Corvinus University of Budapest

Doctoral School of Economics and Business Informatics

Innovations in Actuarial Mortality Forecasting

Thesis Booklet

Szentkereszti Gábor

Supervisor: Vékás Péter, Ph.D.

Budapest, 2026

I. Research background and justification of the topic

1.1 Mortality in historical perspective

To describe the improvement in overall mortality, Omran (1971) introduced the concept of epidemiological transition. According to him, the process of mortality improvement can be divided into three main stages. In the first stage, the age of pestilence and famine, mortality is characterized by cyclicality and greater fluctuations, with major determinants being epidemics, wars, and famine. In the next stage, the age of receding pandemics, infectious diseases begin to recede and famine ceases. Meanwhile, in the final stage, the age of degenerative and man-made diseases, infectious diseases are completely displaced, and degenerative diseases become the primary cause of death.

Analyzing European countries, around 1950 the northern and western countries had significantly better life expectancies than the southern and eastern countries. Over the next two decades, several Southern and Eastern European countries experienced partial convergence. A crisis then emerged in the states belonging to the Soviet Union, leading to stagnation in life expectancy growth and, in some countries, worsening mortality indicators among middle-aged male populations (Meslé and Vallin, 2002). Omran's original theory was subsequently extended to account for later developments, including the cardiovascular revolution. Vallin and Meslé (2004) developed a theory whose final stage is the fight against ageing and a further increase in life expectancy.

1.2 Relevance of mortality forecasting

From a macroeconomic perspective, the sustainability of pension systems has become an increasingly pressing issue in most countries. Uncertainty in mortality patterns complicates expenditure forecasting, which in turn makes the formulation of appropriate policy decisions highly challenging. In pay-as-you-go pension schemes, long-term planning requires expectations regarding the best estimates of mortality and population dynamics, as only then can expected revenues and expenditures be projected. The continuous increase in life expectancy creates a growing financial burden through prolonged benefit payments, while persistently low fertility places pressure on the contribution side.

The evolution of mortality rates across birth cohorts is also highly significant for the life insurance sector. Since both expected revenues and expected expenditures depend on mortality, accurate forecasts of future mortality rates are crucial for insurers. Static premium calculations assume that mortality probabilities remain constant over time. Due to continuous mortality improvement, this assumption is evidently flawed, particularly for annuities and endowment insurance. Dynamic premium calculation instead forecasts mortality probabilities for each age group and uses these projections in premium setting.

Accurate mortality assumptions are likewise essential for the valuation of life insurance contracts and realistic projections of portfolio profitability. Longevity risk, the risk arising from unexpectedly low mortality or unexpectedly high survival rates, is a pressing issue for state pension systems, insurance companies, and individuals. Solvency II quantifies longevity risk as the decrease in net asset value under a permanent 20% reduction in mortality rates. Longevity risk is therefore a significant and complex phenomenon affecting individuals, governments, private companies, and regulators. This makes it essential to employ mathematical and statistical methods to develop accurate expectations concerning the future evolution of mortality rates.

1.3 Research questions and hypotheses

This thesis aims to address the following research questions.

RQ1. Does using more complex time-series methods to forecast the Lee–Carter model's mortality index lead to improved forecast accuracy? Which method provides the most reliable forecasts for actuaries and demographers?

Hypothesis 1: Forecast accuracy can be improved by using spatio-temporal econometric models, which utilize not only temporal information but also spatial information contained in the data.

RQ2. To what extent can forecasts of Visegrád Group countries' mortality rates be improved by using neural networks compared with classical methods? Do neural networks provide more precise forecasts across all Visegrád Group populations? Are the forecast results significantly better?

Hypothesis 2: Mortality-rate forecasts for Visegrád Group countries can be significantly improved using neural networks.

RQ3. Can neural networks produce coherent mortality forecasts? If so, under what conditions?

Hypothesis 3: The coherence of mortality forecasts generated by a neural network depends on identifiable restrictions on its architecture and activation functions.

II. Methods used

2.1 Classical mortality models

The Lee–Carter model, introduced by Lee and Carter (1992), is one of the most widely adopted stochastic models for forecasting age- and time-specific mortality rates. Its log mortality rates are given by

$$\log(m_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t},$$

where a_x and b_x are age-specific parameters, k_t is a time-specific mortality index, and $\varepsilon_{x,t}$ is the residual term. The mortality index is most commonly forecast using a random walk with drift. To overcome the incoherence of forecasts obtained by applying Lee–Carter models separately to related populations, Li and Lee (2005) proposed a coherent multi-population extension with a common mortality factor and population-specific factors.

