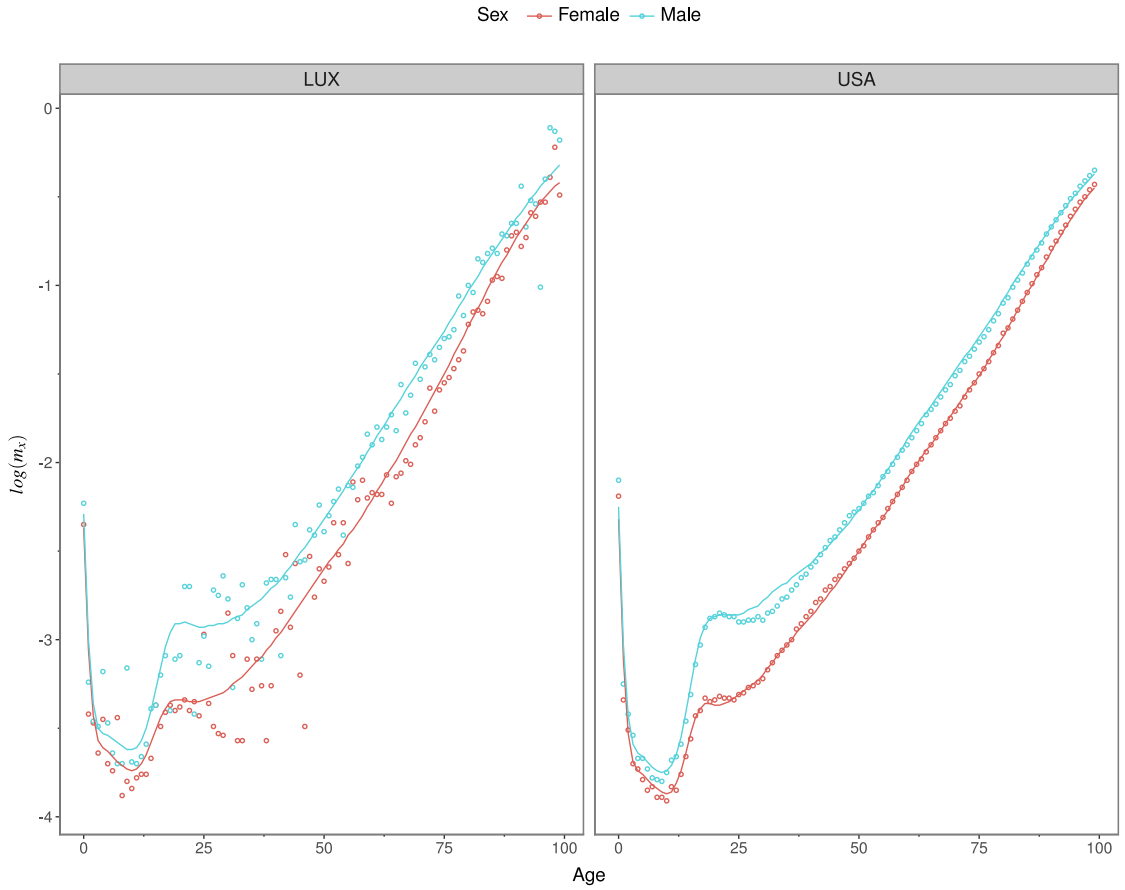


Deep Learning with Whittaker and Henderson

C. Balona / R. Richman

AGENDA

Lee-Carter Method

Lee-Carter Neural Network (LCNN)

Roughness and Whittaker-Henderson

Results and Additional Findings

Why should we care about mortality forecasting?

- Mortality modelling and forecasting is fundamental to management of longevity risks…
- … and useful for diagnosing trends in risk business
- Major topic in overseas markets, see, e.g. CMI Mortality Projections Model 2021
- Especially topical post-COVID to try ascertain future levels of mortality improvements
- Do we need to give more emphasis to this in SA context? Higher rates = no but …
- … IFRS 17 Risk Adjustments and SAM/ORSA requirements ~= uncertainty of forecasts = yes!

Lee-Carter Model – Where it all started?

- Foundational model for mortality forecasting is the Lee-Carter model (Lee and Carter 1992) (LC model)
- Many other approaches; within actuarial literature see Cairns, Blake and Dowd (2006) for an approach(CBD model) suited to old-age mortality (model coefficients of logistic model of qx)
- Mortality over time modelled using:

$$\log\left(u_{x,t}\right) = a_x + b_x k_t$$

- i.e. (log) mortality = average rate + rate of change . time index
- Relies on latent variables that must be estimated from data and then multiplied
- Could use interaction term between the variables Year and Age but this specification would require t.x effects to be fit compared to the t+x effects in the Lee-Carter model.
- => use non-linear/PCA regression to estimate the latent terms (Brouhns, Denuit and Vermunt 2002; Currie 2016; Lee and Carter 1992)

Evolution of neural mortality forecasting

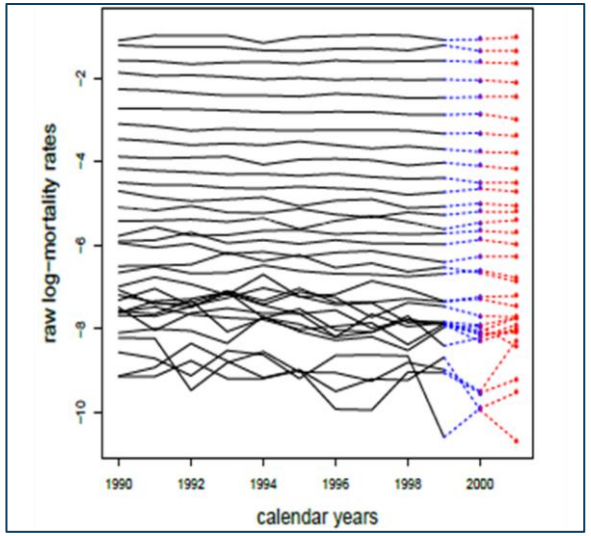
Forecast model coefficients
- Deep learning integrated Lee–Carter mode

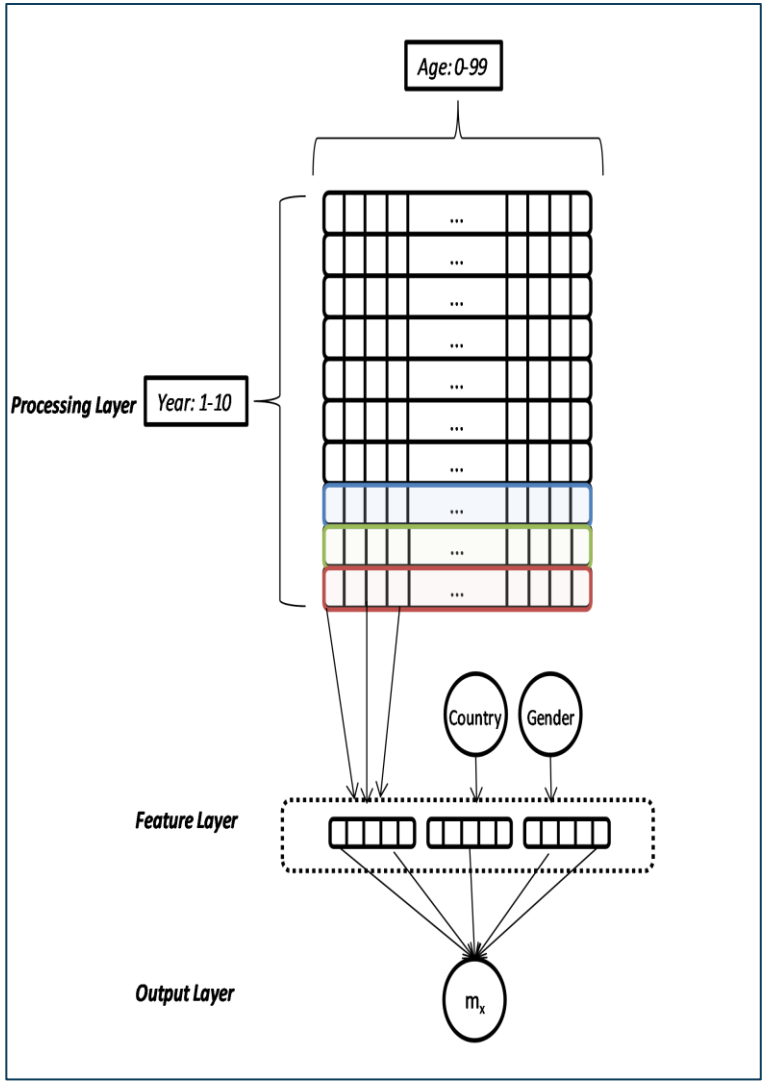
Forecast rates from tabular data
- Lee-Carter Neural Net

