Coherent Forecasting of Mortality Rates: A Spatial-Temporal Approach

Yang Lu
Aix-Marseille School of Economics, Aix-Marseille University

Hong Li
School of Finance, Nankai University

12th International Longevity Risk and Capital Markets Solutions Conference, September 29-30, 2016, Chicago, U.S.A.

Motivation

- Human life expectancy has been increasing substantially since the 20th century.
- Great advance in literature of stochastic mortality modelling in recent decades.
- Single population models
  - Lee and Carter (1992); Cairns, Blake and Dowd (2006); Renshaw and Haberman (2006); Hyndman and Ullah (2007)...
- Multi-population models
  - Li and Lee (2005); Dowd et al. (2011); Zhou et al. (2014); Hyndman et al. (2013); ...

Motivation

- Widely used age-period mortality models (e.g., Lee-Carter and Cairns-Blake-Dowd) are likely to produce diverging mortality forecasts for different ages.
- The Lee-Carter model:

$$\log m_{i,t} = a_i + b_i k_t + \epsilon_{i,t}, \qquad \forall i, t.$$

- $\lim_{t\to\infty} \frac{\log m_{k,t}}{\log m_{j,t}} = \infty$ if $b_k < b_j$.
- Li et al. (2013) propose to gradually rotate b_i-s (in an *ad hoc* way), so that they are equal for the majority of ages at some point in the future.
- A similar issue exists for multi-population models.

Motivation

- Existing studies suggest that goodness-of-fit may be improved when a cohort effect is included (Cairns et al., 2009).
- The cohort extension of Lee-Carter (Renshaw and Haberman, 2006):

$$y_{i,t} = a_i + b_i k_t + c_i \gamma_{t-i} + \epsilon_{i,t}.$$

- Suffered from identifiability issue (see e.g. Kuang et al., 2008, Hunt and Villegas, 2015).
- Should be used with caution when forecasting.

Contribution

- We propose a spatial-temporal vector autoregressive model for the mortality surface.
- The model allows the mortality rates at different ages to feature spatial dependence, i.e. the nearer the two ages, the more substanial the dependence between mortality processes at these ages.
- It has nice dynamic properties, e.g., coherent mortality forecasts among ages.
- It has a flexible structure, and has Lee-Carter and CBD model as special cases.

Table of Contents

A (sparse) Vector Autoregressive Model

$$y_{i,t} = (1 - \alpha_i - \beta_i) y_{i,t-1} + \alpha_i y_{i-1,t-1} + \beta_i y_{i-2,t-1} + m_i + \epsilon_{i,t},$$

- $y_{i,t}$ — mortality quantities, e.g., $\log m_{i,t}$ or $\text{logit} q_{i,t}$;
- $(1 - \alpha_i - \beta_i)$ — the period effect;
- α_i — the cohort effect (from the same cohort);
- β_i — the cohort effect (from the younger neighbouring cohort);
- $(\epsilon_{i,t})$ are multivariate normal and independent over time.

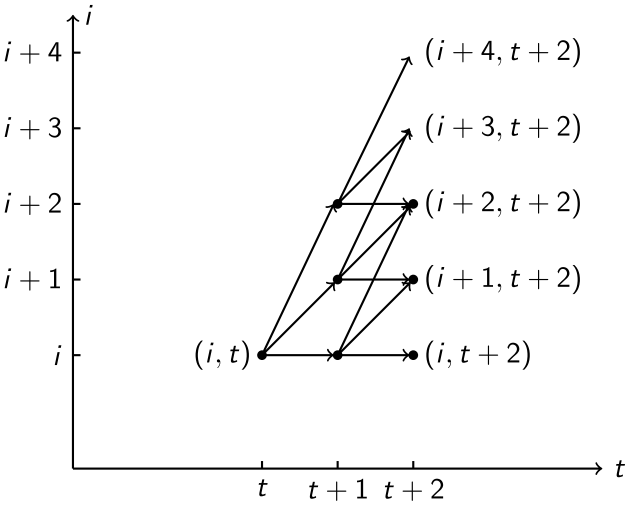
VAR representation

Thus the vector process $(y_{i,t})$ is a large and sparse VAR process:

$$\begin{bmatrix} y_{1,t+1} \\ y_{2,t+1} \\ y_{3,t+1} \\ \vdots \\ y_{I,t+1} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & \dots & \dots \\ \alpha_2 & 1-\alpha_2 & 0 & \dots & \dots \\ \beta_3 & \alpha_3 & 1-\alpha_3-\beta_3 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \dots & 0 & \beta_I & \alpha_I & (1-\alpha_I-\beta_I) \end{bmatrix}}_{R} \begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \\ \vdots \\ y_{I,t} \end{bmatrix} + \begin{bmatrix} m_{1,t} \\ m_{2,t} \\ m_{3,t} \\ \vdots \\ m_{I,t} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t+1} \\ \epsilon_{2,t+1} \\ \epsilon_{3,t+1} \\ \vdots \\ \epsilon_{I,t+1} \end{bmatrix},$$