2.2 Spatial and panel econometric forecasting

The vast majority of researchers, actuaries, and demographers use standard time-series techniques to project time-varying parameters of the Lee–Carter and Li–Lee models. However, spatial dependence can be as significant as temporal autocorrelation, while the panel structure of the data is often neglected. The first study draws on ordinary and spatial dynamic panel linear models, spatio-temporal autoregressive integrated moving-average processes, and spatial eigenvector filters. It also develops a robust selection framework for identifying the best model–technique combinations by country and a bootstrap-based procedure for quantifying projection uncertainty.

The techniques applied to the mortality indices are Random Walk with Drift (RWD), Autoregressive Integrated Moving Average (ARIMA), Dynamic Panel Linear Model (DPLM), Vector Autoregression (VAR), Spatio-Temporal ARIMA (STARIMA), Spatial Dynamic Panel Linear Model (SDPLM), and Eigenvector Spatio-Temporal Filter (ESTF). Moran's I is used to quantify global spatial and spatio-temporal autocorrelation, while Local Indicators of Spatial Association identify spatial clusters and outliers.

The models are tested on mortality data for 22 European countries, spanning 1960–2019 and ages 0–99. The first 45 years, up to 2004, are used as training data by default; the remaining 15 years form the test data, divided into an eight-year validation period and a seven-year evaluation period. Parameters and hyperparameters are estimated separately for the Lee–Carter and Li–Lee frameworks, and observed jump-off rates are used to enhance forecast accuracy.

2.3 Recurrent neural networks for the Visegrád Group

The second study uses mortality data from the Human Mortality Database for the four Visegrád Group countries. Annual deaths and central exposures are available by age and sex. The data cover 1970–2019 and the adult population aged 18–99. The training period is 1970–2009, while the subsequent ten years form the test period. Lee–Carter models and neural networks are fitted separately by country and sex; the multi-population Li–Lee models are fitted separately by sex.

Simple recurrent neural networks, Long Short-Term Memory networks, and Gated Recurrent Units are considered. For all networks, the target is the logarithmic mortality rate for a given age and calendar year, and the inputs are lagged values of the target. Fivefold simple cross-validation and grid search are used for hyperparameter optimization. The grids contain 3, 5, or 8 lags; batch sizes of 64, 128, or 256; and 5, 6, or 7 neurons, with 10, 12, or 14 neurons for the simple RNN. The networks have a single recurrent layer with a hyperbolic tangent activation and an identity output. Training uses Adam with a learning rate of 0.001 and mean squared error loss, early stopping, and 50 runs for each final parameter set. The best run and the ensemble average are both evaluated.

2.4 Feedforward neural networks and coherence

The third study examines whether feedforward neural networks can produce coherent multi-population mortality forecasts. Coherence is a crucial property: forecasts for related populations should not diverge indefinitely. The proposed network uses age and population embeddings together with calendar time and passes them through fully connected hidden layers. The theoretical analysis classifies activation functions according to the asymptotic differences generated between population forecasts.

Linear, ReLU, Leaky ReLU, and ELU activation functions belong to the hypercoherent class; Sigmoid, Tanh, and Softsign belong to the convergent class; Softplus, Swish, and GELU belong to the weakly coherent class. Sufficient conditions are derived for each coherence subtype under positivity restrictions on the relevant weights and regularity conditions for weak coherence.

The empirical illustration uses Human Mortality Database data for England & Wales, both sexes and ages 0–99. Calendar years 1960–1999 are used for training, 2000–2009 for validation, and 2010–2019 for testing. A single-hidden-layer network is optimized with Adam and mean squared error loss. Due to random initialization, 25 optimization runs are performed and a 10% trimmed mean provides the final forecast. Swish is used in the subsequent hyperparameter optimization because of its favorable coherence properties.

III. Scientific results of the dissertation

Thesis 1: Spatial and panel econometric techniques can improve mortality-index forecasts

The analysis of European countries demonstrates that more advanced time-series and econometric methods have clear practical value. Actuaries and demographers can improve forecast accuracy by moving beyond traditional Random Walk or ARIMA models when forecasting mortality indices in Lee–Carter or Li–Lee frameworks. Significant spatial and spatio-temporal autocorrelation is found in the differenced mortality indices, motivating the use of spatio-temporal forecasting techniques.

