Modeling Mortality of Multiple Populations with Vector Error Correction Models: Applications to Solvency II

Rui Zhou[c], Yujiao Wang[b], Kai Kaufhold[a],
Johnny S.-H. Li[b,1], Ken Seng Tan[b]

[a] *Advanced Reinsurance Services GmbH, Cologne, Germany*
[b] *Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada*
[c] *Warren Centre for Actuarial Studies and Research, University of Manitoba, Winnipeg, Manitoba, Canada*

Abstract

Recently, Cairns et al. (2011) introduced a general framework for modeling the dynamics of mortality rates of two related populations simultaneously. Their method ensures that the resulting forecasts do not diverge over the long run by modeling the difference in the stochastic factors between the two populations with a mean-reverting autoregressive process. In this paper, we investigate how the modeling of the stochastic factors may be improved by using a vector error correction model (VECM). This extension is highly intuitive, allowing us to visualize the cross-correlations and the long-term equilibrium relation between the two populations. Another key benefit is that this extension does not require the user to assume which one of the two populations is dominant. This benefit is important because, as we demonstrate in this paper, it is not always easy to identify the dominant population, even if one population is much larger than the other. We illustrate our proposed extension with data from a pair of populations, and apply it to the calculation of Solvency II risk capital.

Keywords: Mortality forecasting; Longevity basis risk; Solvency II

1. Introduction

Mortality improvements in different populations may be correlated with one another. We expect, for example, that a breakthrough in treating a certain dread disease will improve mortality in a number of developed countries at similar, if not the same, time points. Therefore, when we project future mortality in multiple populations jointly, we need a model that takes account of the potential correlations across different populations. The multi-population model should also be structured carefully so that it will not result in diverging long-term forecasts, which do not seem to be biologically reasonable.

Multi-population mortality modeling has a wide range of applications. For instance, when projecting mortality for a smaller population with a thin volume of mortality data, the forecaster may want to enhance credibility by modeling the smaller population jointly with a larger population. The approach of using a larger population to benchmark was considered by Li et al. (2010) when they developed mortality improvement formulas for Canadian insured lives. Multi-population mortality models can also be used to simultaneously forecast mortality of both genders, thereby enforcing greater consistency of the sex differentials. The resulting projection, which has the statistical relationship between male and female mortality

[1]Corresponding author. Tel.: +1 (519) 888 4567 ext. 35542; fax: +1 (519) 746 1875.
E-mail: shli@uwaterloo.ca.

improvements embedded in it, is particularly useful for insurers in the European Union where gender-neutral pricing is enforced.

Multi-population mortality modeling plays a crucial role in securitization of longevity risk. The development of the longevity risk market has been quite slow so far. One reason for the slow development is the conflicting interests between the buy-side and sell-side of longevity risk. Buyers of longevity risk prefer standardized instruments, which are generally easier to analyze and are more conductive to liquidity, but sellers of the risk are often worried about longevity basis risk, which arises from the difference in mortality improvements between the sellers' populations and the populations to which the standardized longevity hedging instruments are linked. Multi-population mortality models help practitioners better estimate the longevity basis risk involved in a standardized longevity hedge, narrowing the gap between both sides of the market.

Recently, we have seen significant developments in multi-population modeling. Li and Lee (2005) developed an augmented common factor model for modeling mortality dynamics in multiple populations. Li and Hardy (2011) considered four possible ways to generalize a single-population mortality model to one that fits two or more populations. Cairns et al. (2011) introduced a general framework for producing consistent mortality forecasts for a pair of related populations. Dowd et al. (2011) designed a gravity mortality model for two related but different sized populations. A similar model has also been proposed by Jarner and Kryger (2011). Zhou et al. (2011) introduced a two-population mortality model with transitory jump effects, and applied it to pricing catastrophic mortality securitizations. All these models are constructed in such a way that forecasts of life expectancies in different populations do not diverge in the long run.

Many existing multi-population mortality models require a subjective assumption about which of the populations under consideration is dominant, driving the mortality dynamics of all other populations. Technically speaking, it is assumed that the mortality dynamics of the dominant population follow a non-stationary stochastic process, and that the difference between the mortality rates of the dominant population and a non-dominant population follows a mean-reverting stochastic process. In this way, over the long run, mortality in all populations will improve at the same speed, which is specified by the stochastic process for the dominant population. In previous studies of two-population mortality modeling, it is often assumed that the larger population is dominant. Nevertheless, as we are going to demonstrate in Section 2, this assumption is not always justifiable, and if the opposite assumption is used, the resulting projections can be very different.

In this paper, we investigate how this subjective assumption may be avoided. More specifically, we model the joint evolution of mortality over time using multivariate stochastic processes with a *symmetric* structure, thereby avoiding the need of making an assumption about which population is dominant. Processes considered include a vector autoregression (VAR) and a vector error correction model (VECM). We estimate these processes to data from a pair of populations, and evaluate the performance of these processes in terms of their goodness-of-fit, robustness with respect to changes in sample period, and performance in in- and out-of-sample forecasts.

The results for stochastic forecasts from both processes are then applied in a second step to a particular type of reinsurance transaction called a longevity swap, in which an insurance company or a pension fund can hedge its longevity risk by transferring it to a reinsurance provider. This type of financial transaction has been prevalent in the UK for the past five years with increasing volume, both in the number and size of the transactions. Most recently (on 10 Dec 2012), Swiss Re announced having completed a longevity swap reinsurance contract with the UK insurance company LV=,[2] taking the UK market's total transaction volume to over GBP 15 billion since 2009.[3] These longevity swaps are typically subject to longevity basis risk, either explicitly, when the hedge instrument is based on an underlying population mortality index, and the hedge purchaser retains the risk that its own portfolio of risks behaves differently from the general population, or implicitly, when the hedge provider has priced the longevity risk using a mortality forecast developed from general population mortality trends. We demonstrate the application of the multi-population mortality models to measuring longevity basis risk by using the stochastic mortality forecasts to calculate the so-called risk margin, as defined in the European Solvency II framework. The risk margin is a measure for the cost of capital which an average insurance company would have to take into account, when determining a risk-adjusted price for insurance policies. We use the risk margin as an estimate for the risk-adjusted price of a longevity swap, that is, the mark-up over best estimate liabilities, which a hedge provider would require in order to take on the risk, and study whether the choice of model has an influence on the price and whether our proposed models are suitable for measuring longevity basis risk.

The remainder of this paper is organized as follows. Section 2 explains in more detail the motivations of this study. Section 3 discusses the VAR and VECM modeling approaches. Section 4 estimates the models to data from a pair of populations and evaluates the performance of the estimated models. Section 5 applies the proposed modeling approaches to calculation of Solvency II risk capital. Section 6 concludes the paper with some suggestions for further research.