Forecast time series of rates
- Lee and Carter go Deep Learning

Forecast from "images" of rates
- Lee-Carter Convolutional Neural Network

Explainable forecast from "images" of rates
- LocalGLMnet





EXTENDING LC – TWO PERSPECTIVES

- Lee Carter model = regression model using features derived from data using PCA
- Perspective 1: Use a neural network to model the regression problem and let it decide on the feature set
  - LC is a linear model once parameters are known => use a NN to derive non-linear model
  - More advanced models explicitly specify the types of interaction between population-level and regional level parameters => use a NN to derive more predictive specification
- Perspective 2: use a more general step function formulation to specify the multi-population model
  - LC uses step functions of single dimension for each parameter:

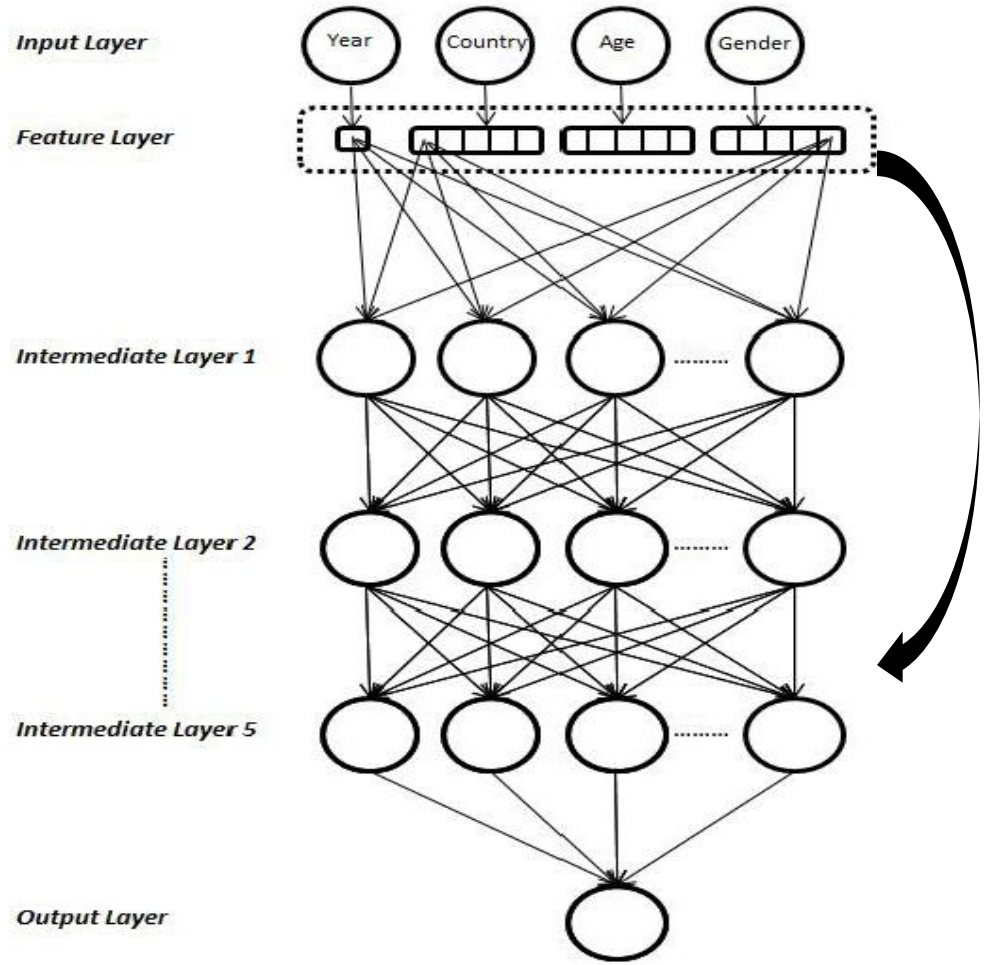
$$\log\left(u_{x,t}\right) = g(x) + h(x)i(t),$$

$$g(x) = \begin{cases} a_1 & \text{for } x = 1, \\ a_2 & \text{for } x = 2, \\ \vdots & \\ a_\omega & \text{for } x = \omega, \end{cases}$$

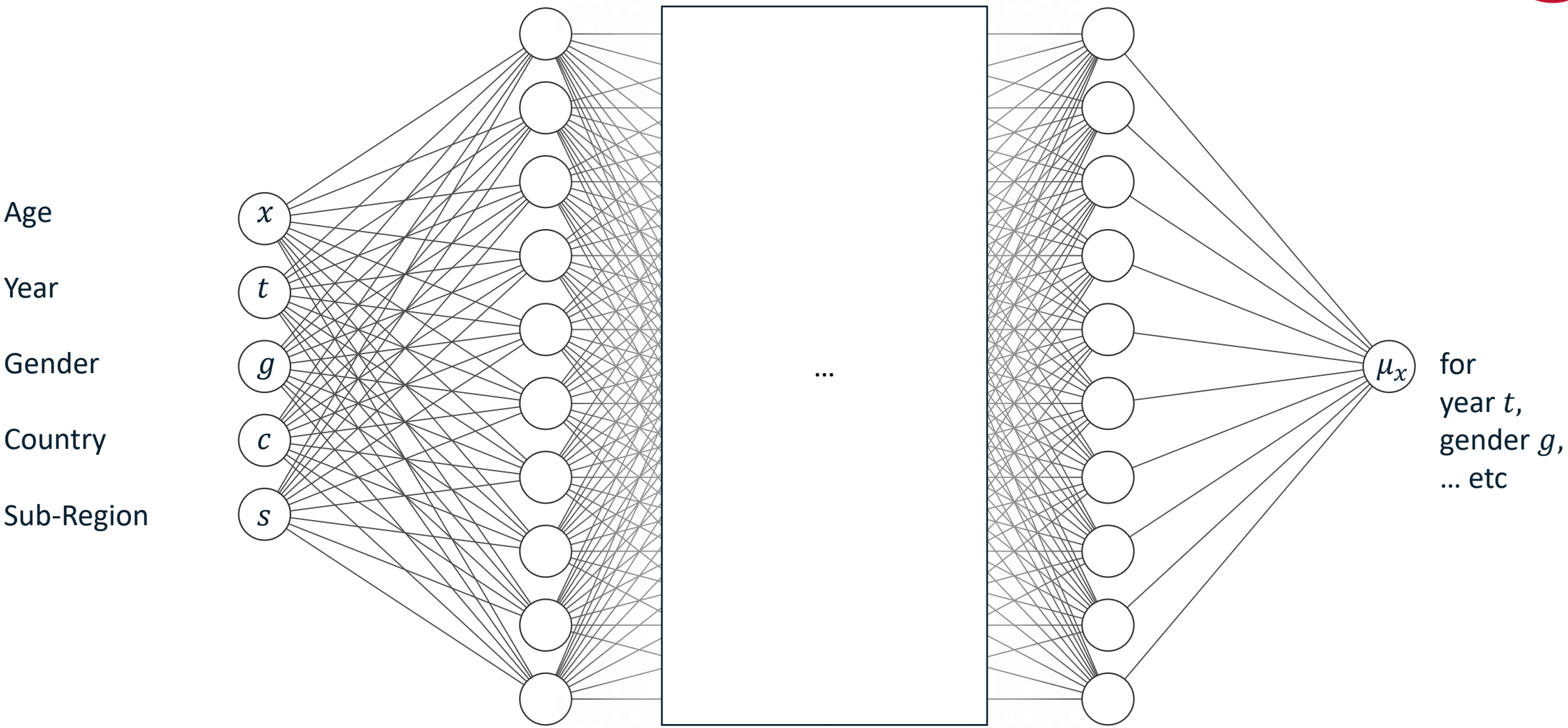
- Use NN to derive multi-dimensional vectors for each parameter using an embedding layer

