A neural network analyzer for mortality forecast

D. Hainaut[1]

[1]ISBA, Université Catholique de Louvain

BNP PARIBAS CARDIFF 21-3-17

Motivation

- Improvement of longevity observed over the last century is a matter of concerns for the insurance industry.
- Reasons: reduction of mortality caused by infectious and chronic diseases at older ages.
- Standard in the industry model of Lee and Carter (LC) (1992) and its extensions:
  - Brouhns et al. (2002) : log-likelihood maximization
  - Renshaw and Haberman (2003) : multi-factor version
  - Renshaw and Haberman (2006) : cohort effect

Motivation

- Hypothesis: a linear dynamic between log-forces of mortality and latent time processes.
- This hypothesis allows to use a Principal Component analysis to estimate parameters.
- PCA : extraction method that attempts to characterize lower-dimensional structure in large multivariate datasets
- Alternative to PCA:
  - Kramer (1991)/ neural network based generalization of PCA to the nonlinear feature extraction problem (NLPCA). Field of application: chemical engineering (Dong and McAvoy 1996), psychology (Fotheringhame and Baddeley 1997), climatic mathematics (Monahan, 2000).
  - Hastie and Stuetzle (1989): principal curves and surfaces (PCS). Equivalent to NLPCA.

Contributions

1) One of the first article that applies the NLPCA to reduce the dimension of the term structure of mortality rates. Sparse literature on this field (Atsalakis et al. (2007) , Garko (2015) , Puddu and Menotti (2009) . Previous attempt use a **neural net to model and forecast a unique time serie** (not the term structure).

2) Calibration of the network with a **genetic algorithm** (McNelis (2005)). Given that the number of parameters, gradient descents, are inefficient and the risk of finding a local minimum is important.

3) Numerical illustration with French mortality rates from 1946 to 2014: **the Neural Network (NN) approach significantly outperforms the LC model**.

Benchmark: the LC model

- The Lee-Carter model is used as benchmark to evaluate the NN approach. The death probability:

$$q(t,x) \approx 1 - \exp\left(-\mu(t,x)\right) .$$

- In multi-factor LC models, the log-force of mortality is a linear combination of d time latent factors $\kappa_t^{i=1,\dots,d}$, with covariates that depend on the age:

$$\ln \mu(t,x) = \alpha_x + \sum_{i=1}^{d} \beta_x^i \kappa_t^i . \tag{1}$$

$\alpha_x \in \mathbb{R}^{x_{max}}$: permanent impact of age. $\beta_x^{i=1\dots d} \in \mathbb{R}^{x_{max}}$: marginal effect of latent factors on mortality at each age.

- For identifiability, we impose:

$$\sum_x \beta_x^i = 1 \qquad \sum_t \kappa_t^i = 0 \qquad i = 1, \dots, d . \tag{2}$$

Benchmark: the LC model

- Dataset of mortality forces $(\mu(t,x))$ range from year t_{min} to t_{max} and from age x_{min} to x_{max}.
- First stage, from the constraint $\sum_{t=t_{min}}^{t_{max}} \kappa_t^i = 0$,

$$\alpha_x = \frac{1}{t_{max} - t_{min} + 1} \sum_{t=t_{min}}^{t_{max}} \ln \mu(t,x) \qquad x = x_{min}, \dots, x_{max}.$$

- The next step consists to perform a PCA on adjusted observations denoted by X:

$$X := (\ln \mu(t,x) - \alpha(t,x))_{t=t_{min}\dots t_{max},\, x=x_{min}\dots x_{max}}$$

to find the d-plet $\kappa_t := (\kappa_t^1, \kappa_t^2, ..., \kappa_t^d)$ and $\left(\beta_x^i\right)_{x=x_{min}\dots x_{max}}$.

