A Step-by-Step Guide to Building Two-Population Stochastic Mortality Models

Johnny Li   Rui Zhou   Mary Hardy

Longevity 9
September 2013, Beijing

Why Multi-Population Models Are Needed?

- Capture the dependence of mortality dynamics between different populations
- Useful for modeling mortality dynamics of:
  - Males and females
  - Different national populations
  - A population and its subpopulation(s)
- Useful for estimating basis risk involved in standardized longevity risk hedges

Existing Two-Population Models

- **APC structure**
  - Dowd et al. (2011) : *A gravity model*
  - Cairns et al. (2011) : *A general framework for two-population mortality modeling*
- **Lee-Carter structure**
  - Yang and Wang (2013) : *VECM for period effects*
  - Zhou, Li, Tan (2013) : *RWAR with transitory jumps*
  - Zhou et al. (2012) : *Symmetric processes for period effects*
  - Li and Hardy (2011) : *A co-integrated Lee-Carter model*
  - Li and Lee (2005) : *The augmented common factor model*
  - Carter and Lee (1992) : *The joint-k model*

Some Comments

- Most existing multi-population models are based on the Lee-Carter and APC structures.
- In many situations, these two structures are not the best performing.
- For instance, Cairns et al. (2009) found that models M2 and M7 significantly outperform the Lee-Carter and APC structures in an application to US and England and Wales (EW) data.

Commonly Used Model Structures

- **M1 (Lee-Carter)**

$$\ln m_{x,t} = \beta_x^{(1)} + \beta_x^{(2)} \kappa_t^{(2)}$$

- **M2 (Renshar-Haberman)**

$$\ln m_{x,t} = \beta_x^{(1)} + \beta_x^{(2)} \kappa_t^{(2)} + \beta_x^{(3)} \gamma_{t-x}^{(3)}$$

- **M3 (APC)**

$$\ln m_{x,t} = \beta_x^{(1)} + n_a^{-1} \kappa_t^{(2)} + n_a^{-1} \gamma_{t-x}^{(3)}$$

- **M5 (Original CBD)**

$$\text{logit } q_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x})$$

- **M6 (CBD with a cohort effect term)**

$$\text{logit } q_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + \gamma_{t-x}^{(3)}$$

- **M7 (CBD with cohort effect and quadratic terms)**

$$\text{logit } q_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + \kappa_t^{(3)}((x - \bar{x}))^2 - \sigma_x^2) + \gamma_{t-x}^{(4)}$$

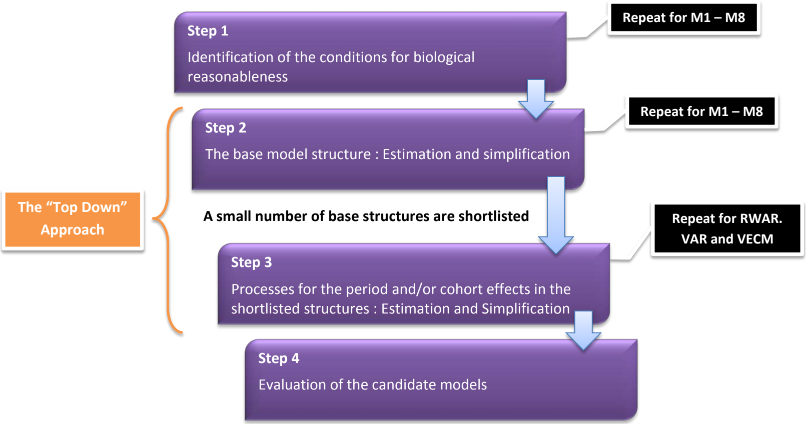
- **M8 (CBD with age-dependent cohort effects)**

$$\text{logit } q_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + \gamma_{t-x}^{(3)}(x_c - x)$$

Research Objectives

- Challenges
  - Structures other than the Lee-Carter and APC are usually more complex
  - In a multi-population set-up, the number of parameters is very large when the correlations among all period and/or cohort effects are modeled.
- Our contributions
  - A step-by-step procedure to generalize existing single-population models to multi-population models
  - A “top-down” approach (proposed as a concept by Cairns (2012)) to improve compactness