EXAMPLE – LEE- CARTER NEURAL NET

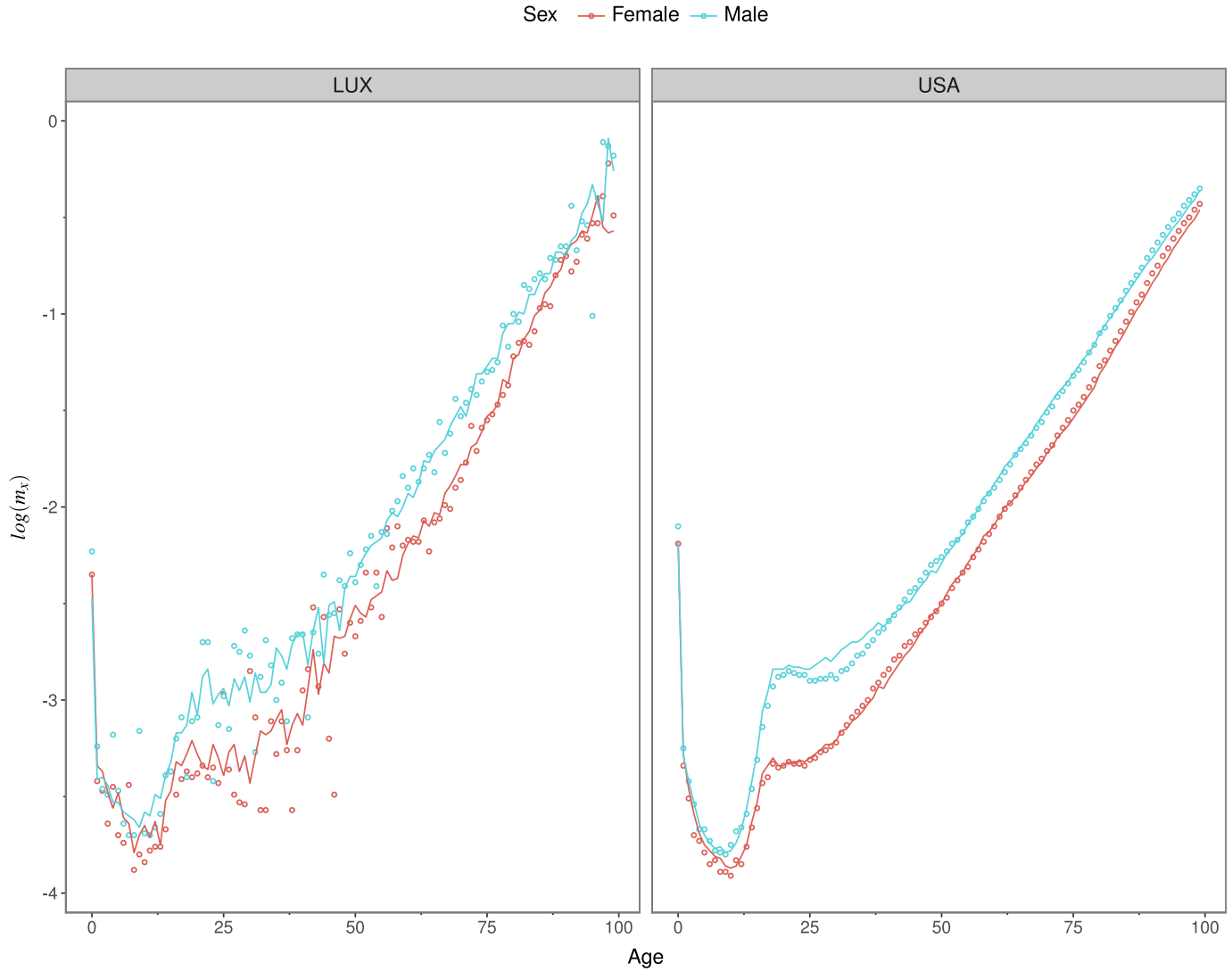
- Multi-population mortality forecasting model (Richman and Wüthrich 2018)
- Supervised regression on HMD data (inputs = Year, Country, Age; outputs = mx)
- 5 layer deep FCN
- Generalizes the LC model in both ways mentioned before
- Note that no time-series forecasting is done
  - Year enters the model as a numerical variable.
  - Forecasts made by predicting using values for Year beyond the range of input data