2. Motivations

We work under the two-population mortality modeling framework of Cairns et al. (2011). The first step in this framework is to specify a parametric structure for the age-specific death rates in each population under consideration. In this paper, we consider the Lee-Carter model structure (Lee and Carter, 1992),

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

where $m_{x,t}^{(i)}$ is the central death rate for population i at age x and in year t, $\alpha_x^{(i)}$ and $\beta_x^{(i)}$ are age-specific parameters, and $\kappa_t^{(i)}$ is a time-varying factor.[4] The second step is to model the evolution of the two time-varying factors $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ with stochastic processes, from which a joint forecast of mortality in both populations can be obtained.

This modeling framework was developed on the hypothesis that death rates in two related populations do not diverge over the long run. To satisfy the hypothesis of non-divergence, it is sufficient to assume that $\beta_x^{(1)} = \beta_x^{(2)}$ and that the spread $\kappa_t^{(1)} - \kappa_t^{(2)}$ is mean-reverting. When modeling the time-varying factors, it is often assumed that one population is dominant over the other. The time-varying factor for the dominant population is modeled by a non-stationary stochastic process, while the spread $\kappa_t^{(1)} - \kappa_t^{(2)}$ is modeled by a stationary one.

[2]Source: http://www.swissre.com/media/news_releases/nr_20121210_longevity_insurance_cover.html.

[3]Source: Hymans & Robertson, "Buy-outs, buy-ins and longevity hedging" 2012 Q1, available online at http://www.hymans.co.uk/media/118860/buy-out_buy-in_report_q12012.pdf.

[4]The Lee-Carter model is referred to as Model M1 in Cairns et al. (2009). This model structure is also used in an earlier (and fuller) version of Cairns et al. (2011) to demonstrate two-population mortality modeling. The earlier version of Cairns et al. (2011) is available at http://www.ma.hw.ac.uk/_andrewc/papers/ajgc54.pdf.

For instance, Cairns et al. (2011) suggested that if one population (say population 1) is much larger than the other, then we may model the time-varying factor $\kappa_t^{(1)}$ for the larger population with a random walk and the spread $\kappa_t^{(2)} - \kappa_t^{(1)}$ with a first order autoregressive process. This means that, over the long run, the evolution of mortality rates for the smaller population will follow that of the larger population.

Nevertheless, the assumption that the larger population is dominant over the smaller one is not always justifiable. To illustrate, let us consider the following pair of populations:

- Population 1: English and Welsh male population
  The data for this population are obtained from the Human Mortality Database (2012).

- Population 2: The population of UK male insured lives
  The data for this population are obtained from the Continuous Mortality Investigation (CMI) Bureau of the Institute and Faculty of Actuaries.

A sample period of $t_0 = 1961$ to $t_1 = 2005$ and a sample age range of $x_0 = 60$ to $x_1 = 84$ are used. This data set is also used some other papers on two-population mortality modeling (e.g., Cairns et al., 2011; Dowd et al., 2011).[5] Readers can therefore compare the results in this and the aforementioned papers readily.

Note that population 2 is much smaller than and is approximately a subset of population 1. In 2005, the number of exposures (over the age range of 60 to 84) in population 1 is about 4.7 million, while that in population 2 is about 260,000.

We fit equation (1) with $\beta_x^{(1)} = \beta_x^{(2)}$ for all x to the data set using the method of maximum likelihood. The log-likelihood function is derived under the assumption that the death count for population i at age x and time t follows a Poisson distribution with mean $m_{x,t}^{(i)} E_{x,t}^{(i)}$, where $E_{x,t}^{(i)}$ is the number of persons-at-risk for population i at age x and time t. We refer readers to Wilmoth (1993) for details about the log-likelihood function and to Li et al. (2009) for an algorithm for maximizing the log-likelihood function. Two constraints, $\sum_x \beta_x = 1$ and $\kappa_{t_1}^{(i)} = 0$, where t_1 is the last year in the sample period, are applied to stipulate parameter uniqueness.[6] The estimated values of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ within the sample period are shown graphically in Figure 1.

Of our interest is $\Delta\kappa_t^{(i)} = \kappa_t^{(i)} - \kappa_{t-1}^{(i)}$, which can be interpreted as the change in the overall level of mortality for population i from time $t-1$ to time t. The sample cross-correlations between $\Delta\kappa_t^{(1)}$ and $\Delta\kappa_t^{(2)}$ are displayed in Figure 2.

The cross-correlation at lags 3 and 13 are significant, indicating that $\Delta\kappa_t^{(1)}$ is significantly correlated with $\Delta\kappa_{t-3}^{(2)}$ and $\Delta\kappa_{t-13}^{(2)}$. This means that population 1's mortality improvement at present is significantly related to population 2's in the past. Equivalently speaking, in this example, it is the smaller population (population 2) that leads the bigger one, rather than the other way around. The above analysis motivates us to consider alternative modeling approaches that do not require us to make a subjective assumption about which population is dominant.

[5]Dowd et al. (2011) used exactly the same sample period and sample age range. Cairn et al. (2011) used the same sample period but a slightly different age range (60-89).

[6]Other parameter constraints may be used, but as Lee and Miller (2001) pointed out, the constraints we use produce more accurate mortality forecasts than other typical parameter constraints do. Over the long run, the difference between the expectations of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ will be a constant that is not necessarily zero, so using the constraint $\kappa_{t_1}^{(i)} = 0$, $i = 1, 2$, does not mean we are assuming that mortality rates in both populations have reached the long-term equilibrium at the forecast origin.

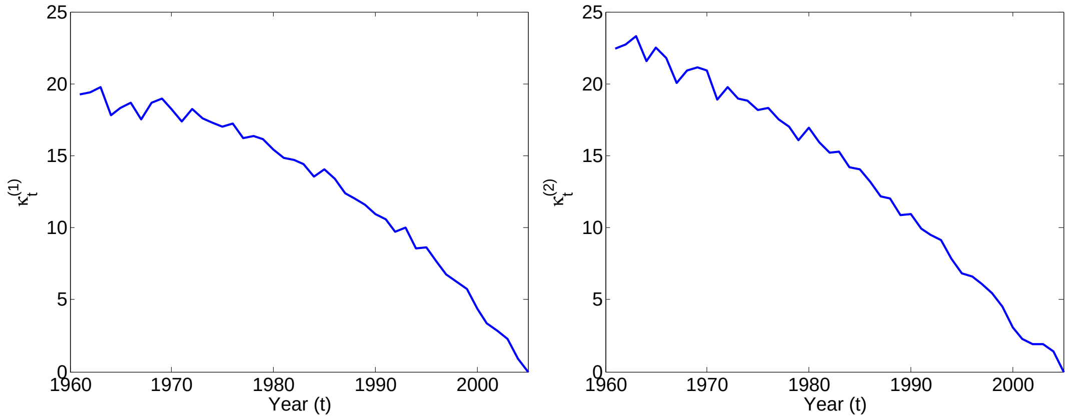

Figure 1: Estimated values of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ over the sample period of 1961-2005.