The Step-by-Step Procedure at a Glance



Data and Notation

Data

- Illustration I : EW males and CMI males
  Age range: 60-84; Year range: 1947-2005
- Illustration II : EW males and US males
  Age range: 60-84; Year range: 1947-2005

Notation

- Superscript (i, j) denotes the i-th term in the base model structure for the j-th population.
- E.g. M2:

$$\ln m_{x,t}^{(j)} = \beta_x^{(1,j)} + \beta_x^{(2,j)}\kappa_t^{(2,j)} + \beta_x^{(3,j)}\gamma_{t-x}^{(3,j)}$$

Step I

Identification of the conditions for biological reasonableness

Considerations of Biological Reasonableness

- A global convergence in mortality levels has been observed in the second half of the 20th century (White, 2002; Wilson, 2001).
- Constraints are needed to avoid an increasing divergence in life expectancy forecasts in the long run.
- Cairns et al. (2011) derived the non-divergence conditions for multi-population Lee-Carter and APC models.
- We derive the non-divergence conditions for the multi-population versions of M2, M5, M6, M7 and M8.

The Non-Divergence Conditions for M2 and M7

M2: $\ln m_{x,t}^{(j)} = \beta_x^{(1,j)} + \beta_x^{(2,j)}\kappa_t^{(2,j)} + \beta_x^{(3,j)}\gamma_{t-x}^{(3,j)}$

- $\beta_x^{(2,1)} = \beta_x^{(2,2)}$
- $\beta_x^{(3,1)} = \beta_x^{(3,2)}$
- $\kappa_t^{(2,1)} - \kappa_t^{(2,2)}$ is mean-reverting.
- $\gamma_{t-x}^{(3,1)} - \gamma_{t-x}^{(3,2)}$ is mean-reverting.

M7: logit $q_{x,t}^{(j)} = \kappa_t^{(1,j)} + \kappa_t^{(2,j)}(x - \bar{x}) + \kappa_t^{(3,j)}((x - \bar{x}))^2 - \sigma_x^2) + \gamma_{t-x}^{(4,j)}$

- $\kappa_t^{(i,1)} - \kappa_t^{(i,2)}$, $i = 1, 2, 3$, is mean-reverting.
- $\gamma_{t-x}^{(4,1)} - \gamma_{t-x}^{(4,2)}$ is mean-reverting.

Step II

The base model structure: Estimation and simplification

Estimating the Base Model Structure

- The maximum likelihood method is used.
- Necessary non-divergence conditions (if any) are taken into account.
- Assume that

$$D_{x,t}^{(j)} \sim \text{Poisson}\left(E_{x,t}^{(j)} m_{x,t}^{(j)}\right),$$

where $D_{x,t}^{(j)}$ and $E_{x,t}^{(j)}$ are the death and exposure counts.
- The log-likelihood is

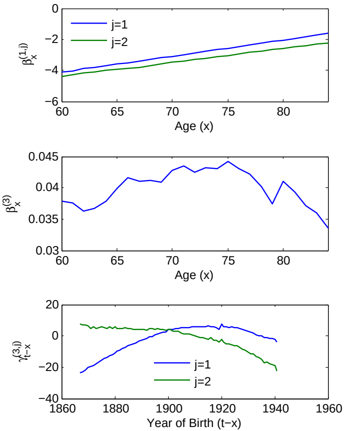
$$\textstyle\sum_{i,x,t} w_{x,t}^{(j)} \left(D_{x,t}^{(j)} \ln m_{x,t}^{(j)} - E_{x,t}^{(j)} m_{x,t}^{(j)}\right) + \text{constant},$$

where $w_{x,t}^{(j)}$ is the weight function.