Lee-Carter Neural Network



Lee-Carter Neural Network

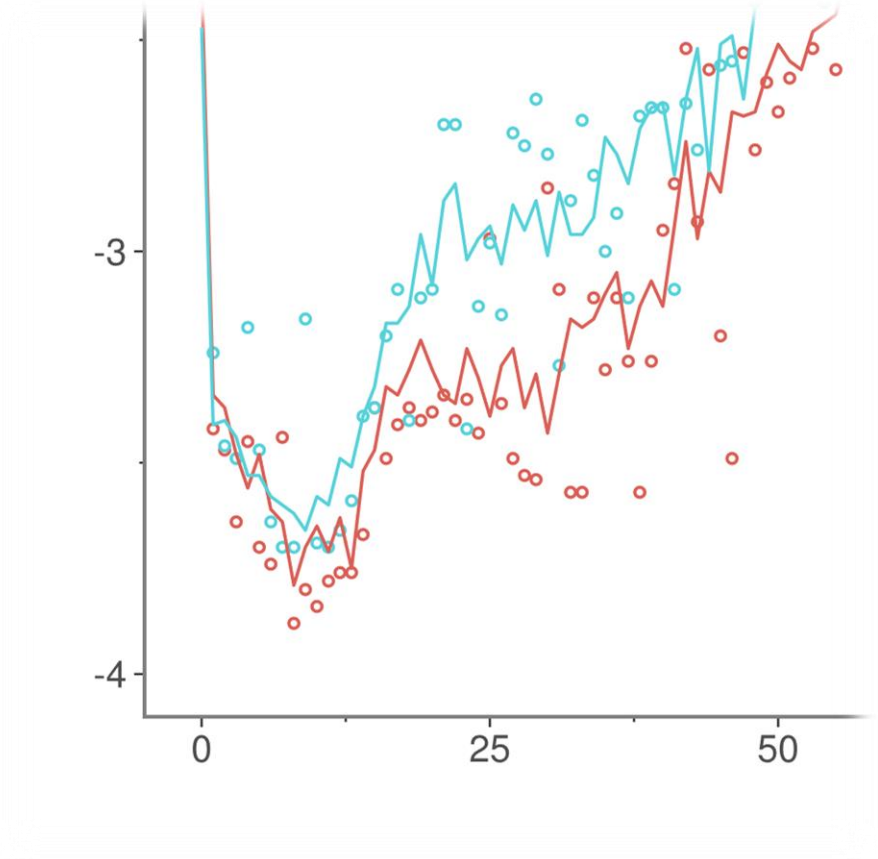


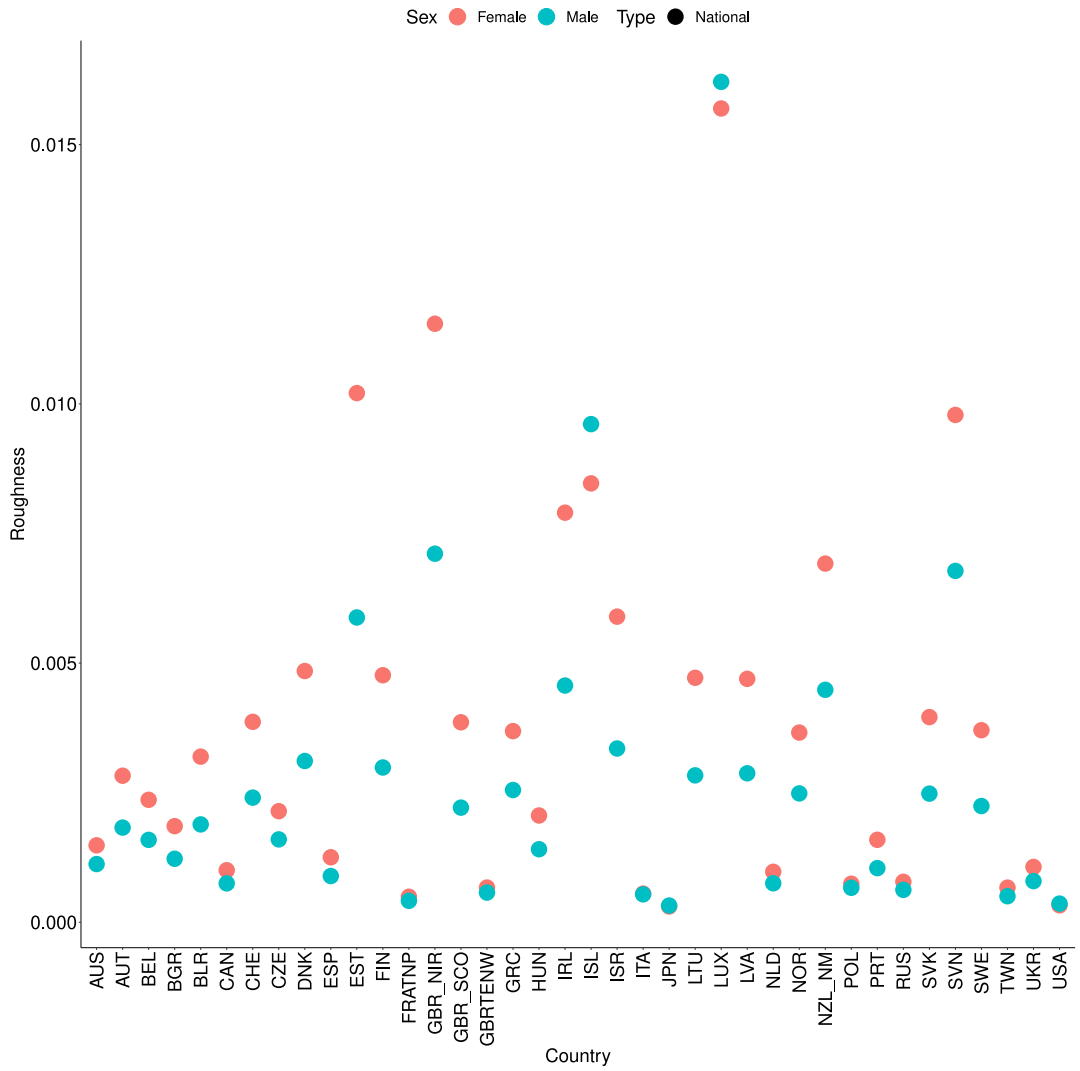
Lee-Carter Neural Network

Model	Type[1]	Average MSE	Median MSE
Lee-Carter (LC)	National	5.56	2.47
LC	Sub-national	22.10	1.48
LCNN	National	**3.05**	**1.53**
LCNN	Sub-national	**20.43**	**0.86**

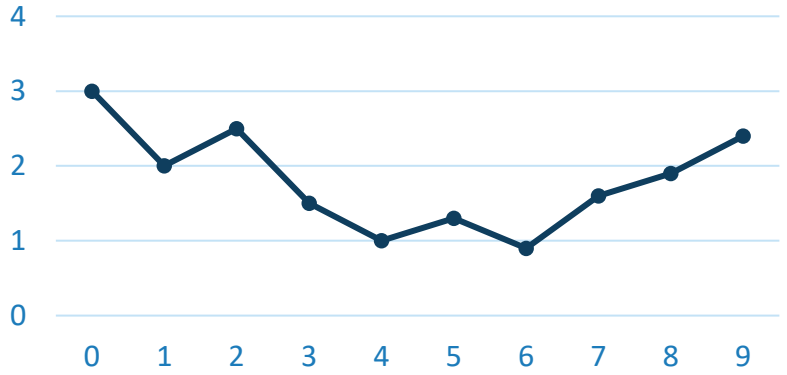
1. Refers to performance on national population (e.g. USA), or the sub-regions of the population (e.g. Arkansas).