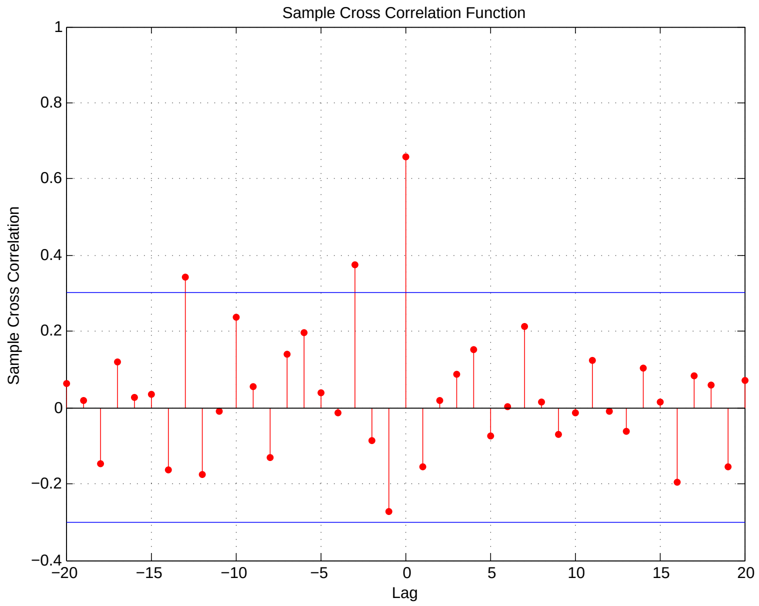

Figure 2: Sample cross-correlations between $\Delta\kappa_t^{(1)}$ and $\Delta\kappa_t^{(2)}$. The two horizontal solid lines are the 95% confidence bounds. The cross-correlations at lags -3 and -13 are significant.

3. Modeling the Time-Varying Factors

We continue to work with the framework of Cairns et al. (2011) and the Lee-Carter model structure, which is specified in equation (1). To satisfy the non-divergence hypothesis, we set $\beta_x^{(1)} = \beta_x^{(2)}$ for all x and model the time-varying factors in such a way that the difference $\kappa_t^{(1)} - \kappa_t^{(2)}$ will revert to a long-term mean.

We use the Lee-Carter model structure because it is relatively simple, allowing us to focus on our primary objective, namely to improve the joint modeling of the time-varying factors for different populations. We do knowledge that the Lee-Carter model may not be the best among all prevalent stochastic mortality models. Depending on the data and evaluation criteria, some more sophisticated models may outperform it (see, e.g., Cairns et al., 2009). However, these models contain additional features, which require further explanations and may distract readers from the main objective of this paper.

In what follows, we consider three methods for modeling $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$, namely a random walk plus a first order autoregressive process (RWAR), a vector autoregression (VAR) and vector error correction model (VECM). As in typical univariate modeling of the Lee-Carter time-varying factors, these three modeling approaches assume that $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ are first order stationary.

3.1. RWAR

This approach was used by Cairns et al. (2011) to illustrate their two-population modeling framework. In this approach, $\kappa_t^{(1)}$ is modeled by a random walk with drift, while the difference between $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ is modeled by a stationary first order autoregressive process:

$$\kappa_t^{(1)} = \kappa_{t-1}^{(1)} + \mu + \epsilon_t^{(1)}, \tag{2}$$

$$\kappa_t^{(1)} - \kappa_t^{(2)} = \mu_\Delta + \phi(\kappa_{t-1}^{(1)} - \kappa_{t-1}^{(2)}) + \epsilon_t^{(2)}, \tag{3}$$

where μ, μ_Δ and ϕ are model parameters. The innovations $\epsilon_t^{(1)}$ and $\epsilon_t^{(2)}$ are not serially correlated. They follow a bivariate normal distribution with a zero mean vector and a constant covariance matrix.

The RWAR approach assumes that population 1 is dominant. Over the long run, $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ are expected to change at the same speed μ, the drift for the dominant population.

As we mentioned in Section 2, the assumption that one of the two populations is dominant is not always justifiable. To overcome this problem, we propose the following two alternative specifications. They both have a symmetric model structure, thereby avoiding the need of assuming which of the two populations is dominant.

3.2. VAR

The VAR approach models $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ jointly with following structure:

$$\Delta\kappa_t^{(1)} = \phi_0 + \phi_1\Delta\kappa_{t-1}^{(1)} + \phi_2\Delta\kappa_{t-1}^{(2)} + \epsilon_t^{(1)}, \tag{4}$$

$$\Delta\kappa_t^{(2)} = \theta_0 + \theta_1\Delta\kappa_{t-1}^{(1)} + \theta_2\Delta\kappa_{t-1}^{(2)} + \epsilon_t^{(2)}, \tag{5}$$

where ϕ_0, ϕ_1, ϕ_2, θ_0, θ_1 and θ_2 are model parameters. As in RWAR, the innovations $\epsilon_t^{(1)}$ and $\epsilon_t^{(2)}$ have no serial correlation and follow a bivariate normal distribution with a zero mean vector and a constant covariance matrix. Cairns et al. (2011) mentioned the possibility of using this modeling approach, but they did not explain how it can be implemented.

We say that this model structure is symmetric, because equations (4) and (5) take exactly the same form. In the VAR approach, the cross-correlations between $\Delta\kappa_t^{(1)}$ and $\Delta\kappa_t^{(2)}$ are explicitly modeled. For instance, equation (4) implies that the mortality improvement in population 1 in a certain year is related to the previous mortality improvements in not only population 1 (through parameter ϕ_1) but also population 2 (through parameter ϕ_2).

The non-divergence condition is achieved if the expectations of $\Delta\kappa_t^{(1)}$ and $\Delta\kappa_t^{(2)}$ are equal when t tends to infinity. It can be shown that this condition is equivalent to requiring

$$\frac{\phi_0}{1-\phi_1-\phi_2} = \frac{\theta_0}{1-\theta_1-\theta_2}. \tag{6}$$

To ensure the estimated model meets the non-divergence condition, equation (6) is included in the estimation process as a parameter constraint.