The “Top-Down” Approach

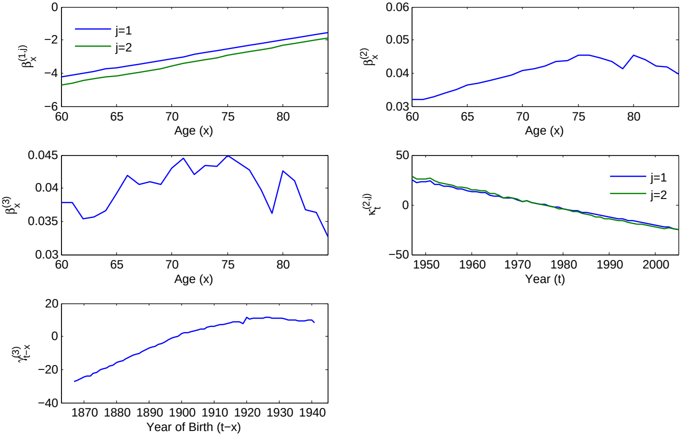
- The full model contains a large number of parameters.
- Can we combine the period and/or cohort effect terms for the two populations?
- For example, in Model M7:
  - $\kappa_t^{(3,1)} = \kappa_t^{(3,2)}$?
  - $\gamma_{t-x}^{(4,1)} = \gamma_{t-x}^{(4,2)}$?
- The “Top-Down” approach:
  - Fit the full model first
  - Gradually combine parameters for the two populations; fit the simplified model
  - Select the model structure that gives the best BIC value

Illustration I : EW and CMI



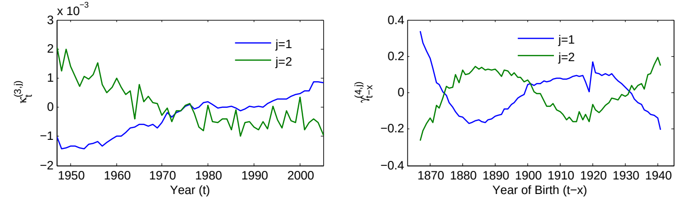
Parameter estimates for the two-population version of M2 (full model)

Illustration I : EW and CMI



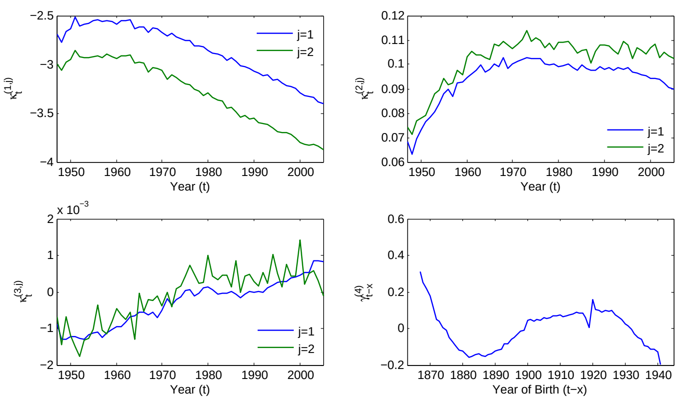
Parameter estimates of the two-population version of M2
(same cohort effects for both populations)

Illustration I : EW and CMI



Parameter estimates for the two-population version of M7 (full model)

Illustration I : EW and CMI



Parameter estimates for the two-population version of M7 (with same cohort effects for both populations)

Illustration I : EW and CMI

Model	**Simplification**	**BIC** $(0.5k\ln(n) - l)$
M1	None	19,760
M2	None	16,964
	$\gamma_{t-x}^{(3,1)} = \gamma_{t-x}^{(3,2)}$	16,898
M3	None	17,583
	$\gamma_{t-x}^{(3,1)} = \gamma_{t-x}^{(3,2)}$	17,409
M5	None	20,311
M6	None	17,716
	$\gamma_{t-x}^{(3,1)} = \gamma_{t-x}^{(3,2)}$	17,548

Illustration I : EW and CMI

Model	**Simplication**	**BIC** $(0.5k\ln(n) - l)$
M7	None	17,956
	$\gamma_{t-x}^{(4,1)} = \gamma_{t-x}^{(4,2)}$	17,535
	$\kappa_t^{(3,1)} = \kappa_t^{(3,2)}$	17,788
	$\kappa_t^{(3,1)} = \kappa_t^{(3,2)},\ \gamma_{t-x}^{(4,1)} = \gamma_{t-x}^{(4,2)}$	17,619
M8	None	17,725
	$\gamma_{t-x}^{(3,1)} = \gamma_{t-x}^{(3,2)}$	17,553

