications: list of populations

#	Country	Male	Female
1	USA		
2	RUS		
3	JPN		
4	GER-W		
5	GER-E		
6	ITA		
7	FRA		
8	GBR-TENW		
9	ESP		
10	UKR		
11	POL		
12	CAN		
13	AUS		
14	NLD		
15	BEL		
16	SWE		X
17	CHE	X	X
18	CZE		
19	GRC		
20	PRT		

#	Country	Male	Female
21	HUN		
22	AUT	X	X
23	DNK		X
24	NOR		X
25	IRL		X
26	FIN	X	X
27	SVK	X	X
28	LUX	X	X
29	SVN	X	X
30	EST	X	X
31	GBR-SCO	X	X
32	GBR-NIR	X	X
33	ISL	X	X
34	NZL-NM	X	X
35	BLR		
36	BGR		
37	LTU		
38	LVA		
39	ISR		
40	TWN		

Neural Networks

NNs are high-dimensional non-linear regression models.

Let $\boldsymbol{y} = (y_1, y_2, \dots, y_{q_0})^\top \in \mathbb{R}^{q_0}$, a FCN layer with $q_1 \in \mathbb{N}$ units is a function

$$\boldsymbol{z} : \mathbb{R}^{q_0} \to \mathbb{R}^{q_1}, \qquad \boldsymbol{y} \mapsto \boldsymbol{z}(\boldsymbol{y}) = (z_1(\boldsymbol{y}), z_2(\boldsymbol{y}), \dots, z_{q_1}(\boldsymbol{y}))^\top .$$

Each new feature component $z_j(\boldsymbol{y})$ is a non-linear function of $\boldsymbol{y}$

$$\boldsymbol{y} \;\mapsto\; \boldsymbol{z}_j(\boldsymbol{y}) = \phi\left(w_{j,0} + \sum_{l=1}^{q_0} w_{j,l} y_l\right) = \phi\left(w_{j,0} + \langle \boldsymbol{w}_j, \boldsymbol{y} \rangle\right), \qquad j = 1, \dots, q_1,$$

where $\phi : \mathbb{R} \to \mathbb{R}$ is a (non-linear) activation function, $w_{j,l} \in \mathbb{R}$ represent the network parameters and $\langle \cdot, \cdot \rangle$ denotes the scalar product in $\mathbb{R}^{q_0}$.

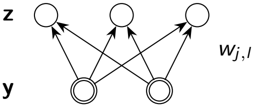


Figure: A FCN layer.

Deep Neural Networks

Let $\boldsymbol{q} = \{q_k\}_{1 \le k \le m} \in \mathbb{N}^m$ and $m \in \mathbb{N}$. A deep FCN can be formalised as:

$$\boldsymbol{y} \;\mapsto\; \boldsymbol{z}^{(m:1)}(\boldsymbol{y}) = \left(\boldsymbol{z}^{(m)} \circ \cdots \circ \boldsymbol{z}^{(1)}\right)(\boldsymbol{y}) \;\in\; \mathbb{R}^{q_m},$$

where all mappings $\boldsymbol{z}^{(k)} : \mathbb{R}^{q_{k-1}} \to \mathbb{R}^{q_k}$ are FCN layers with weights $W^{(k)} = (\boldsymbol{w}_j^{(k)})_{1 \le j \le q_k} \in \mathbb{R}^{q_k \times q_{k-1}}$ and biases $\boldsymbol{w}_0^{(k)} \in \mathbb{R}^{q_k}$, for $1 \le k \le m$.

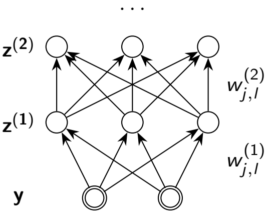


Figure: A deep FCN network.

Deep NNs are powerful tools for function approximation; see the universal approximation theorem (Hornik et al. 1989).

Joint Calibration of multiple Renshaw and Haberman Models

We reformulate the RH model as follows:

$$D_{x,t}^{(i)} \sim Poisson(E_{x,t}^{(i)} \eta_{x,t}^{(i)})$$

where

$$\log \eta_{x,t}^{(i)} = a(x,i) + b(x,i)k(t,i) + c(x,i)\gamma(x-t,i).$$

The RH parameters are expressed as functions that depend on the input data and can be learned using supervised learning methods.

In this setting, the model calibration is carried out by solving:

$$\underset{a(),b(\cdot),c(\cdot),k(\cdot),\gamma(\cdot)}{\arg\max} \sum_{i\in\mathcal{I}} \sum_{x\in\mathcal{X}} \sum_{t\in\mathcal{T}} \left(D_{x,t}^{(i)} \log \eta_{x,t}^{(i)} - E_{x,t}^{(i)} \eta_{x,t}^{(i)} \right) + H$$

where $H \in \mathbb{R}$.

- Human Mortality Database:
  - Ages: $\mathcal{X} = \{x \in \mathbb{N}_0 : 0 \leq x \leq 99\}$.
  - Period: $\mathcal{T} = \{t \in \mathbb{N} : 1950 \leq t \leq 2018\}$.
  - 80 populations considered.