3.3. VECM

The VECM approach is often used for modeling dynamics of multiple economic variables over time. It adds to the VAR model an error-correction term, a linear combination of the variables being modeled. When applied to economic modeling, the error-correction term defines the equilibrium relations between different economic variables. It is assumed that if the system is not in equilibrium, economic agents will react to the disequilibrium error until equilibrium is restored.

We can use the VECM approach to model $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$. Let us suppose that the long-term equilibrium relation between $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ is given by $a\kappa_t^{(1)} + b\kappa_t^{(2)} + c = 0$, where a, b and c are constants. According to Granger's Representation Theorem (Engle and Granger, 1987), there exists the following VECM specification for $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$:

$$\begin{aligned}
\Delta\kappa_t^{(1)} &= \rho^{(1)}(a\kappa_{t-1}^{(1)} + b\kappa_{t-1}^{(2)} + c) + \phi_0 + \phi_1\Delta\kappa_{t-1}^{(1)} + \phi_2\Delta\kappa_{t-1}^{(2)} + \epsilon_t^{(1)}, \\
\Delta\kappa_t^{(2)} &= \rho^{(2)}(a\kappa_{t-1}^{(1)} + b\kappa_{t-1}^{(2)} + c) + \theta_0 + \theta_1\Delta\kappa_{t-1}^{(1)} + \theta_2\Delta\kappa_{t-1}^{(2)} + \epsilon_t^{(2)}.
\end{aligned}$$

This specification is formed by adding an error-correction term $a\kappa_{t-1}^{(1)} + b\kappa_{t-1}^{(2)} + c$ to equations (4) and (5). Parameters $\rho^{(1)}$ and $\rho^{(2)}$ are called adjustment coefficients. They can be interpreted to mean the forces at which $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ are being pulled to the long-term equilibrium. As in the previous two modeling approaches, the innovations $\epsilon_t^{(1)}$ and $\epsilon_t^{(2)}$ have no serial correlation and follow a bivariate normal distribution with a zero mean vector and a constant covariance matrix.

To ensure non-divergence, over the long run, the difference between the expectations of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ has to be a constant. This is equivalent to requiring a long-term equilibrium of $\kappa_t^{(1)} - \kappa_t^{(2)} + c = 0$. Incorporating this constraint, the VECM specification can be written as follows:

$$\begin{aligned}
\Delta\kappa_t^{(1)} &= \rho^{(1)}(\kappa_{t-1}^{(1)} - \kappa_{t-1}^{(2)}) + \phi_0' + \phi_1\Delta\kappa_{t-1}^{(1)} + \phi_2\Delta\kappa_{t-1}^{(2)} + \epsilon_t^{(1)}, \\
\Delta\kappa_t^{(2)} &= \rho^{(2)}(\kappa_{t-1}^{(1)} - \kappa_{t-1}^{(2)}) + \theta_0' + \theta_1\Delta\kappa_{t-1}^{(1)} + \theta_2\Delta\kappa_{t-1}^{(2)} + \epsilon_t^{(2)},
\end{aligned}$$

where $\phi_0' = \phi_0 + c\rho^{(1)}$ and $\theta_0' = \theta_0 + c\rho^{(2)}$. As t tends to infinity, the expectation of $\kappa_t^{(1)} - \kappa_t^{(2)} + c$ tends to zero. It follows that the long-term expectations of $\Delta\kappa_t^{(1)}$ and $\Delta\kappa_t^{(2)}$ are equal, and that equation (6) is automatically satisfied.

3.4. Comments on the Three Modeling Approaches

In all three modeling approaches, the non-divergence hypothesis is satisfied. The resulting forecasts of $m_{x,t}^{(1)}$ and $m_{x,t}^{(2)}$ are guaranteed not to diverge over the long run.

The VAR and VECM approaches have a few advantages over the RWAR approach. The first and the most important advantage is that the model structures in both approaches are symmetric, which means the user does not need to make a subjective assumption about which population is dominant. The second advantage is that both approaches are able to capture cross-correlations between $\Delta\kappa_t^{(1)}$ and $\Delta\kappa_t^{(2)}$.

There are two desirable features that are unique to the VECM approach. First, relative to the VAR approach, the VECM approach is more intuitive, as it describes explicitly the

RWAR							
μ	-0.4378 (0.1213)	μ_Δ	-0.0259 (0.1234)	$\sigma^{(1)}$	0.6492 (0.0849)		
		ϕ	0.9396 (0.0690)	$\sigma^{(2)}$	0.5520 (0.1051)		
				ρ	0.2545 (0.2544)		
VAR							
ϕ_0	-0.6315 (0.1219)	θ_0	-0.6389 (0.1570)	$\sigma^{(1)}$	0.6233 (0.0963)		
ϕ_1	-0.1098 (0.2974)	θ_1	0.1571 (0.2635)	$\sigma^{(2)}$	0.6605 (0.1060)		
ϕ_2	-0.1890 (0.2317)	θ_2	-0.4713 (0.2945)	ρ	0.6259 (0.1418)		
VECM							
ϕ_0	-0.8016 (0.1868)	θ_0	-0.8200 (0.2130)	$\sigma^{(1)}$	0.5454 (0.0755)	$\rho^{(1)}$	-0.2078 (0.0848)
ϕ_1	-0.1559 (0.2542)	θ_1	0.1523 (0.2535)	$\sigma^{(2)}$	0.6408 (0.1062)	$\rho^{(2)}$	-0.1050 (0.0863)
ϕ_2	-0.3007 (0.2211)	θ_2	-0.5391 (0.2656)	ρ	0.6145 (0.1525)		

Table 1: RWAR, VAR and VECM parameter estimates. Standard errors are shown in parentheses.

long-term equilibrium of the system. Second, the VECM approach incorporates more information in the modeling process, because it includes not only the first differences but also the original values of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$. By incorporating more information, the VECM approach may potentially lead to a significantly better goodness-of-fit.

4. Comparing the Fitted Models

4.1. Estimation

We now fit the three stochastic processes to the values of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ we obtained in Section 2. For each process, parameter estimates are obtained by maximizing the log-likelihood function, which is derived from the fact that $(\epsilon_t^{(1)}, \epsilon_t^{(2)})$ follows a bivariate normal distribution with a zero mean vector and a constant covariance matrix. The log-likelihood functions for the three processes are provided in the Appendix.

The estimated parameters are shown in Table 1. In the table, ρ denotes the correlation coefficient between $\epsilon_t^{(1)}$ and $\epsilon_t^{(2)}$, and $\sigma^{(1)}$ and $\sigma^{(2)}$ denote the standard deviations of $\epsilon_t^{(1)}$ and $\epsilon_t^{(2)}$, respectively.