Illustration II : EW and US

Model	**Simplification**	**BIC** $(0.5k\ln(n) - l)$
M1	None	29,903
M2	None	22,929
	$\gamma_{t-x}^{(3,1)} = \gamma_{t-x}^{(3,2)}$	24,710
M3	None	25,340
	$\gamma_{t-x}^{(3,1)} = \gamma_{t-x}^{(3,2)}$	26,961
M5	None	35,716
M6	None	26,057
	$\gamma_{t-x}^{(3,1)} = \gamma_{t-x}^{(3,2)}$	30,002

Illustration II : EW and US

Model	**Simplification**	**BIC** $(0.5k\ln(n) - l)$
M7	None	24,974
	$\gamma_{t-x}^{(4,1)} = \gamma_{t-x}^{(4,2)}$	25,861
	$\kappa_t^{(3,1)} = \kappa_t^{(3,2)}$	24,906
	$\kappa_t^{(3,1)} = \kappa_t^{(3,2)}$, $\gamma_{t-x}^{(4,1)} = \gamma_{t-x}^{(4,2)}$	28,875
M8	None	25,579
	$\gamma_{t-x}^{(3,1)} = \gamma_{t-x}^{(3,2)}$	29,536

Step III

Processes for the period and/or cohort effects in the shortlisted structures: Estimation and simplification

Modeling Period and Cohort Effects: M7 as an Example

- In Step II, we found that M7 with $\gamma_{t-x}^{(4,1)} = \gamma_{t-x}^{(4,2)}$ gives the best BIC value for EW and CMI populations.
- In this model, the two populations share a common cohort effect term, but different period effect terms.
- A generic ARIMA process is used to model $\gamma_{t-x}^{(4)}$ (common cohort effect term).
- We require $\kappa_t^{(i,1)} - \kappa_t^{(i,2)}$, $i = 1, 2, 3$, to be mean-reverting.

Modeling Period and Cohort Effects: M7 as an Example

Possible processes for $\kappa_t^{(i,1)}$ and $\kappa_t^{(i,2)}$:

1. **RWAR**

$$\begin{aligned}\Delta\kappa_t^{(i,1)} &= \mu^{(i,1)} + Z_t^{(i,1)} \\ \kappa_t^{(i,1)} - \kappa_t^{(i,2)} &= \mu^{(i,2)} + \phi^{(i,2)}(\kappa_{t-1}^{(i,1)} - \kappa_{t-1}^{(i,2)}) + Z_t^{(i,2)}\end{aligned}$$

2. **VAR**

$$\Delta\kappa_t^{(i,j)} = \mu^{(i,j)} + \phi_1^{(i,j)}\Delta\kappa_{t-1}^{(i,1)} + \phi_2^{(i,j)}\Delta\kappa_{t-1}^{(i,2)} Z_t^{(i,j)}$$

Modeling Period and Cohort Effects: M7 as an Example

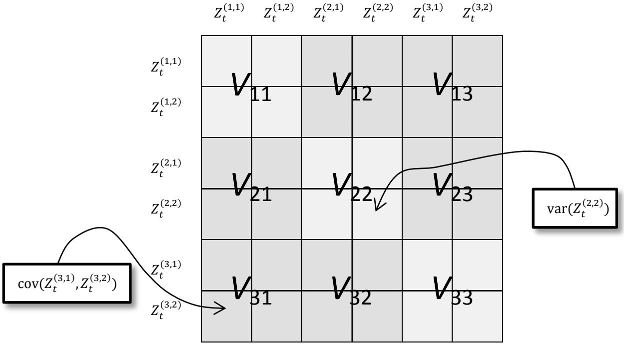
3. **VECM**

$$\begin{aligned}\Delta\kappa_t^{(i,j)} &= \mu^{(i,j)} + \alpha^{(i,j)}(\kappa_{t-1}^{(i,1)} - \kappa_{t-1}^{(i,2)}) \\ &\quad + \phi_1^{(i,j)}\Delta\kappa_{t-1}^{(i,1)} + \phi_2^{(i,j)}\Delta\kappa_{t-1}^{(i,2)} + Z_t^{(i,j)}\end{aligned}$$

- One bivariate process for each pair of period effect terms.
- The three processes are fitted jointly under the assumption that $Z_t^{(i,j)}$, $i = 1, 2, 3$ and $j = 1, 2$, follow a multivariate normal distribution.