Benchmark: the LC model

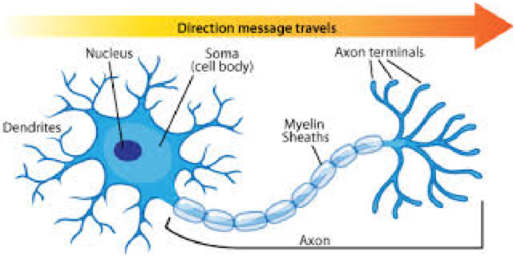
- Last stage : a time series model is specified for the latent processes.
- Most of authors use an AR(1) model or a random walk with drift.
- In this paper, we opt for the second choice and assume that increments of κ_t^i are Gaussian random variables with a mean γ_i and a variance σ_i^2:

$$\kappa_t^i - \kappa_{t-1}^i \quad = \quad \gamma_i + \sigma_i \, \epsilon_t \quad i = 1, ..., d \tag{3}$$

where ϵ_t is a standard normal random variable.

The neural net analyzer

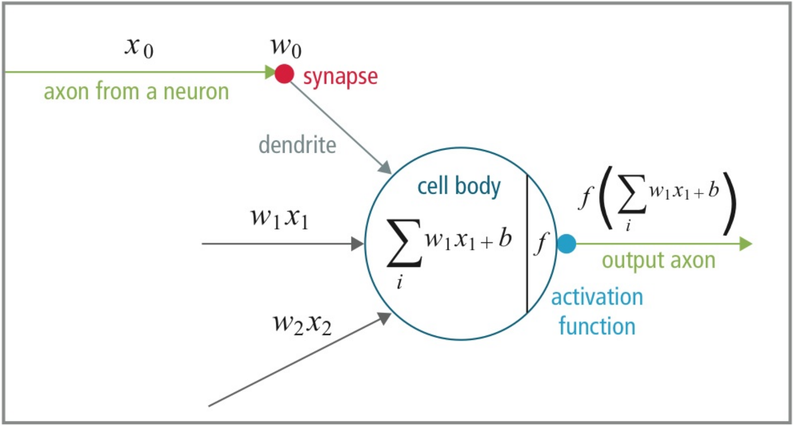
- Most neurons receive input signals throughout their dendritic trees (sometimes many thousands of input signals).
- Whether or not a neuron is excited into firing an impulse depends on the sum of all of the excitatory and inhibitory signals.



- If the neuron does end up firing, the nerve impulse, or action potential, is conducted down the axon.

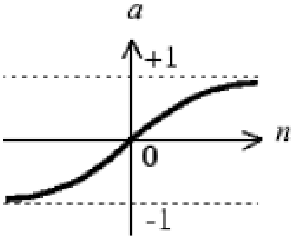
The neural net analyzer

The signals (e.g.,x_0) interact multiplicatively with the dendrites of the other neuron based through synaptic weights (e.g., $w_0 x_0$). The dendrites carry input signals to the cell body, where they all are summed. If the final sum is above a threshold, the neuron fires, sending a spike along its axon.



The neural net analyzer

The activation function f is a sigmoid:



$$a = tansig(n) = \frac{2}{(1 + e^{-2n})} - 1$$

The neural net analyzer

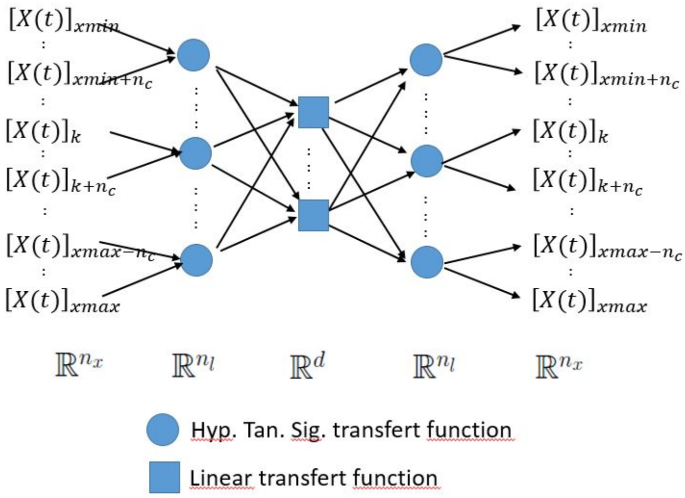
- Input signals = centered log-forces of mortality:

$$X(t) := \begin{pmatrix} \ln \mu(t, x_{min}) - \alpha_{x_{min}} \\ \vdots \\ \ln \mu(t, x) - \alpha_x \\ \vdots \\ \ln \mu(t, x_{max}) - \alpha_{x_{max}} \end{pmatrix} \quad t = t_{min}, ..., t_{max} .$$

- We aim to calibrate the **weights** $(w_i)_{i=...l}$ defining an encoding and a decoding functions: $f^{enc} : \mathbb{R}^{n_x} \to \mathbb{R}^d$ and $f^{dec} : \mathbb{R}^d \to \mathbb{R}^{n_x}$.

The neural net analyzer

- Architecture of the neural net:



The neural net analyzer

- The encoding function, $f^{enc}(.)$, projects curves of mortality rates at time t, $X(t) \in \mathbb{R}^{n_x}$ in $\mathbb{R}^d$. As in the LC model, projections in $\mathbb{R}^d$ are in a d-plet $\kappa_t^{nn} := (\kappa_t^{nn,1}, ..., \kappa_t^{nn,d})$:

$$\kappa_t^{nn} := f^{enc}(X(t)) \quad t = t_{min}, ..., t_{max} \,.$$

- The decoding function $f^{dec}(.)$ uses κ_t^{nn} to build an approximation $\hat{X}(t) \in \mathbb{R}^{n_x}$ of the initial curve of log-mortality rates:

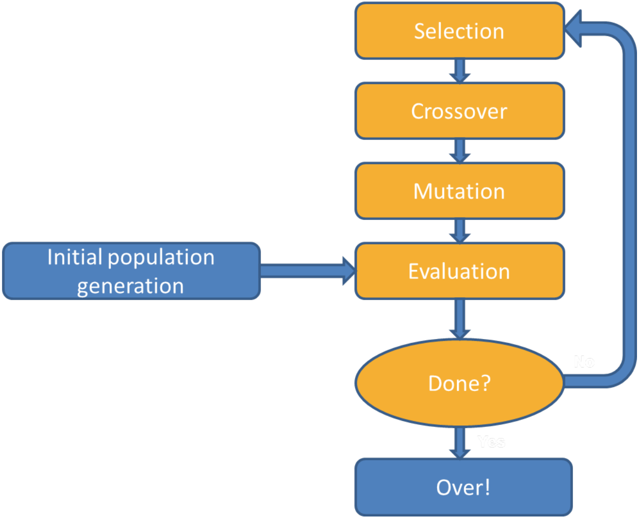
$$\hat{X}(t) \quad := \quad f^{dec}\left(\kappa_t^{nn}\right) \,.$$

- Compared to the original LC model, the linear relation $\beta_x \kappa_t$ is replaced by a non-linear function:

$$\ln \mu(t,x) = \alpha_x + f^{dec}\left(x, \kappa_t^{nn}\right) \tag{4}$$

- Calibration of weights by minimization of the quad. error: $\arg\min_{weights} \sum_{t=t_{min}}^{t_{max}} \left\| X(t) - \hat{X}(t) \right\|_2^2$.