The estimation of these three stochastic processes is conditioned on the estimates of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ in equation (1). It follows that when we calculate standard errors of the RWAR, VAR and VECM parameter estimates, some correction is necessary to account for the uncertainty surrounding the estimates of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$. We make the necessary correction on the basis of the theoretical results in Murphy and Topel (2002). The standard error of each parameter estimate is displayed in Table 1.

In the following sub-sections, we evaluate the goodness-of-fit, forecasting performance and robustness of the three modeling approaches. We focus on the time-series processes for $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ only, because all three modeling approaches share the same Lee-Carter

Model	RWAR	VAR	VECM
Maximized log-likelihood, $\hat{l}$	−78.2448	−73.1794	−66.6272
Effective number of parameters, k	6	8	11
Effective sample size, n	44	43	43
BIC	179.1948	176.4485	174.6276

Table 2: The values of BIC for the three estimated processes.

parameters.

4.2. Goodness-of-fit

We now provide, on the basis of the maximized log-likelihood, $\hat{l}$, an evaluation of the goodness-of-fit. A higher value of $\hat{l}$ generally means a better fit to historical data. However, under the principle of parsimony, one should use the least possible number of parameters for an adequate representation. A fairer comparison between the estimated models should be based on a selection criterion that takes into account of not only the maximized log-likelihood but also the number of parameters. We consider the Bayesian Information Criterion (BIC) (Schwarz, 1978), which is defined by

$$BIC = k \ln(n) - 2\hat{l},$$

where k is the effective number of parameters and n is the effective sample size. A model is more desirable if it gives a good fit (large $\hat{l}$) and is parsimonious (small k). Hence, we prefer the model with the smaller value of BIC. The values of BIC for the three estimated processes are displayed in Table 2.[7] The VECM approach yields the smallest value of BIC.

Next, we analyze the standardized residuals. They are calculated by dividing the crude residuals with their corresponding standard deviations. If the fit is adequate, the standardized residuals should exhibit no trend over time. Figures 3 and 4 show the standardized residuals for populations 1 and 2, respectively. We observe significant downward trends in the standardized residuals generated from RWAR and VAR models. By contrast, VECM generates standardized residuals with less apparent trends, indicating that it gives a more adequate fit.

4.3. Forecasting

In this subsection, we evaluate the models' forecasting performance by comparing their in- and out-of-sample forecasts.

4.3.1. In-sample forecasts

In-sample forecasts are obtained as follows. The three time-series processes are fitted to $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ for years 1961 to 1990. The fitted models are then used to obtain 'forecasts' of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ for years 1991 to 2005. These 'forecasts' are compared against the actual values of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ for years 1991 to 2005. The results are shown graphically in Figure 5.

[7] In fitting the Lee-Carter model, we used data for the period of 1961-2005. Hence, we have 45 pairs of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ in total. The log-likelihood for RWAR is conditioned on the the first pair of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$, which means for RWAR, $n = 45 - 1 = 44$. The log-likelihoods for VAR and VECM are conditioned on the first pair of $\Delta\kappa_t^{(1)}$ and $\Delta\kappa_t^{(2)}$, which means for VAR and VECM, $n = 45 - 2 = 43$. Readers are referred to the Appendix for further information about the log-likelihoods for the models.

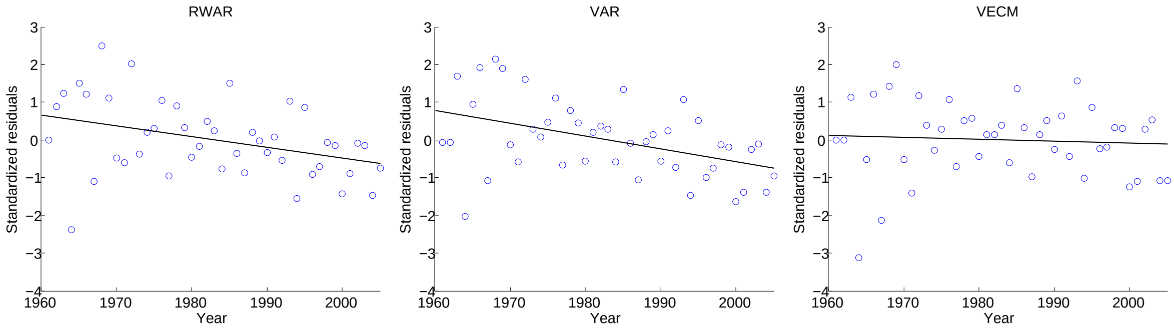

Figure 3: Standardized residuals of the estimated RWAR, VAR and VECM models, population 1. The solid lines are the least squares fits to the standardized residuals.

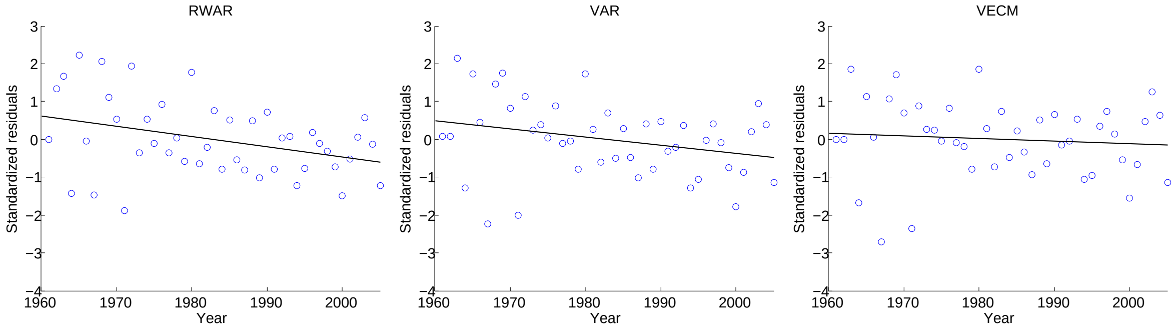

Figure 4: Standardized residuals of the estimated RWAR, VAR and VECM models, population 2. The solid lines are the least squares fits to the standardized residuals.

We observe that the RWAR approach gives the least desirable result. The mean forecasts for both populations are significantly biased high. After year 2000, the 95% confidence intervals no longer contain the actual values. The VAR approach gives a similar result, but it leads to a smaller bias for population 2. This improvement may be attributed to its symmetric structure, which spares us from the need of assuming $\kappa_t^{(2)}$ will always follow $\kappa_t^{(1)}$.