Modeling Period and Cohort Effects: M7 as an Example

There are 21 parameters in the covariance matrix. We divide it into nine 2-by-2 matrices.



Modeling Period and Cohort Effects: M7 as an Example

The “top-down” approach is applied to the covariance matrix:

- The full model is fitted first.
- Fit models with reduced number of parameters in the covariance matrix:
  - $V_{12} = 0$
  - $V_{13} = 0$
  - $V_{23} = 0$
  - $V_{12} = 0$ and $V_{13} = 0$
  - $V_{12} = 0$ and $V_{23} = 0$
  - $V_{13} = 0$ and $V_{23} = 0$
  - $V_{12} = 0$, $V_{13} = 0$ and $V_{23} = 0$
- Choose the specification with best (lowest) BIC value.

Modeling Period and Cohort Effects: M7 as an Example

The "top-down" approach is applied to the AR coefficients:

- Consider the first pair of period effect terms ($i = 1$). The AR coefficients are $\phi_1^{(1,1)}$, $\phi_2^{(1,1)}$, $\phi_1^{(1,2)}$ and $\phi_2^{(1,2)}$.
  - Estimate the process with all AR coefficients.
  - Estimate the process with one or both cross-correlation parameters (i.e., $\phi_2^{(1,1)}$, $\phi_1^{(1,2)}$) equal zero.
  - Estimate the process with one or both auto-correlation parameters (i.e., $\phi_1^{(1,1)}$, $\phi_2^{(1,2)}$) equal zero.
- Repeat for the other two pairs of period effect terms (i.e., $i = 2, 3$).

Illustration I (EW and CMI), Revisited

- The three models with the best BIC values in Step II:
  - M2 with a common cohort effect term
  - M3 with a common cohort effect term
  - M7 with a common cohort effect term

- We fit RWAR, VAR and VECM processes to the period effect terms in the three shortlisted models.

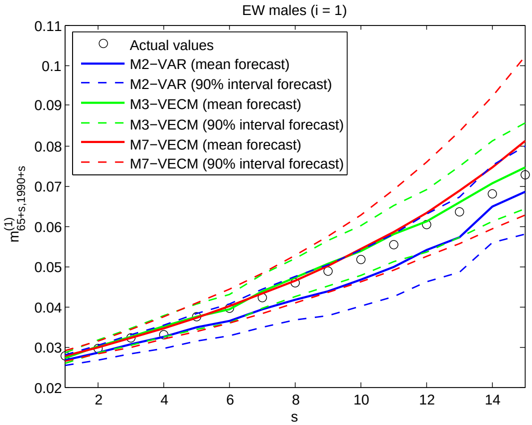
- The resulting BIC values:

	RWAR	VECM	VAR
M2	126.09	120.49	118.38
M3	127.48	120.28	120.44
M7	-1593.4	-1599.6	-1597.5

Step IV

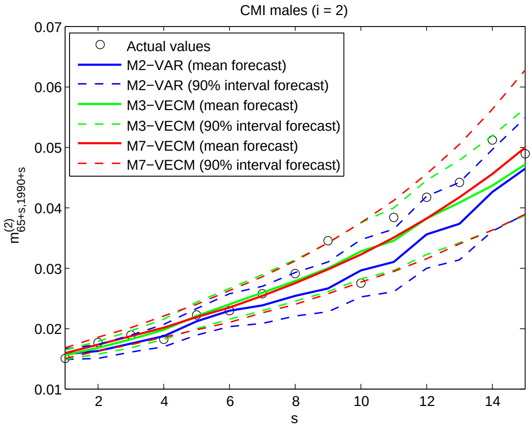
Evaluation of the candidate models

Evaluating In-Sample Forecasting Performance



"Forecasts" of the central death rates for the cohort aged 65 in 1990, based on models fitted to data from 1947 to 1990