The neural net analyzer



The neural net analyzer

- Create a population of candidate parameters (weights that multiply input signals) (100 individuals).
- **Selection**: choose randomly two pairs of individuals from the population. The two winning individuals P_1, $P_2 \in \Omega$ are retained for "breeding" purposes.
- **Crossover** of P_1 and P_2 (3 types: Shuffle, Arithmetic ,Single-point), 2 children: C_1 and C_2 .
- **Mutation** of children. With a decreasing probability,

$$\tilde{C}_{i,k} = \begin{cases} C_{i,k} + s\left(1 - r_2^{(1-\frac{G}{G*})^b}\right) & \textit{if } r_1 > 0.5 \\ C_{i,k} - s\left(1 - r_2^{(1-\frac{G}{G*})^b}\right) & \textit{if } r_1 \leq 0.5 \end{cases}$$

  where r_1 and r_2 are r.v. on [0, 1] and s is a standard normal r.v.
- The "family" (P_1, P_2, C_1, C_2) engage in a **tournament**. The two vectors with the best goodness of fit, whether parents or children, survive and pass to the next generation.

Application

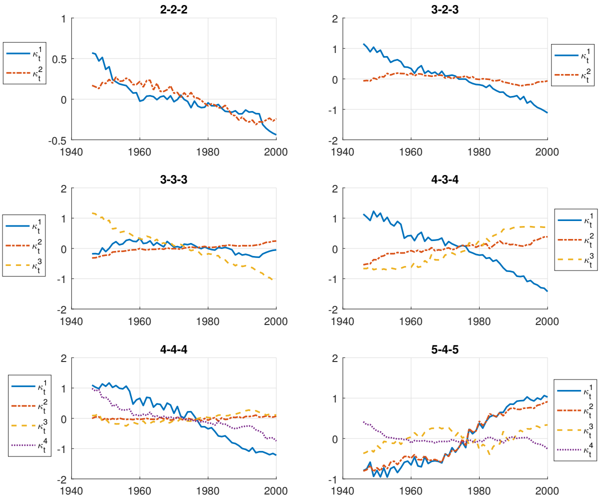
- We work with raw mortality rates observed for the French population over the period 1946 to 2014 (HMD.org).
- Calibration: from 1946 up to 2000. To compare the predictive capability, projection by simulation over fourteen years (10 000 simulations)
- Comparison with mortality over the period 2001-2014.

Application

Lee Carter			Neural Net			
Dim.	Coef.	Avg. $\left\|X - \hat{X}\right\|_2$	n_l	n_d	Coef.	Avg. $\left\|X - \hat{X}\right\|_2$
1	180	0.0024	2	2	368	0.0010
2	270	0.0023	**3**	**2**	**55**	**0.0009**
3	360	0.0023	3	3	558	0.0009
4	450	0.0023	4	3	744	0.0008
5	540	0.0023	4	4	752	0.0008
6	630	0.0023	5	4	940	0.0008

French population, 1946-2000.

Application



Application

Jarque Bera test applied to increments of latent factors over the period 1970-2000, for neural analyzer with three neurons in input / output layers and two neurons in the intermediate layer.

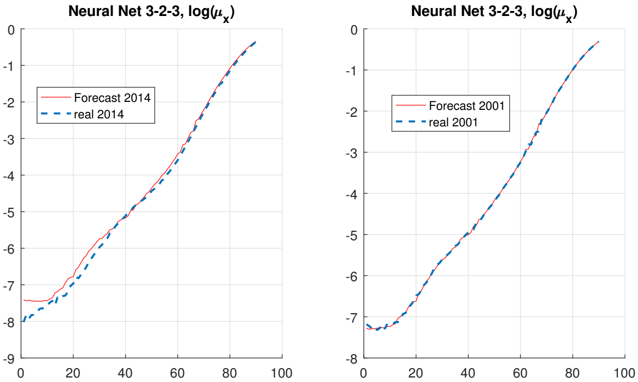
		Jarque Bera statistics for a 3-2-3 Neural Net		
factors	Normality	p-value	JB statistic	Critical Value 5%
$\kappa_t^{nn,1} - \kappa_{t-1}^{nn,1}$	Accept	0.5000	0.0762	4.4496
$\kappa_t^{nn,2} - \kappa_{t-1}^{nn,2}$	Accept	0.3698	1.2296	4.4496