The VECM approach captures a small part of the recent acceleration in reductions of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$. The resulting forecasts are still biased high, but the bias is less severe. The 95% confidence intervals, although narrower, are able to capture all true values over the entire period of 1991 to 2005.

4.3.2. Out-of-sample forecasts

Out-of-sample forecasts are produced using the entire data set that covers years 1961 to 2005. The results are shown graphically in Figure 6. Among all three approaches, the VECM approach yields forecasts that are the most in line with the recent trends.

4.4. Robustness

In this subsection, we perform three tests to compare the robustness of the three modeling approaches.

The first test investigates how forecasts will change if a different look-back window is used. In this test, we fit the three models to two different look-back windows: 45 years (1961-2005) and 30 years (1976-2005). We then produce, for each model and each look-

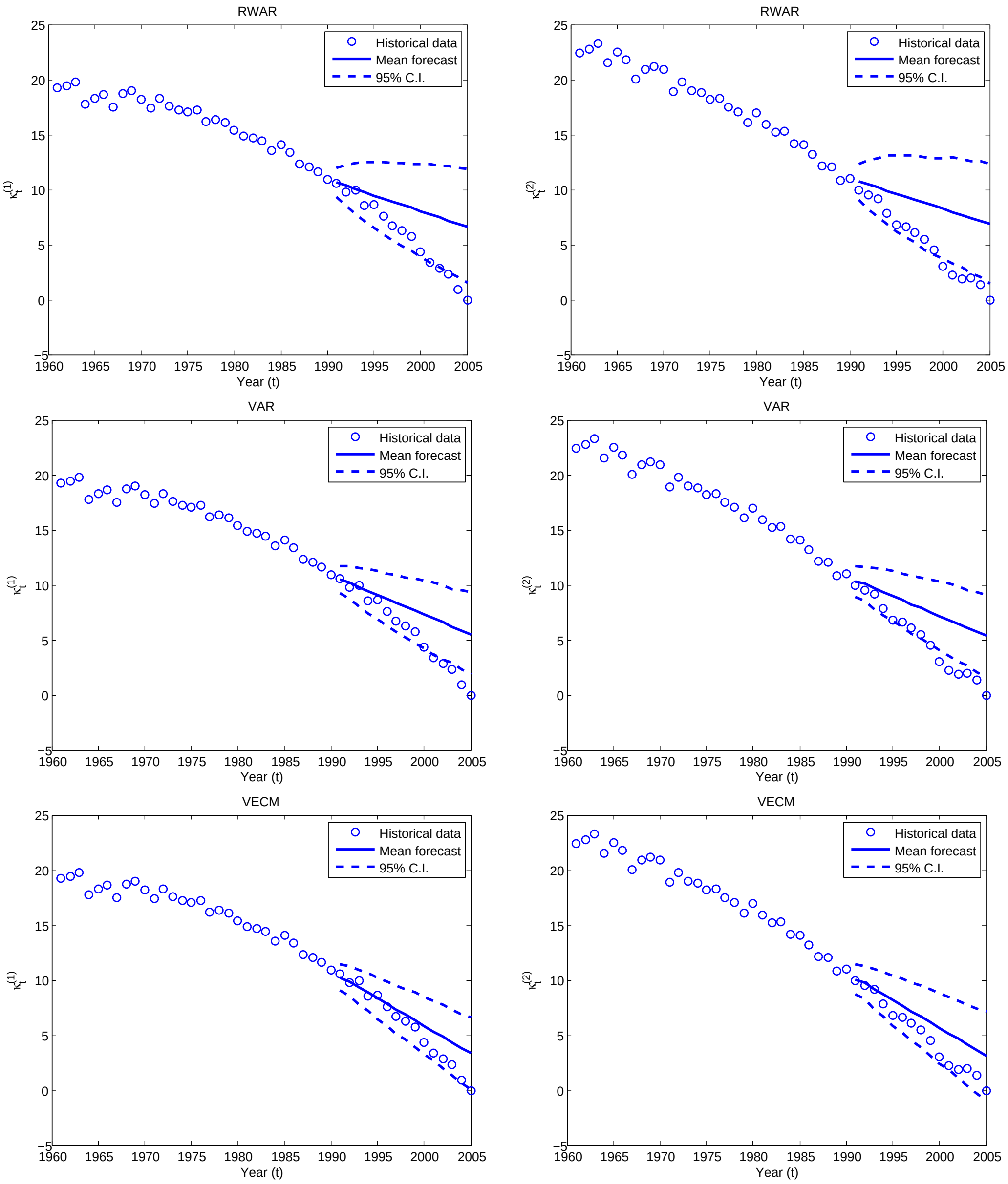

Figure 5: In-sample forecasts of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$, based on the RWAR, VAR and VECM models that are fitted to a restricted sample period of 1961-1990, both populations.

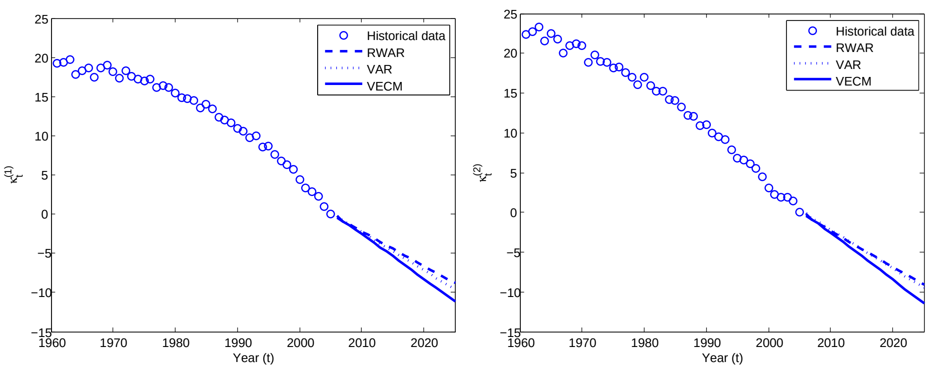

Figure 6: Out-of-sample forecasts of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$, based on the RWAR, VAR and VECM models that are fitted to the entire sample period of 1961-2005, both populations.

back window, mean and 95% interval forecasts of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$. The results are presented in Figure 7. In the diagram, the solid lines represent forecasts based on a 45-year look-back window, while the dashed lines represent forecasts based on a 30-year look-back window. If a model is robust, the two look-back windows should generate similar forecasts.

We observe from Figure 7 that none of the three modeling approaches is completely robust with respect to a change in the look-back window. The lack of robustness is the most severe when the RWAR approach is used. The problem is less severe when the VAR or VECM approach, which is based on a symmetric model structure, is used. Among all three modeling approaches, the VECM approach appears to be the most robust.