  Data Partitioning:
  - Train data $\mathcal{T}_{train} = \{t \in \mathbb{N} : 1950 \leq t \leq 1999\}$;
  - Test data $\mathcal{T}_{test} = \{t \in \mathbb{N} : 2000 \leq t \leq 2018\}$.

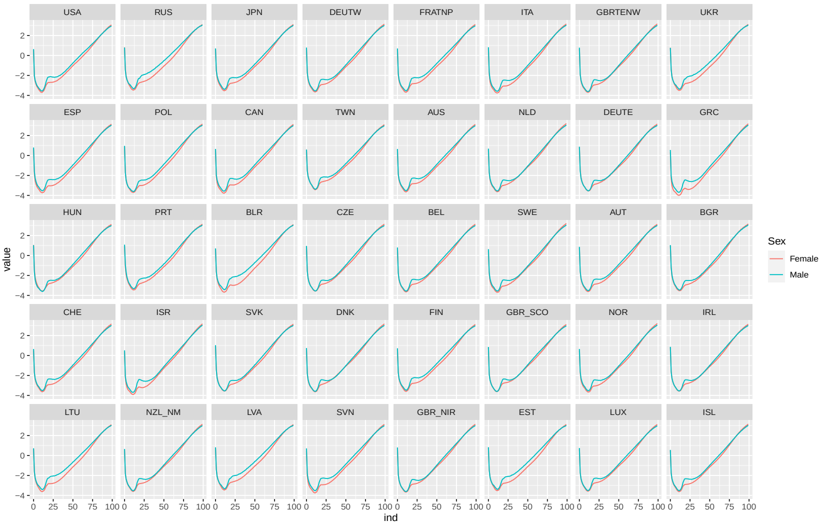
We compare:

- **RH**: the RH model individually calibrated using `StMoMo`;
- **JCRH**: the RH model jointly calibrated using Neural Networks.

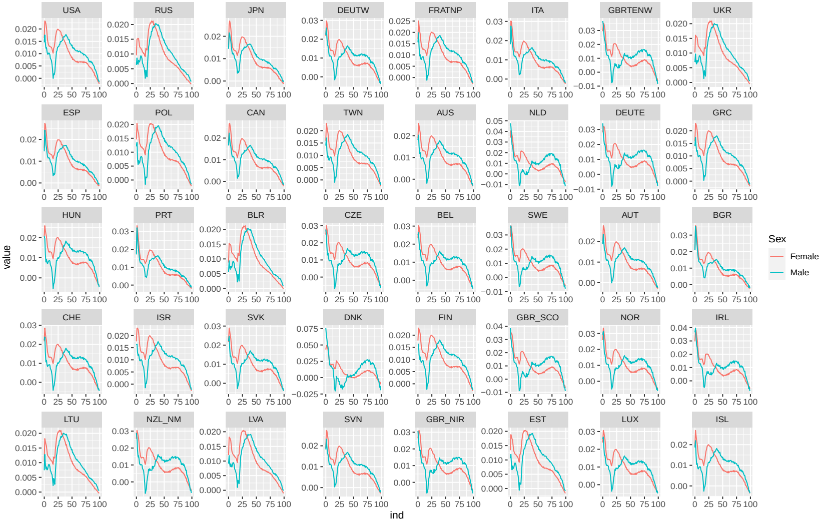
Forecasting Mean Squared Error:

$$MSE_i = \frac{1}{|\mathcal{T}_{test}| \cdot |\mathcal{X}|} \sum_{t \in \mathcal{T}_{test}} \sum_{x \in \mathcal{X}} (m_{x,t}^{(i)} - \hat{m}_{x,t}^{(i)})^2, \quad \forall i \in \mathcal{I}.$$

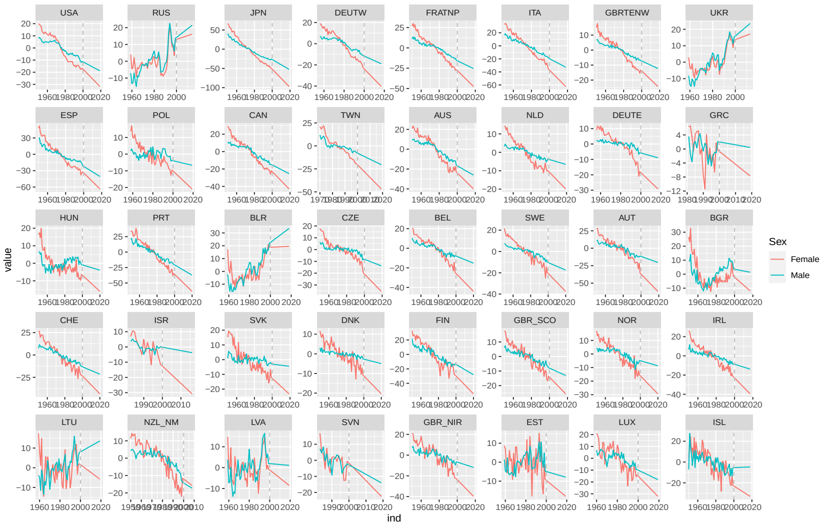
estimates of $(a_x^{(i)})_{x \in \mathcal{X}}$