- A spatial-temporal autoregressive model (Pace et al., (1998)).
- For given t, closer i and j leads to closer $y_{i,t}$ and $y_{j,t}$.
- Shocks $(\epsilon_{i,t})$ from one cohort enters other cohorts progressively.

Propagation of shocks



Long-term coherence among ages

Theorem

If $0 < \alpha_i + \beta_i < 1$ *for all* i*, then for any ages* $i \neq j$*, the mortality rates* $y_{i,t}$ *and* $y_{j,t}$ *are such that:*

$$y_{i,t} - y_{j,t}$$

is stationary.

The Covariance matrix: a simultaneous equation

$$
\begin{aligned}
\epsilon_{I,t} &= c_I \epsilon_{I-1,t} + \eta_{I,t} \\
&\cdots\cdots \\
\epsilon_{i,t} &= a_i \epsilon_{i+1,t} + c_i \epsilon_{i-1,t} + \eta_{i,t} \\
&\cdots\cdots \\
\epsilon_{1,t} &= a_1 \epsilon_{2,t} + \eta_{1,t},
\end{aligned}
$$

- $\eta_{i,t}$ are i.i.d. white noise with variance σ_i^2.
- a_i — effect from the older neighbouring cohort;
- c_i — effect from the younger neighbouring cohort;
- Closer i and j leads to closer $\epsilon_{i,t}$ and $\epsilon_{j,t}$.
- A flexible structure with a reasonable amount of parameters ($3I - 2$ in total).

The Covariance matrix

- In matrix term:

$$\underbrace{\begin{bmatrix} 1 & -c_I & 0 & \dots & \dots \\ -a_{I-1} & 1 & -c_{I-1} & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \dots & 0 & -a_2 & 1 & -c_2 \\ \dots & \dots & 0 & -a_1 & 1 \end{bmatrix}}_{M} \epsilon_t = \eta_t.$$

- The covariance matrix:

$$\Sigma_\epsilon = (M^{-1})' Diag(\sigma_1^2, ..., \sigma_I^2) M^{-1}.$$

- M is invertible when $a_i + c_i < 1$ for all i.

Special cases of the spatial VAR model

- The Lee-Carter model
  - $\alpha_i = \beta_i = 0$;
  - $\epsilon_{i,t} = m_i \epsilon_t$;

$$y_{i,t} = y_{i,1} + m_i \underbrace{\left(t - 1 + \sum_{\tau=1}^{t} \epsilon_\tau\right)}_{k_t}.$$

- The Cairns-Blake-Dowd model
  - $\alpha_i = \beta_i = 0$;
  - $\epsilon_{i,t} = \epsilon_{1,t} + i\epsilon_{2,t}$;
  - $m_i = \tilde{m}_i + i\hat{m}_i$

$$y_{i,t} = y_{i,1} + \underbrace{\left(t\tilde{m}_i + \sum_{\tau=1}^{t} \epsilon_{1,\tau}\right)}_{k_{1,t}} + i \times \underbrace{\left(t\hat{m}_i + \sum_{\tau=1}^{t} \epsilon_{2,\tau}\right)}_{k_{2,t}}.$$

Table of Contents

A penalized least square approach

- The model can be directly estimated by (two-stage) ordinary least square
  1. estimates the causality matrix R, calculate $\hat{y}$;
  2. calculate $\hat{\epsilon} = y - \hat{y}$, and estimate the variance-covariance matrix.
- Nevertheless, the resulting parameter estimates would be rather erratic, due to limitation of observations.
- We propose a penalized least square method
  - to smooth the parameter estimates;
  - When the penalty term is quadratic, closed form solutions are available.