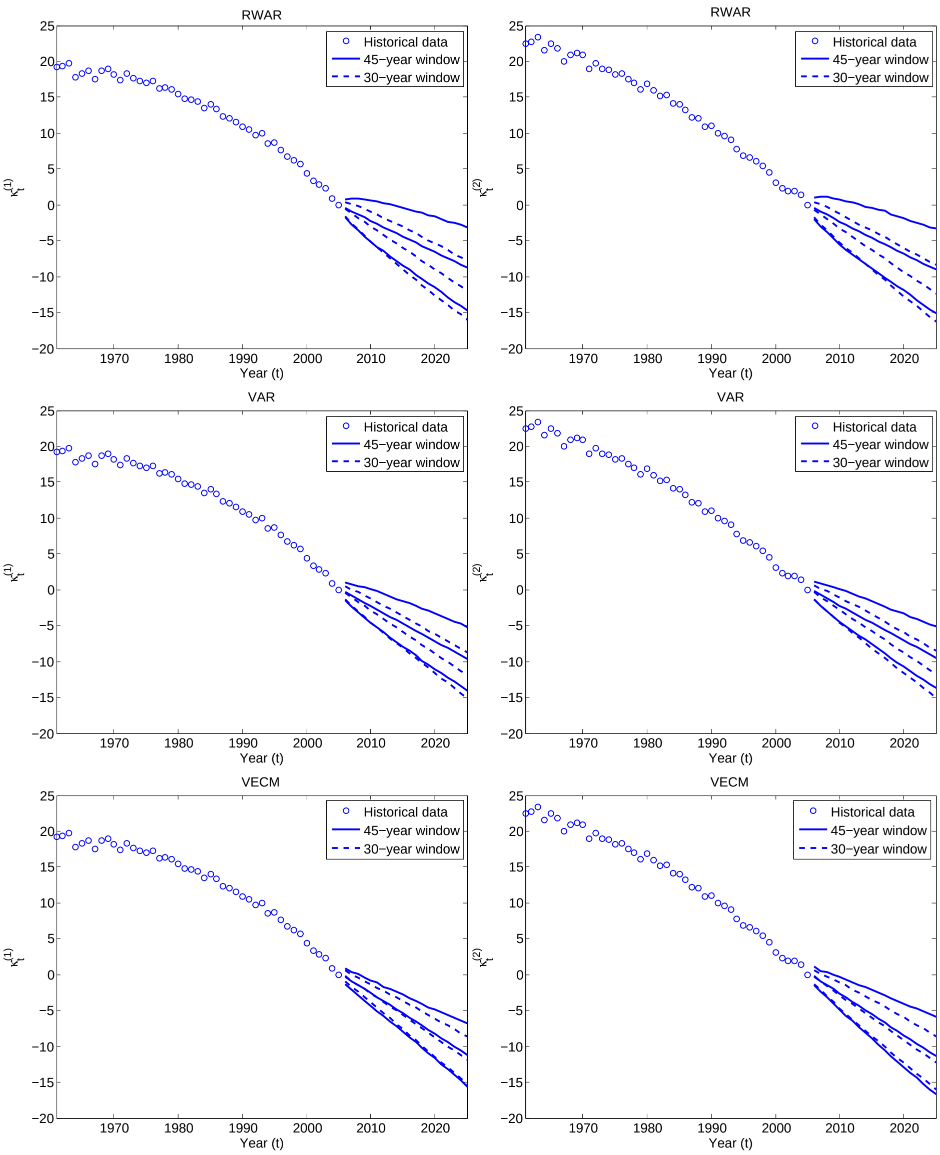

Figure 7: Forecasts of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$, based on look-back windows of 45 years and 30 years. For each forecast, the middle line represents the mean forecast, while the two lines surrounding the middle line represent the 95% interval forecast.

The second test examines how forecasts will change if new data are incorporated into the estimation process. In this test, we mimic the situation when we update a mortality model with new data. Specifically, we fit the models to data over three rolling sample periods, namely (1) 1961-1990, (2) 1966-1995 and (3) 1971-2000. For each fitted model, forecasts of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ beyond the sample period are made. The results of this test are depicted in Figure 8. The dotted, dashed and solid lines represent mean forecasts that are based on sample periods (1), (2) and (3), respectively. For ease of exposition, the 95% interval forecasts are not shown. If a model is robust, then the three mean forecasts should be close the one another.

Again, none of the three models is completely robust with respect to shifts in the sample period. The gaps between the three forecasts produced by the RWAR approach are very wide. In addition, the RWAR forecasts that are based on the two earlier sample periods are rather far from the actual values $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$. The problem is less severe if the VECM or VAR approach is used. Among all three approaches, the VECM approach yields mean forecasts that are closest to one another and are closet to the actual values.