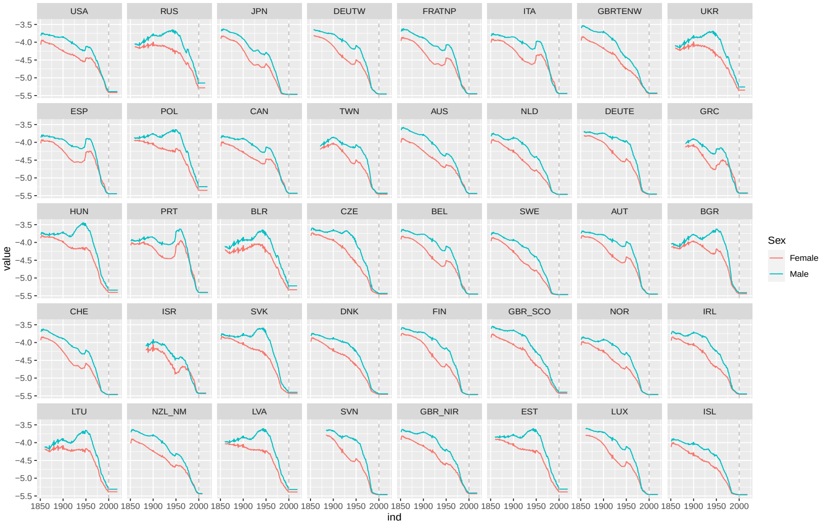
estimates of $(b_x^{(i)})_{x\in\mathcal{X}}$



estimates of $(k_t^{(i)})_{t\in\mathcal{T}}$



estimates of $(\gamma_l^{(i)})_{l=t_0-x_\omega}^{t_n-x_0}$



Forecasting MSE by population

Country	Female		Male	
	RH	JCRH	RH	JCRH
USA	**0.30**	1.01	**0.48**	0.81
RUS	**2.52**	3.29	**4.90**	6.54
JPN	16.13	**0.51**	24.42	**1.59**
DEUTW	1.78	**1.51**	1.27	**0.68**
DEUTE	3.31	**0.78**	**3.78**	3.21
ITA	10.60	**0.79**	8.59	**0.57**
FRA	3.91	**1.23**	4.63	**1.30**
GBRTENW	0.52	**0.19**	**0.57**	1.45
ESP	17.00	**1.48**	19.82	**0.98**
UKR	32.39	**4.74**	5.02	**3.67**
POL	1.53	**0.93**	7.89	**6.75**
CAN	11.54	**0.44**	12.25	**0.57**
AUS	**0.65**	1.02	0.91	**0.90**
NLD	8.71	**3.01**	**1.16**	1.45
BEL	4.62	**1.53**	5.19	**1.66**
SWE		**2.00**	**1.93**	3.04
CHE		**4.03**		**2.95**
CZE	**0.76**	0.77	**3.96**	5.69
GRC	3.17	**1.16**	2.04	**1.45**
PRT	12.88	**1.13**	10.44	**2.21**

Country	Female		Male	
	RH	JCRH	RH	JCRH
HUN	**1.25**	1.57	13.71	**10.30**
AUT		**1.64**		**2.28**
DNK		**0.41**	**1.96**	2.64
NOR		**1.89**	**2.89**	2.90
IRL		**1.20**	**4.30**	4.76
FIN		**1.46**		**3.24**
SVK		**3.77**		**13.05**
LUX		**5.48**		**15.64**
SVN		**2.34**		**10.63**
EST		**3.55**		**20.39**
GBR-SCO		**0.45**		**3.53**
GBR-NIR		**1.44**		**6.68**
ISL		**9.68**		**22.75**
NZL-NM		**1.17**		**3.79**
BLR	1107.41	**4.24**	12.58	**5.42**
BGR	11.55	**3.65**	**7.76**	6.60
LTU	**22.11**	**10.38**	**10.43**	**8.90**
LVA	**3.48**	4.84	16.09	**12.95**
ISR	3.45	**1.73**	5.25	**1.46**
TWN	7.01	**1.53**	2.70	**3.37**

Forecasting Mean Squared Error of the RH and the JCRH models for different populations. Values are multiplied by 10^5.

Future Research

In future research, we intend to

- Jointly calibrate multiple RH models without the assumption $c_x^{(i)} = 1$.
- Extend the joint calibration approach to other stochastic mortality models with cohort effects.
- Investigate the use of modern deep learning models, such as transformers.

Future Research

In future research, we intend to

- Jointly calibrate multiple RH models without the assumption $c_x^{(i)} = 1$.
- Extend the joint calibration approach to other stochastic mortality models with cohort effects.
- Investigate the use of modern deep learning models, such as transformers.

For comments or suggestions:

salvatore.scognamiglio@uniparthenope.it