A two-step penalized least square approach

- First, obtain $(\hat{\alpha}_i)_{i=2,\dots,I}, (\hat{\beta}_i)_{i=3,\dots,I}, (\hat{m}_i)_{i=1,\dots,I}$ by minimizing

$$L_1 = \sum_{i=2}^{I-1} \sum_{t=1}^{T-1} \epsilon_{i,t}(\alpha_i, \beta_i, m_i)^2 + \lambda_\alpha \sum_{i=3}^{I-1} (\alpha_{i+1} - \alpha_i)^2 + \lambda_\beta \sum_{i=3}^{I-1} (\beta_{i+1} - \beta_i)^2 + \lambda_m \sum_{i=3}^{I-1} (m_{i+1} - m_i)^2$$

- Second, given $\hat{\epsilon}(\alpha, \beta, m)$, obtain $(\hat{a}_i)_{i=1,\dots,I-1}, (\hat{c}_i)_{i=2,\dots,I}$ by minimizing

$$L_2 = \sum_{i=2}^{I-1} \sum_{t=1}^{T} \eta_{i,t}(a_i, c_i)^2 + \lambda_a \sum_{i=2}^{I-1} (a_{i+1} - a_i)^2 + \lambda_c \sum_{i=2}^{I-1} (c_{i+1} - c_i)^2$$

where $(\lambda_\alpha, \lambda_\beta, \lambda_m, \lambda_a, \lambda_c)$ are *known* smoothing coefficients.

A penalized least square approach

- When $(\lambda_\alpha, \lambda_\beta, \lambda_m, \lambda_a, \lambda_c)$ are all zero, we get the OLS estimator.
- The solution is unique and has closed form expression, which depends on the choice of the smoothing coefficients.
- The optimal value of $(\lambda_\alpha, \lambda_\beta, \lambda_m, \lambda_a, \lambda_c)$ is determined by a cross-validation method.

Two-population extension, the casualty matrix

The model can be easily extended to the two-population case.

$$y_{j,i,t} = \rho_{j,i}\Big[(1-\alpha_{j,i})y_{j,i,t-1} + \alpha_{j,i}y_{j,i-1,t-1}\Big] + (1-\rho_{j,i})y_{-j,i,t-1} + m_{j,i} + \epsilon_{j,i,t},$$

and

$$y_{j,1,t} = \rho_{j,1}y_{j,1,t-1} + (1-\rho_{j,1})y_{-j,1,t-1} + m_{j,1} + \epsilon_{j,1,t}.$$

- $\rho_{j,i}$ — weight assigned to the own population;
- $(1-\alpha_{j,i})$ — period effect;
- $\alpha_{j,i}$ — cohort effect;
- Identical estimation strategy.

Two-population extension, the covariance matrix

$$\epsilon_{j,I,t} = \theta_{j,I} c_{j,I} \epsilon_{j,I-1,t} + (1-\theta_{j,I})\epsilon_{-j,I,t} + \eta_{j,I,t}$$

$$\cdots\cdots$$

$$\epsilon_{j,i,t} = \theta_{j,i}\left(a_{j,i}\epsilon_{j,i+1,t} + c_{j,i}\epsilon_{j,i-1,t}\right) + (1-\theta_{j,i})\epsilon_{-j,i,t} + \eta_{j,i,t}$$

$$\cdots\cdots$$

$$\epsilon_{j,1,t} = \theta_{j,1} a_{j,1} \epsilon_{j,2,t} + (1-\theta_{j,1})\epsilon_{-j,1,t} + \eta_{1,t},$$

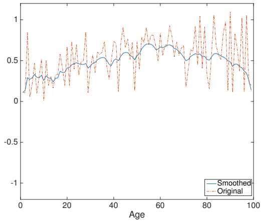
- $\theta_{j,i}$ — weight assigned to the own population;
- $a_{j,i}$ — effect of the older neighbouring age;
- $c_{j,i}$ — effect of the younger neighbouring age.

Table of Contents

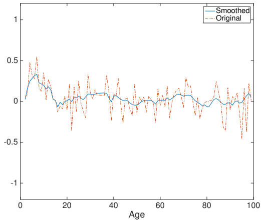
Data

- US and UK uni-sex mortality data ($\log m$);
- Age 0 to 99;
- Year 1950 to 2013.
- Both the single-population and two-population model.