Within the Lee–Carter framework, LC–ESTF is the strongest overall performer, with six and seven country-level wins under average and maximum underperformance, respectively. LC–SDPLM achieves four wins under both criteria. Spatio-temporal techniques account for 17 and 14 wins out of 22 populations. Within the Li–Lee model, LL–DPLM has six and five wins, and LL–STARIMA has three and five wins. The results are summarized in Table 1.

Table 1: Number of wins by technique within each mortality model.

	Lee–Carter		Li–Lee	
Technique	AU	MU	AU	MU
RWD	1	5	3	3
ARIMA	1	1	1	0
DPLM	1	1	6	5
VAR	2	1	3	3
STARIMA	1	0	3	5
SDPLM	4	4	0	0
ESTF	6	7	2	2
ENS1	4	1	1	3
ENS2	2	2	3	1

Based on multiple robust metrics, LL–DPLM, LC–ESTF, LL–VAR, and LL–STARIMA emerge as particularly strong performers in point forecasts, with notable variation by country. Of these frameworks, all but LL–VAR are proposed in this context for the first time to the best of our knowledge. The mean-calibrated, post hoc adjusted bootstrap produces prediction intervals with correct nominal coverage. In interval forecasting, the newly introduced techniques outperform classical benchmarks, with LL–ESTF and LL–DPLM achieving the lowest median average interval scores. The forecasts of period life expectancy at birth in 2019 are sufficiently robust across model–technique combinations. Overall, spatial and panel econometric techniques may yield measurable improvements in accuracy for several populations.

Thesis 2: Neural networks are competitive, but their advantage varies by population and age

The Visegrád study compares Lee–Carter, Li–Lee, RNN, LSTM, and GRU models, both overall and by age group, for four countries and both sexes. Based on aggregated results, the best-trained neural networks consistently achieve the highest performance on the testing set. Ensemble models are also constructed to ensure robustness. For the female populations, machine-learning methods outperform the two classical models in aggregate, including with the more robust ensemble forecasts; Table 2 reports the corresponding errors.

Table 2: Forecasting errors for female populations. B denotes the best neural-network run and E the ensemble.

	LC	LL	LSTM B	LSTM E	GRU B	GRU E	RNN B	RNN E
CZE	2.21	3.24	1.94	2.07	1.96	2.02	1.98	2.20
HUN	7.40	4.70	3.31	3.91	3.08	3.78	3.46	4.08
POL	1.58	1.91	1.16	1.29	1.06	1.25	1.22	1.46
SVK	4.91	4.63	4.27	4.46	4.34	4.82	4.22	4.57

For male populations, the neural networks perform best in the Czech Republic and Hungary. In Poland, competition is closer: Li–Lee approaches the best neural-network forecasts and outperforms the ensemble models. In Slovakia, the classical models perform better than this neural-network approach. The aggregated errors are shown in Table 3.

Table 3: Forecasting errors for male populations. B denotes the best neural-network run and E the ensemble.

	LC	LL	LSTM B	LSTM E	GRU B	GRU E	RNN B	RNN E
CZE	1.83	2.62	1.31	1.49	1.28	1.52	1.35	1.57
HUN	17.75	6.11	1.77	2.99	2.11	3.91	2.07	3.44
POL	1.39	1.17	1.10	1.41	0.97	1.39	1.05	1.38
SVK	3.07	2.40	3.20	4.02	3.30	4.08	3.16	4.15

A more granular analysis by age reveals a different pattern. The Li–Lee model tends to provide the most accurate forecasts for older age groups, whereas neural-network methods achieve superior performance mainly for younger and middle-aged populations. The results therefore do not imply that neural networks should generally be preferred to classical mortality models. Their relative advantage varies considerably across countries, sexes, and ages. Classical stochastic mortality models remain competitive, are more parsimonious and interpretable, and provide a natural framework for forecast uncertainty. Neural networks may be preferable when they deliver a sufficiently large and robust improvement in out-of-sample point forecast accuracy.

The practical calculations lead to the same qualified conclusion. For life annuities beginning at age 65, Lee–Carter generally produces the lowest premium after the static approach, neural networks produce intermediate premiums, and Li–Lee produces the highest. Hungary is the exception because the neural network estimates higher mortality at ages 85–90 than the classical methods.