Application

Lee Carter			Neural Net			
Dim.	Coef.	Avg. $\left\| X - \hat{X} \right\|_2$	n_l	n_d	Coef.	Avg. $\left\| X - \hat{X} \right\|_2$
1	180	0.0049	2	2	368	0.0030
2	270	0.0049	**3**	**2**	**55**	**0.0032**
3	360	0.0049	3	3	558	0.0034
4	450	0.0049	4	3	744	0.0035
5	540	0.0049	4	4	752	0.0038
6	630	0.0049	5	4	940	0.0042

French population, 2001-2014, forecast

Application

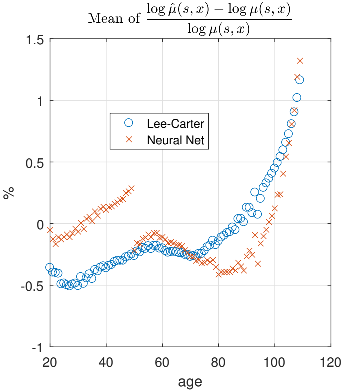


Comparison of real log-mortality rates in 2014/2001 to the average log-forces of mortality simulated with the 3-2-3 neural analyzer

Application

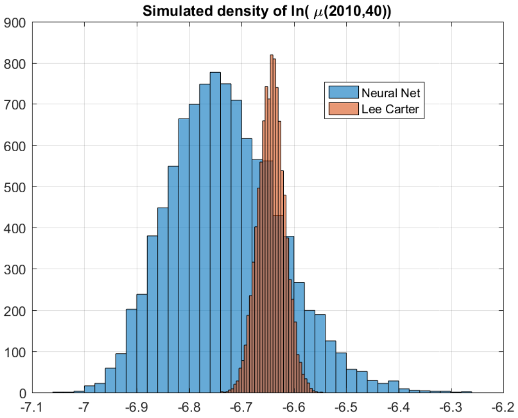
	$e_{20}^{LC}(t)$	$e_{40}^{LC}(t)$	$e_{60}^{LC}(t)$
2014	60.919	41.762	24.077
	$e_{20}^{Obs}(t) - e_{20}^{LC}(t)$	$e_{40}^{Obs}(t) - e_{40}^{LC}(t)$	$e_{60}^{Obs}(t) - e_{60}^{LC}(t)$
2014	2.4362	2.1789	1.7942
	$e_{20}^{NN}(t)$	$e_{40}^{NN}(t)$	$e_{60}^{NN}(t)$
2014	62.231	42.992	25.163
	$e_{20}^{Obs}(t) - e_{20}^{NN}(t)$	$e_{40}^{Obs}(t) - e_{40}^{NN}(t)$	$e_{60}^{Obs}(t) - e_{60}^{NN}(t)$
2014	1.1243	0.9485	0.7076

Application



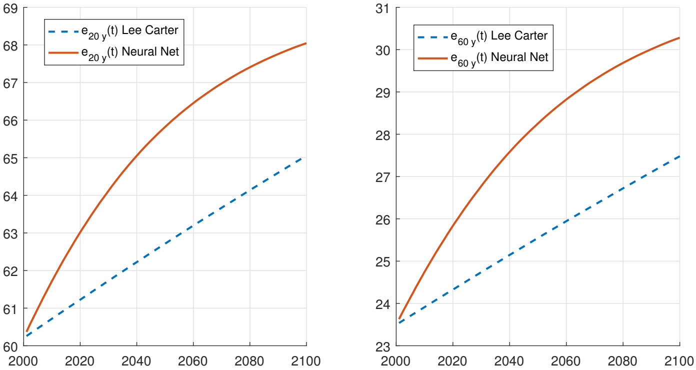
Breeakdown of means and standard deviations of relative average errors of calibration, per age. We compare the 3 Dimensions Lee Carter model to the 3-2-3 Neural network.