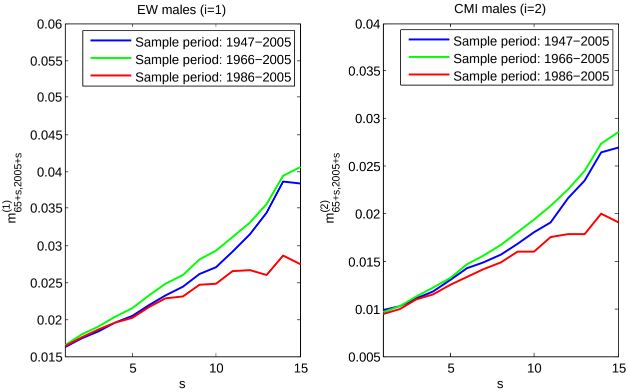
Evaluating In-Sample Forecasting Performance



"Forecasts" of the central death rates for the cohort aged 65 in 1990, based on models fitted to data from 1947 to 1990

Evaluating Robustness

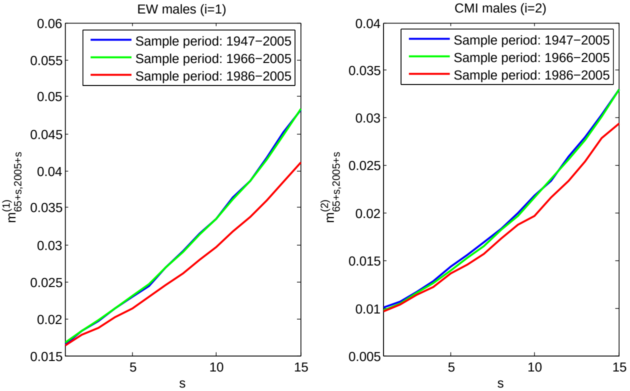
M2-VAR



"Forecasts" of the central death rates for the cohort aged 65 in 2005, based on models fitted to three different sample periods

Evaluating Robustness

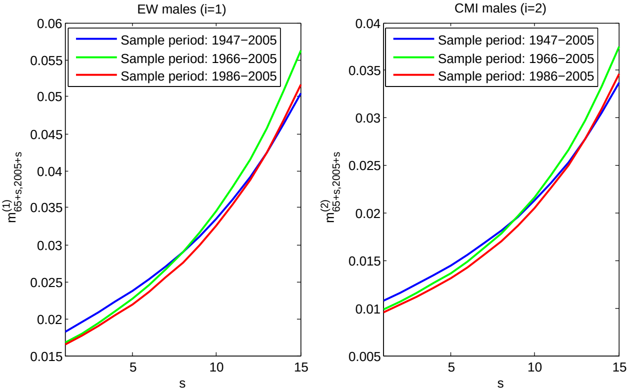
M3-VECM



"Forecasts" of the central death rates for the cohort aged 65 in 2005, based on models fitted to three different sample periods

Evaluating Robustness

M7-VECM



"Forecasts" of the central death rates for the cohort aged 65 in 2005, based on models fitted to three different sample periods

Summary

Summarizing the Procedure for EW and CMI Populations

- **Step 1**: Non-divergence conditions are derived.
- **Step 2**: Shortlisted the best three base structures:
  - M2 with a common cohort effect term
  - M3 with a common cohort effect term
  - M7 with a common cohort effect term
- **Step 3**: Identified processes for cohort and period effect terms in the shortlisted models:
  - M2 – VAR
  - M3 – VECM
  - M7 – VECM
- **Step 4**: Evaluated the three models in Step 3. M7 is recommended for this pair of populations.

Concluding Remarks

- We proposed a step-by-step approach to generalize single-population models to multi-population models.
- A “top down” is used to improve compactness.
- The proposed methods may be applied to single-population models other than the collection we considered.
- Jump effects may be introduced.