Thesis 3: Feedforward neural networks can produce coherent mortality forecasts

It can be formally shown that feedforward neural networks produce coherent forecasts for certain activation functions when the architecture described in the study is applied. Under positive time weights and positive inter-layer weight matrices, a network with coherent-class activation functions produces coherent mortality forecasts. If at least one hidden-layer activation belongs to the convergent class, population forecasts converge. If all activations belong to the hypercoherent class, population differences become constant in the long run. If all belong to the weakly coherent class and the regularity conditions are satisfied, forecasts are weakly coherent. A mixture without convergent activations, containing at least one weakly coherent activation, is also weakly coherent under the same conditions.

The empirical results also support the practical relevance of the proposed neural-network framework. The empirical analysis uses annual death and exposure data from the Human Mortality Database for 26 countries, with male and female observations treated as separate populations, resulting in a panel of 52 country–sex populations. Ages 0–99 are retained, with 1960–2009 used for model estimation and 2010–2019 as an out-of-time evaluation period. Two constrained feedforward neural networks with ReLU and Swish activation functions are compared with Li–Lee models in which the population-specific mortality indices follow either a Random Walk or an AR(1) process. The neural networks use calendar year, age, and population embeddings as inputs and impose positive weights to preserve the required coherence properties. Each neural specification is estimated using ten random initializations, and the resulting forecasts are averaged to form ensemble predictions.

The ReLU ensemble achieves the lowest median and mean country-level absolute MSE underperformance, at 0.001404 and 0.003024, respectively, and is the best-performing method in 20 countries. The Swish ensemble follows closely, with median and mean underperformance of 0.001604 and 0.003441, and performs best in 16 countries, whereas each Li–Lee specification is best in eight countries. However, the Li–Lee model with a Random Walk produces the lowest maximum underperformance (0.015894), compared with 0.019469 for the ReLU ensemble. Thus, the neural-network ensembles provide the strongest overall country-level performance, while the Li–Lee Random Walk offers the most favorable worst-case result.

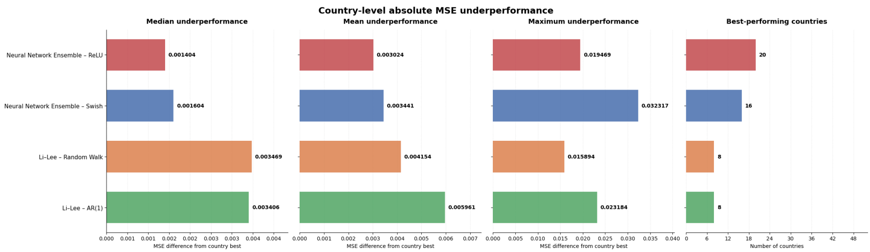

Figure 1: Country-level absolute forecast underperformance of the ReLU and Swish neural-network ensembles and the two Li–Lee benchmarks, 2010–2019.

IV. Main references

Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, 87(419), 659–671. https://doi.org/10.1080/01621459.1992.10475265

Li, N. and Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee–Carter method. *Demography*, 42(3), 575–594. https://doi.org/10.1353/dem.2005.0021

Meslé, F. and Vallin, J. (2002). Mortality in Europe: The divergence between east and west. *Population*, 57(1), 157–197. https://doi.org/10.2307/3246630

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1–2), 17–23. https://doi.org/10.2307/2332142

Omran, A. R. (1971). The epidemiological transition. *Milbank Memorial Fund Quarterly*, 49(4), 509–538. https://doi.org/10.2307/3349375

Richman, R. and Wüthrich, M. V. (2019). A neural network extension of the Lee–Carter model to multiple populations. *Annals of Actuarial Science*, 13(2), 370–389.

Vallin, J. and Meslé, F. (2004). Convergences and divergences in mortality: A new approach to health transition. *Demographic Research*, Special Collection 2, 11–44.

V. Publications related to the dissertation

Szentkereszti, G. and Vékás, P. (2022). Magyar halandósági ráták előrejelzése visszacsatolt neurális hálózatokkal. *Statisztikai Szemle*, 100(10), 905–922. https://doi.org/10.20311/stat2022.10.hu0905

Szentkereszti, G. and Vékás, P. (2024). A visegrádi országok mortalitási rátáinak előrejelzése neurális hálózatokkal. *Szigma*, 55(2–3), 293–312. https://doi.org/10.15170/SZIGMA.55.1242

Szentkereszti, G. and Vékás, P. (2026). Mortality forecasting in geographical space and time. *ASTIN Bulletin: The Journal of the International Actuarial Association*.