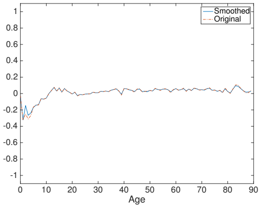
Single-population US, casualty matrix



(a) α

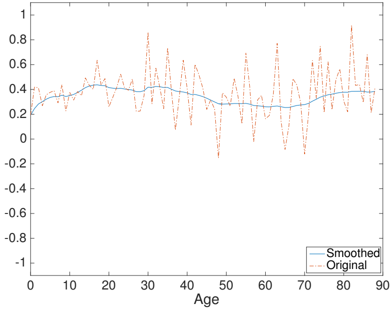


(b) β

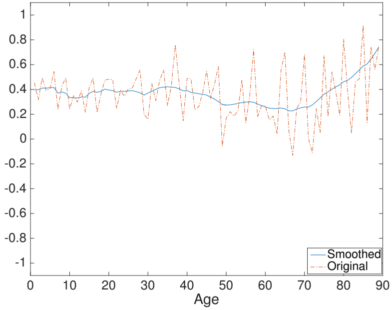


(c) m

Single-population US, covariance matrix

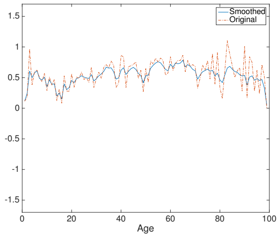


(a) a

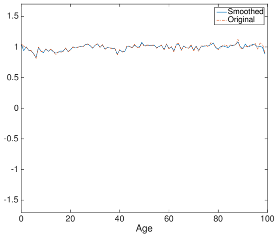


(b) c

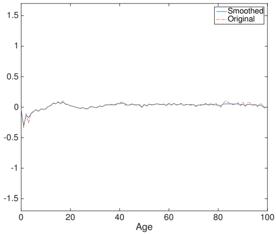
Two-population US, casualty matrix



(a) α

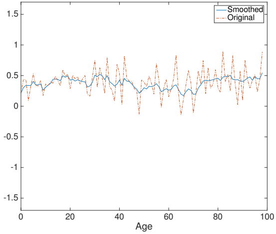


(b) ρ

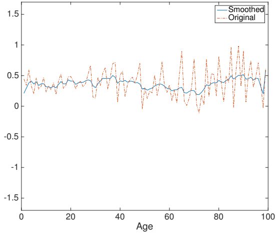


(c) m

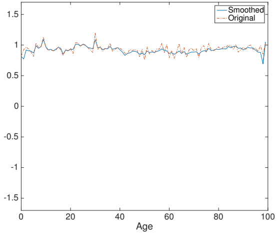
Two-population US, covariance matrix



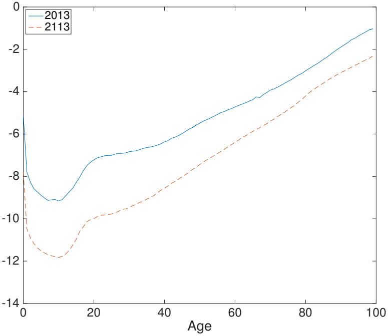
(a) *a*



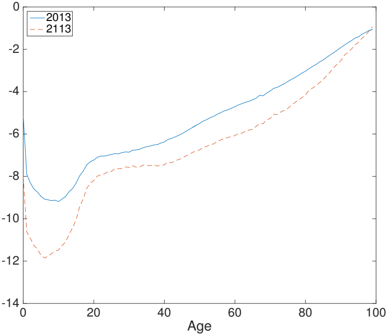
(b) *c*



Projected log m, single population, US



(a) SAR



(b) Lee-Carter

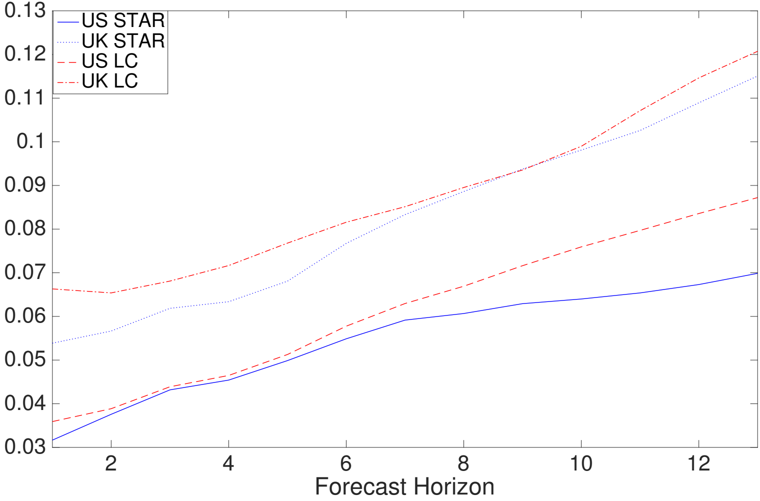
Out-of-sample forecast

- Estimate the models using data up to year 2000, and forecast the mortality rates from year 2001 to 2013.
- Root mean squared forecast error (RMSFE)