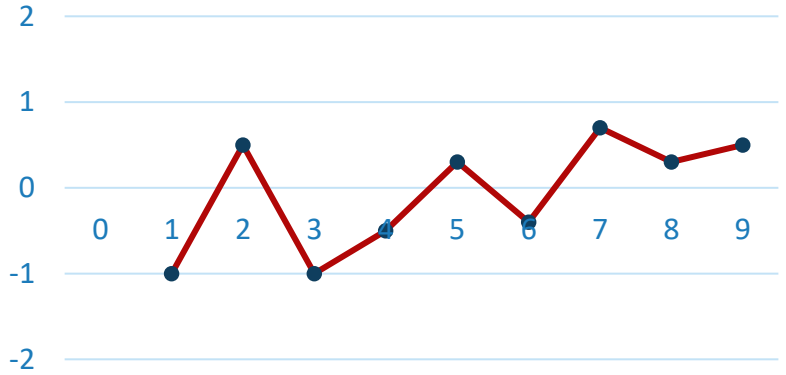
Roughness

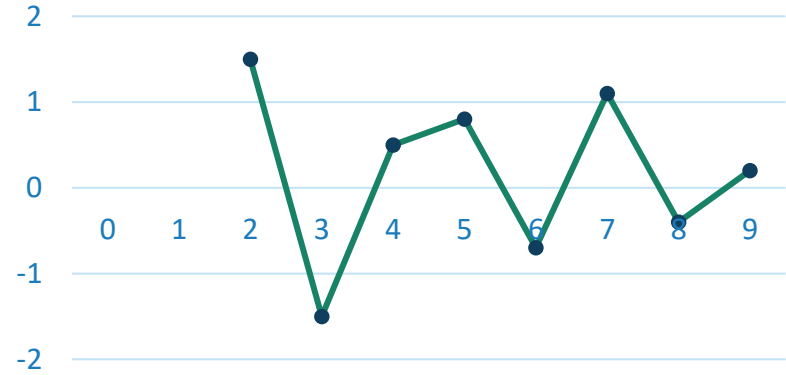


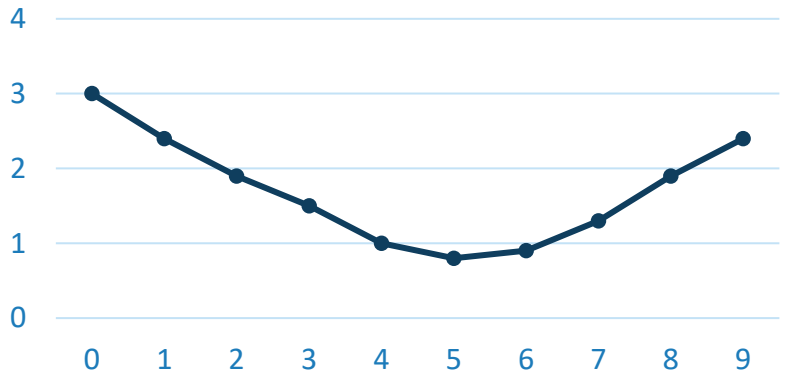


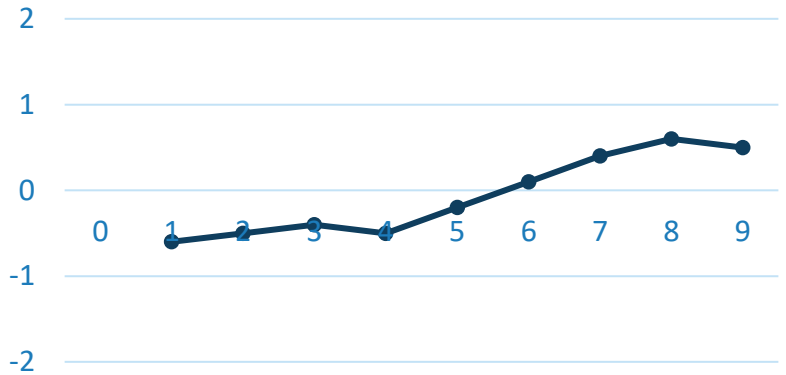
Roughness

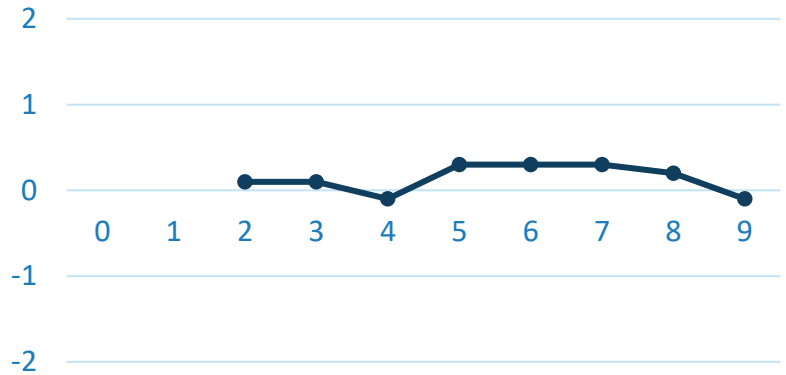












1st Difference

2nd Difference

Roughness

Given a vector of points, find a line through the vector that is **smooth**

Minimise: Error + Roughness

$$MSE\left(\vec{v}, \hat{\vec{v}}\right) + \alpha R(\vec{v})$$

α: Weighting added to focus on smoothness

Roughness

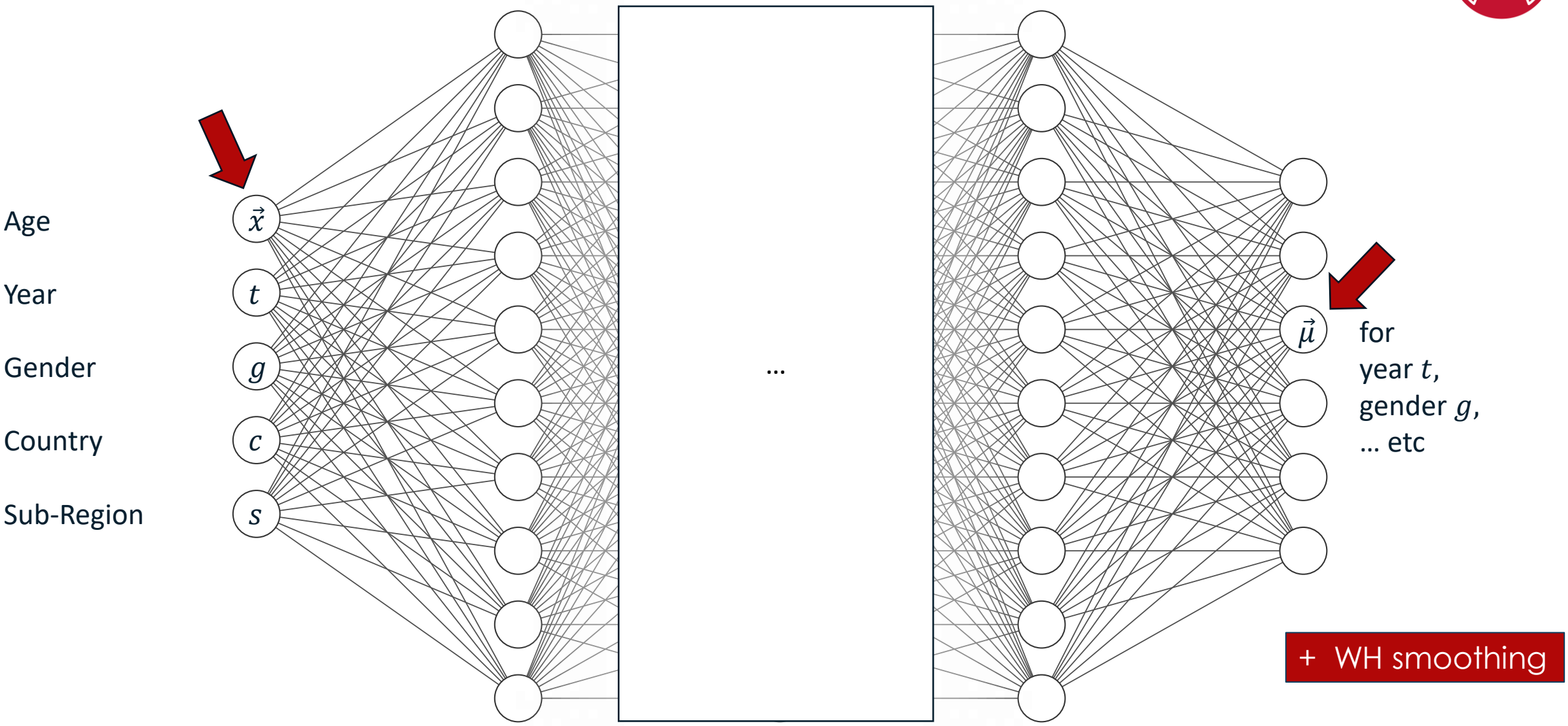
Roughness:

$$\Delta^1 v_x = v_x - v_{x-1}$$
$$\Delta^2 v_x = \Delta(v_x - v_{x-1})$$
$$\dots$$

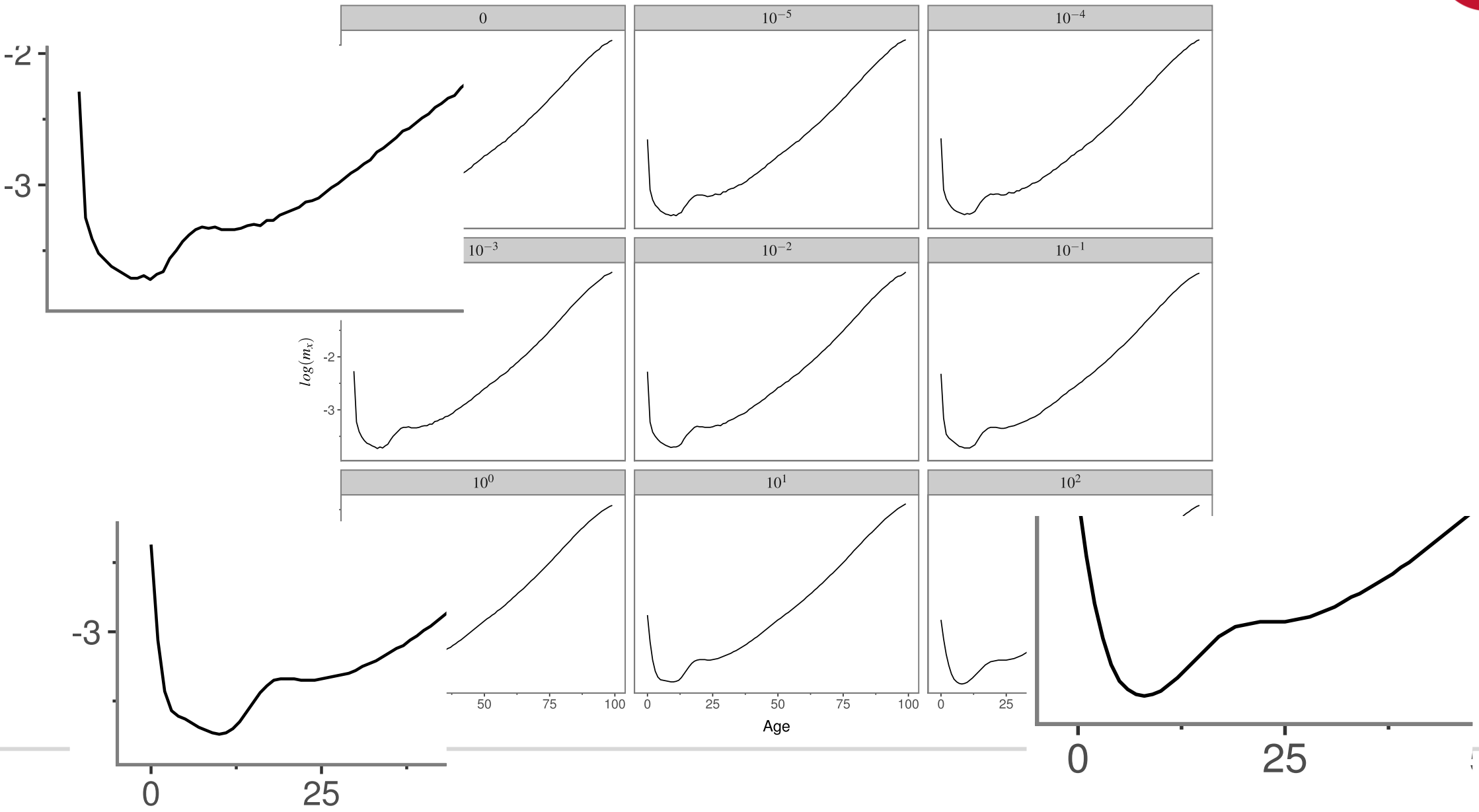
So…

$$L = MSE(\vec{v}, \hat{\vec{v}}) + \alpha \left\langle (\Delta^d v)^2 \right\rangle$$