Application



Comparison of simulated densities for $\ln \mu(2010, 40)$, yield by the LC and Neural models.

Application



Predicted cross-sectional lifetime expectancies with the LC and neural models, over the period 2001-2100.

Application

	$e_{20}^{NN}(t)$	$e_{40}^{NN}(t)$	$e_{60}^{NN}(t)$	$e_{80}^{NN}(t)$
2001	60.36	41.23	23.62	8.982
2025	63.59	44.29	26.32	10.76
2050	65.81	46.42	28.25	12.25
2100	**68.04**	**48.58**	**30.28**	**14.08**
	$e_{20}^{LC}(t)$	$e_{40}^{LC}(t)$	$e_{60}^{LC}(t)$	$e_{80}^{LC}(t)$
2001	60.25	41.12	23.53	8.925
2025	61.47	42.29	24.53	9.485
2050	62.71	43.48	25.54	10.06
2100	**65.05**	**45.73**	**27.48**	**11.20**

Long term predictions of cross-sectional life expectancies, computed by simulations with the LC and Neural models, over the period 2001-2100.

Conclusion

- Totally new approach based on a neural network to predict and simulate the future human mortality.
- The neural analyzer outperforms the LC model and its multi-factor extensions. The average error calibration error is three times lower than the error for the 3 dimension LC model.
- Excellent predictive power of the neural approach. However, the number of neurons must be chosen carefully to avoid over-parameterization.
- On the other hand, the longevity risk predicted by the neural analyzer is three times bigger than the one forecast by the LC model.
- Neural networks are promising for other applications in actuarial sciences (loss reserving)

References

Abdulkarim SA, Garko AB, 2015 Forecasting maternal mortality rate using particle Swarm optimization based artificial neural network. Dutse Journal of Pure and Applied Sciences 1(1): 55 - 59.

Antonio K., Bardoutsos A., Ouburg W. 2015. Bayesian Poisson log-bilinear models for mortality projections with multiple populations. European Actuarial Journal 5:245–281

Atsalakis G, Nezis D, Matalliotakis G, Ucenic CI, Skiadas C, 2007. Forecasting Mortality Rate Using a Neural Network with Fuzzy Inference System. Working paper 0806. University of Crete.

Brouhns N, Denuit M, Vermunt JK 2002. A Poisson log-bilinear regression approach to the construction of projected lifetables. Insurance: Mathematics and Economics 31(3):373-393.

Dong D, McAvoy TJ, 1996. Nonlinear principal component analysis—Based on principal curves and neural networks. Comp. Chem. Eng., 20, 65–78.

Fotheringhame D, Baddeley R, 1997. Nonlinear principal components analysis of neuronal spike train data. Biol. Cybernetics, 77, 282–288.

Khachaturyan A, Semenovskaya S, Vainshtein B, 1979. "Statistical-Thermodynamic Approach to Determination of Structure Amplitude Phases". Sov.Phys. Crystallography. 24 (5): 519–524.

Kramer MA, 1991. Nonlinear principal component analysis using autoassociative neural networks. AIChE J., 37, 233–243.

References

Lee RD , Carter L, 1992. Modelling and forecasting the time series of US mortality. Journal of the American Statistical Association, 87 659-671.
McNelis PD, 2005. Neural networks in finance: gaining predictive edge in the market. Elsevier academic press.
Puddu PE, Menotti A, 2009. Artificial neural network versus multiple logistic function to predict 25-year coronary heart disease mortality in the Seven Countries. European Journal of Cardiovascular.16(5):583-91.
Renshaw AE, Haberman S, 2003. Lee-Carter mortality forecasting with age-specific enhancement. Insurance: Mathematics and Economics, 33 p255-272
Renshaw A, Haberman S, 2006. A cohort-based extension to the Lee–Carter model for mortality reduction factors. Insurance: Mathematics and Economics 38:556–570.