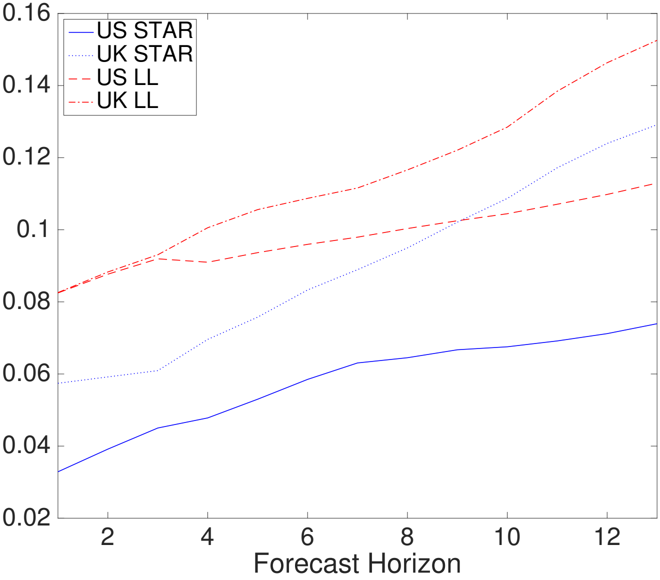
$$RMSFE(j) = \sqrt{\frac{1}{I(U - \hat{u})} \sum_{u=\hat{u}}^{U} \sum_{i=1}^{I} (y_{j,i,u} - \widehat{y}_{j,i,u})^2},$$

- Compare with the Lee-Carter and the Li-Lee model, respecitvely.

RMSFE, Single-population



RMSFE, Two-population



Conclusion

We propose a spatial temporal vector autoregressive model which

- ensures coherent mortality forecasts among ages;
- includes a cohort effect which is easy to explain;
- is easy to estimate (closed form penalized least square estimation).
- can be naturally extended to the multiple population case.
- produces reasonable forecasting performance.

Cairns, A., Blake, D., and Dowd, K. (2006).
A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration.
Journal of Risk and Insurance, 73(4):687–718.

Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., and Balevich, I. (2009).
A quantitative comparison of stochastic mortality models using data from england and wales and the united states.
North American Actuarial Journal, 13(1):1–35.

Dowd, K., Cairns, A., Blake, D., Coughlan, G., and Khalaf-Allah, M. (2011).
A Gravity Model of Mortality Rates for Two Related Populations.
North American Actuarial Journal, 15(2):334–356.

Hunt, A. and Villegas, A. M. (2015).
Robustness and convergence in the lee–carter model with cohort effects.
Insurance: Mathematics and Economics, 64:186–202.

Hyndman, R. J., Booth, H., and Yasmeen, F. (2013).
Coherent mortality forecasting: the product-ratio method with functional time series models.
Demography, 50(1):261–283.

Kuang, D., Nielsen, B., and Nielsen, J. (2008).
Identification of the Age-period-Cohort Model and the Extended Chain-Ladder Model.
Biometrika, 95(4):979–986.

Lee, R. and Carter, L. (1992).
Modeling and Forecasting US Mortality.
Journal of the American Statistical Association, 87(419):659–671.

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.

Li, N., Lee, R., and Gerland, P. (2013).

Extending the lee-carter method to model the rotation of age patterns of mortality decline for long-term projections.
Demography, 50(6):2037–2051.

Pace, R. K., Barry, R., Clapp, J. M., and Rodriquez, M. (1998).
Spatiotemporal autoregressive models of neighborhood effects.
Journal of Real Estate Finance and Economics, 17(1):15–33.

Renshaw, A. and Haberman, S. (2006).
A Cohort-Based Extension to the Lee-Carter Model for Mortality Reduction Factors.
Insurance: Mathematics and Economics, 38(3):556–570.

Zhou, R., Wang, Y., Kaufhold, K., Li, J. S.-H., and Tan, K. S. (2014).
Modeling period effects in multi-population mortality models: Applications to solvency ii.
North American Actuarial Journal, 18(1):150–167.