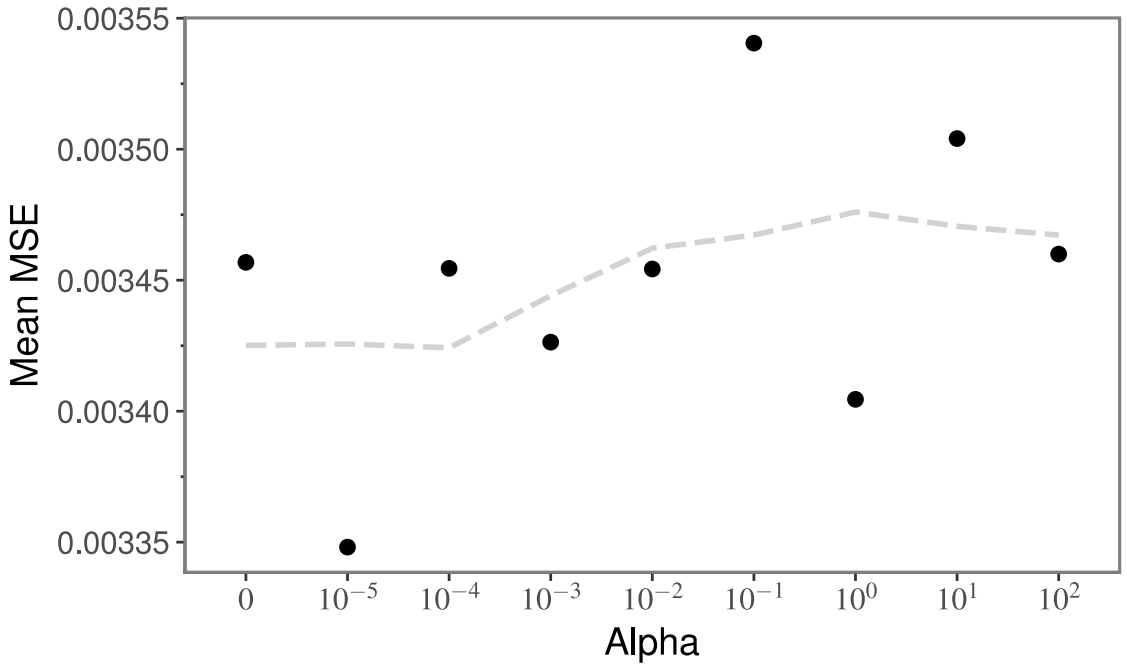
Vector Lee-Carter Neural Network

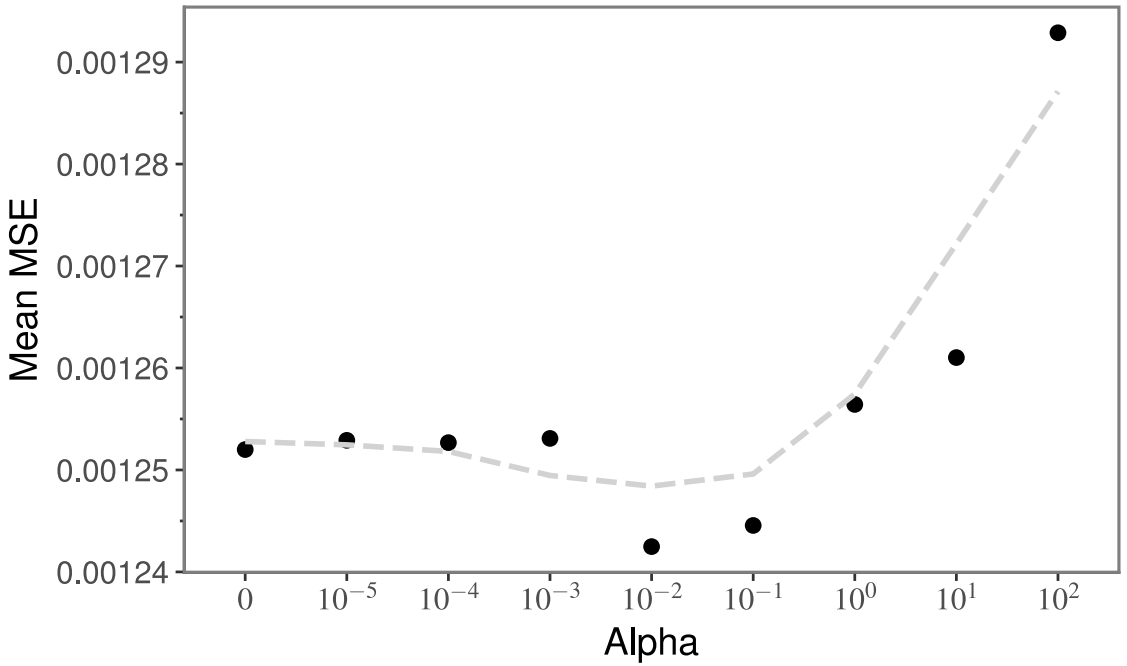


Vector Lee-Carter Neural Network

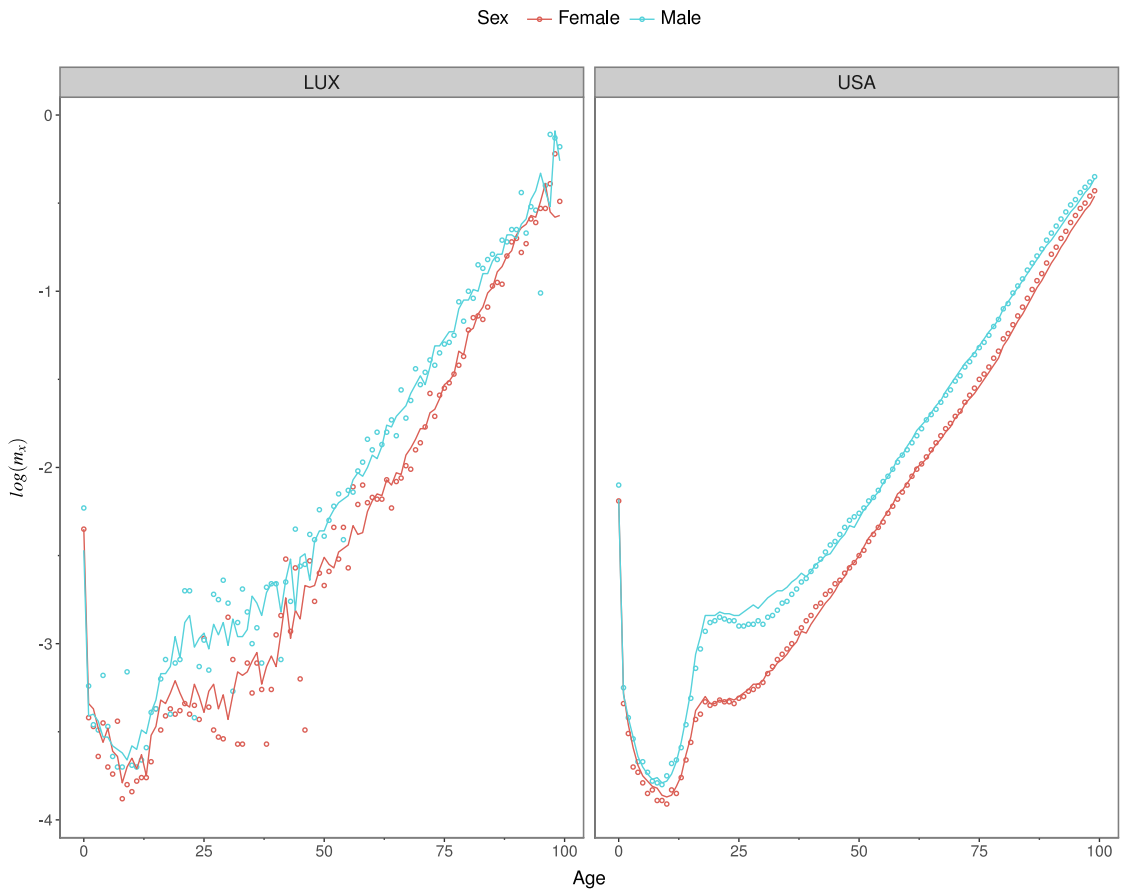


Vector Lee-Carter Neural Network

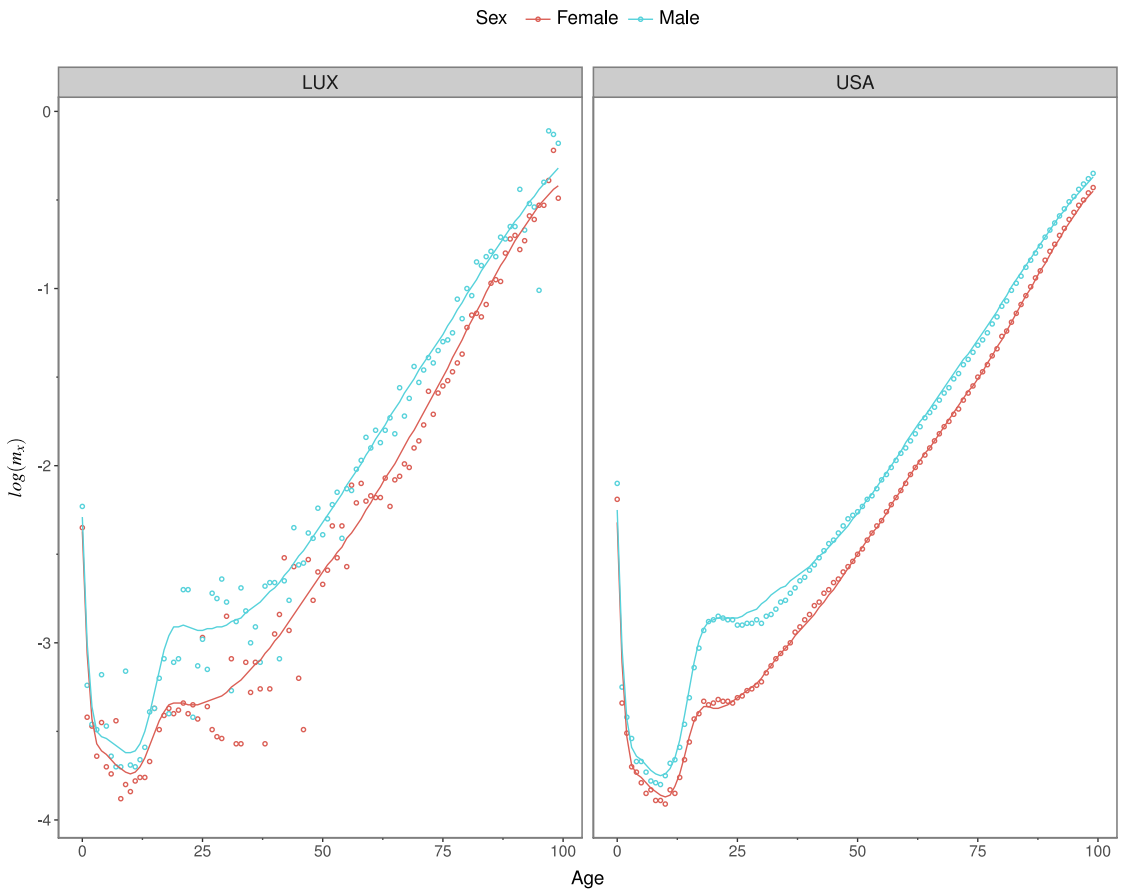




Vector Lee-Carter Neural Network



LCNN

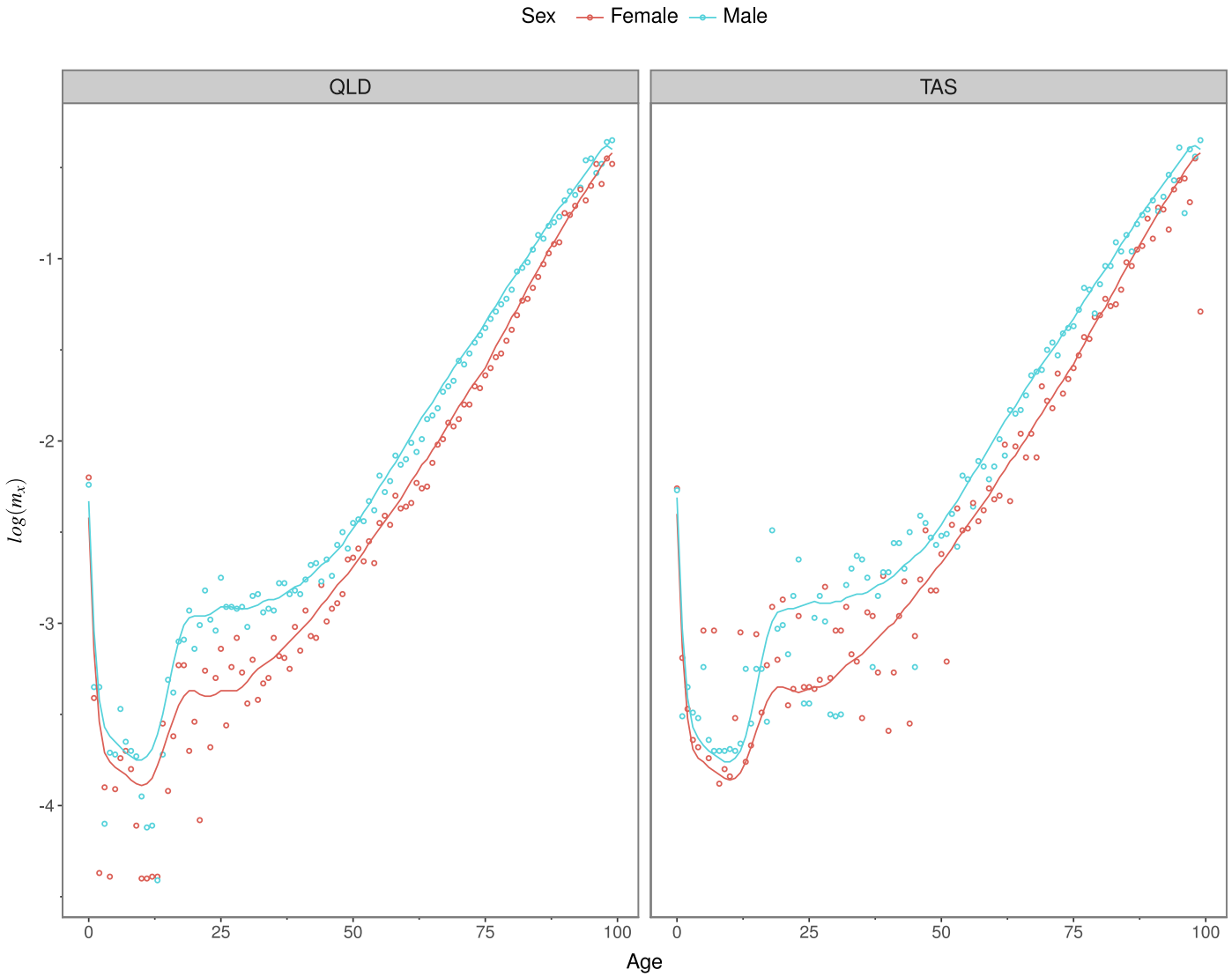


V-LCNN ($\alpha = 1$)

Vector Lee-Carter Neural Network

Model	Type[1]	α	Average MSE	Median MSE
Lee-Carter (LC)	National		5.56	2.47
LC	Sub-national		22.10	1.48
LCNN	National		3.05	**1.53**
LCNN	Sub-national		**20.43**	**0.86**
V-LCNN	National	1	**2.91**	1.80
V-LCNN	Sub-national	1	20.66	1.34

Transfer Learning

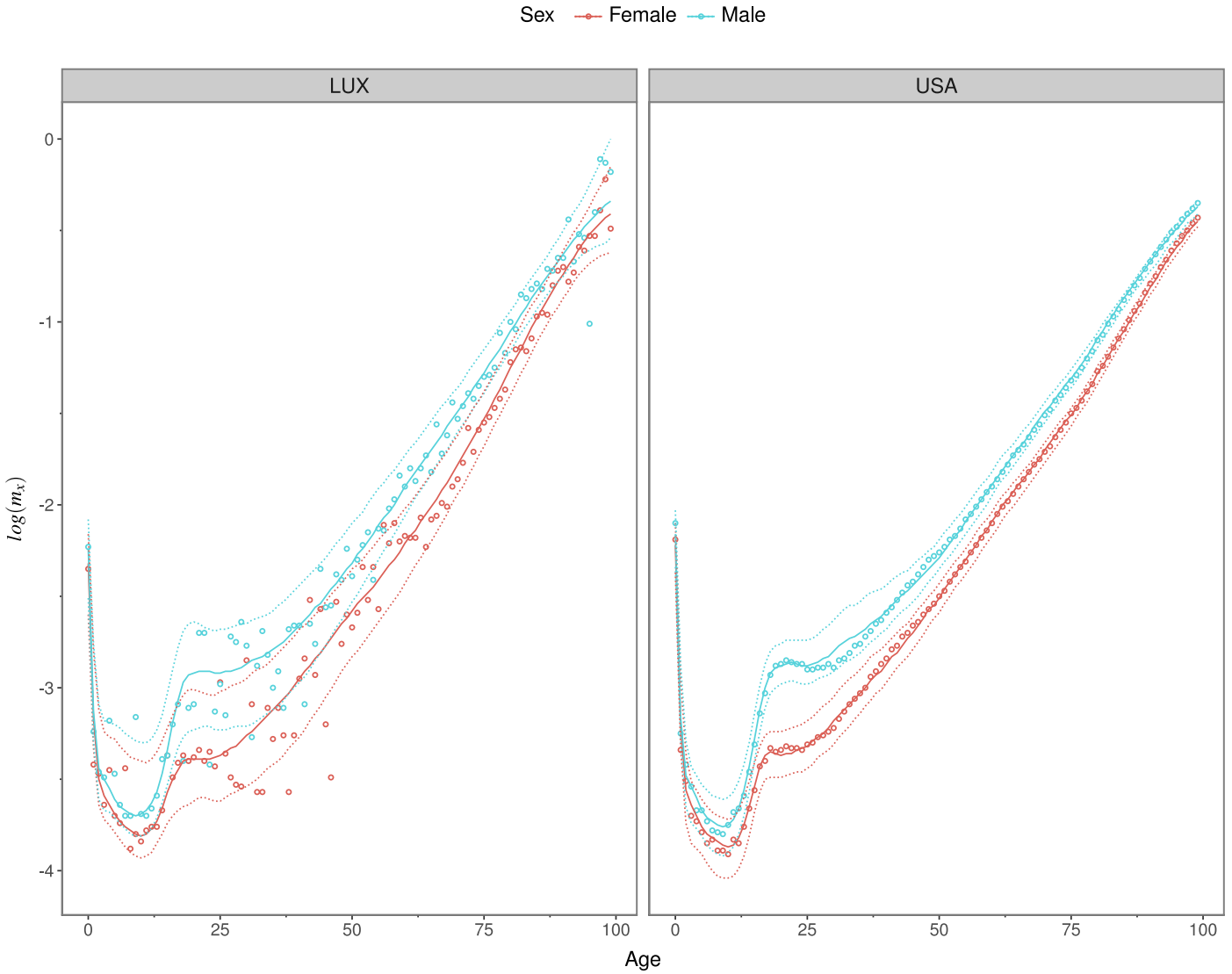


Transfer Learning

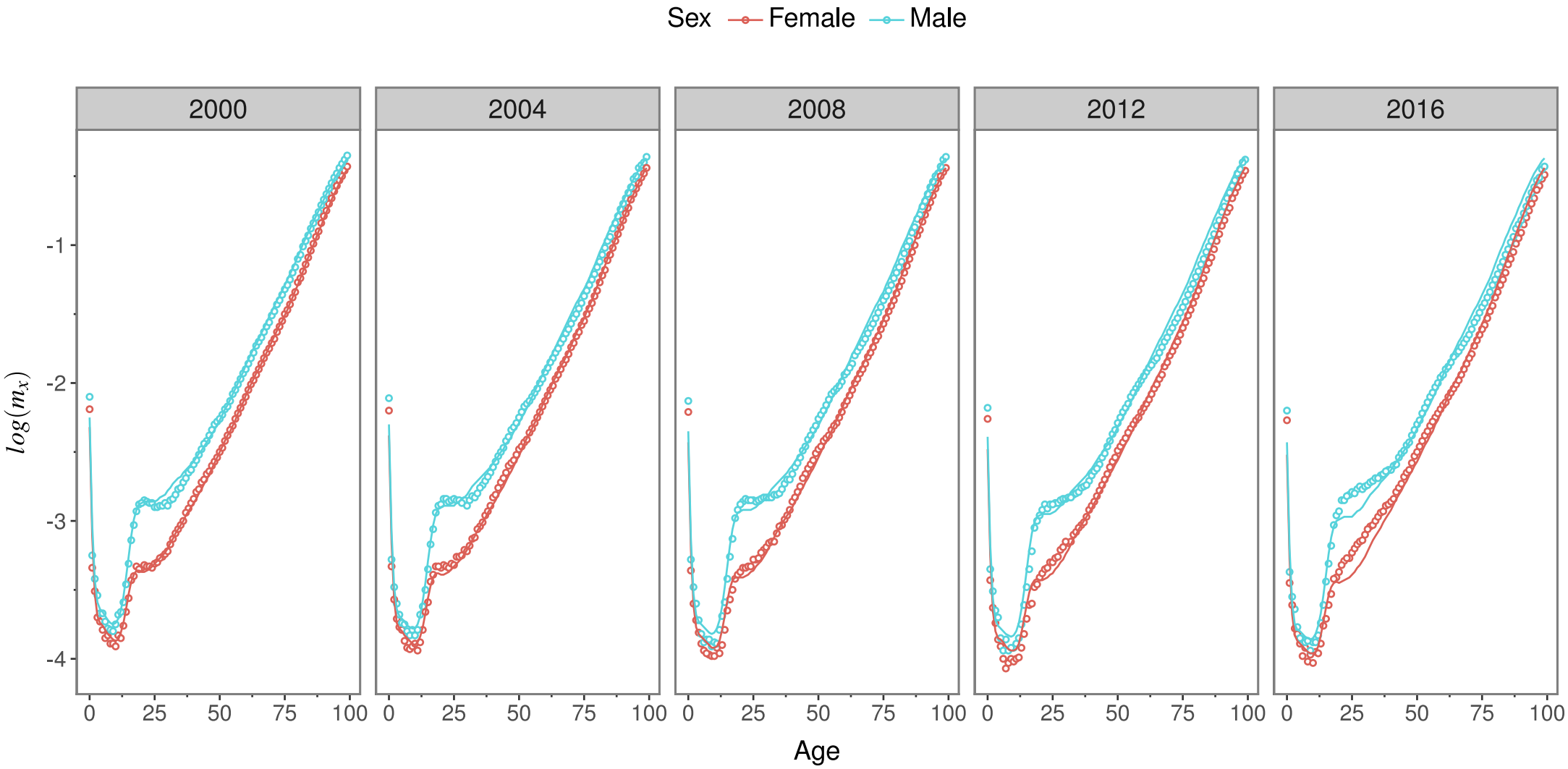
Model	Country	Sub-Region	Transfer	α	MSE
V-LCNN	Australia	QLD	False	1	1.62
V-LCNN	Australia	QLD	True	1	**1.61**
V-LCNN	Australia	TAS	False	1	11.52
V-LCNN	Australia	TAS	True	1	**10.62**

Transfer to unseen populations with improved performance

Quantile Regression



Forecasting



Thank You