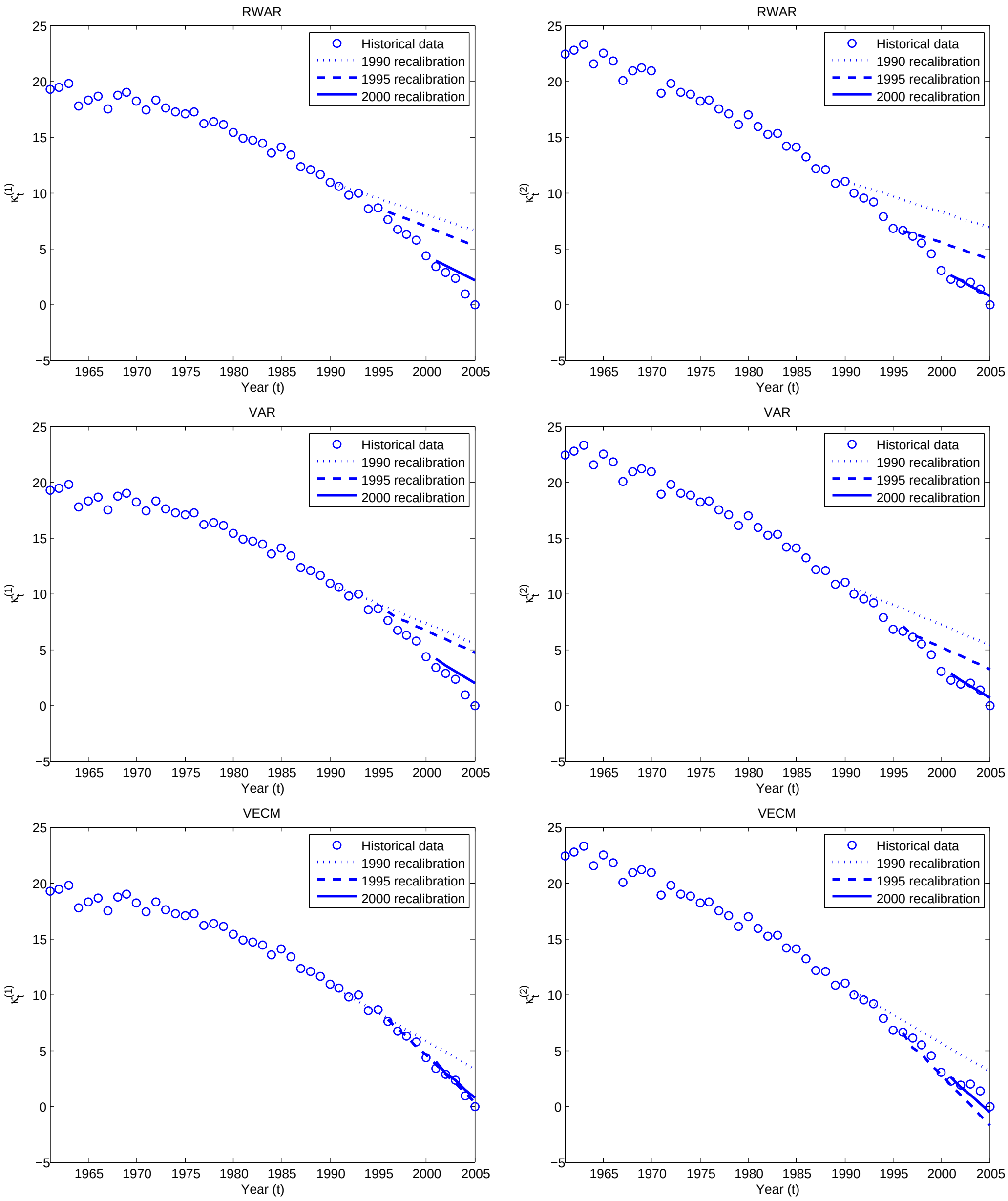

Figure 8: Forecasts of $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$, based on models that are fitted to three different rolling sample periods, (1) 1961-1990, (2) 1966-1995 and (3) 1971-2000, both populations

	Assumed dominant population	
	EW males	UK insured lives
Long-term drift	-0.4378	-0.5094
25-year forecast of κ_t for EW males	-10.9104	-12.8905
25-year forecast of κ_t for UK insured lives	-11.2374	-12.7528
50-year forecast of κ_t for EW males	-21.9316	-25.6862
50-year forecast of κ_t for UK insured lives	-22.3152	-25.5393

Table 3: RWAR forecasts based on different assumed dominant populations

The third test studies how the forecasts will change if we set population 1 to UK insured lives instead of English and Welsh males. Swapping the two populations will not result in any change to the VAR and VECM forecasts, because they are based on symmetric model structures. However, in using the RWAR approach, setting population 1 to UK insured lives (i.e., assuming that UK insured lives is the dominant population) will lead to a different forecast. Table 3 summarizes the effect of changing the assumed dominant population on RWAR forecasts. Assuming the dominant population is UK insured lives leads to a long-term drift with a larger magnitude, implying faster mortality improvements for both populations in the long run. For instance, the change in the assumed dominant population will lead to a 17% reduction in the 50-year forecast of κ_t for the population of English and Welsh males.

5. Applications to Solvency II

In this section, we apply the models discussed in this paper (RWAR, VAR and VECM) to a particular kind of financial instrument, longevity swap reinsurance. We study whether multi-population models can be used to measure longevity basis risk and which impact the choice of model has on the pricing of such financial instruments. In the first subsection, we define longevity basis risk and our proposal for measuring it. This is followed in the second subsection by a brief introduction to longevity swap reinsurance transactions and their pricing. The basic concepts of the European Solvency II as they apply to our example are outlined in the third subsection, with an overview of the calculation methods and results in the fourth and fifth subsections.

5.1. Longevity Basis Risk

Longevity basis risk in general arises from a mismatch between the mortality experience over time of two different populations. For an institution exposed to longevity risk, that is, the financial risk associated with greater than expected payouts to pensioners or annuitants due to longer than expected survival of these annuitants or pensioners, longevity basis risk can arise, when the institution purchases a hedging instrument based on population mortality, while the institution's actual exposure is to the survival of a particular portfolio of lives, which may not share the same mortality characteristics as the general population. For instance, the Continuous Mortality Investigation Bureau (2009) has demonstrated that the UK insured lives mortality is lower and has historically improved at a greater rate than the England and Wales general population mortality.

Longevity basis risk also arises for the hedge provider in our example above, when they provide a longevity hedge corresponding to the exact development of a specific portfolio, but for certain reasons have decided not to take into account the exact mortality trend of the insured portfolio. In other words, the hedge provider has priced the longevity hedge using general population mortality improvements as an approximation for the underlying mortality trend.

For the purpose of this analysis, we define longevity basis risk to be the risk associated with mispricing a longevity swap, which occurs when the hedge provider uses the general population mortality trend rather than a trend calibrated to the specific mortality experience of the portfolio of lives to be hedged. We measure longevity basis risk as the difference between the price of a longevity swap as predicted by a multi-population mortality model, and the price obtained by applying the projected general population (population 1) mortality trend to the current mortality of the sub-population (population 2). In our example, the general population mortality forecast is generated from England and Wales mortality data, and the sub-population is represented by UK insured lives mortality experience for the same period.

5.2. Longevity Swap Reinsurance

In the UK, a form of indemnity contract has become prevalent between insurance companies or pension schemes and reinsurers, which transfers longevity risk exposure from one party to another, but which does not require a transfer of the assets backing the annuity or pension liabilities. This type of contract is termed longevity swap reinsurance, because the hedge purchaser ("ceding company") exchanges uncertain future pension or annuity cash-flows for a set of fixed cash-flows. The fixed cash-flows are also referred to as the reinsurance premiums or the "fixed-leg" of the longevity swap, and are typically equal to a set of best estimate annuity (or pension) payouts plus a margin.

There are numerous ways a reinsurer may determine its required margin to take on the longevity risk. For the sake of having an objective comparison, we have chosen the risk margin as defined within the European Solvency II framework as a measure for the reinsurer's cost of capital. An illustration of a sample longevity swap reinsurance contract is shown in Figure 9. The figure shows projected best-estimate annuity cash-flows for a 65-year-old male, the reinsurance premiums, which are equal to the projected best estimate cash-flows plus a margin to cover the risk. The figure also displays the so-called Solvency Capital Requirements (SCR) for longevity risk according to Solvency II, which will be explained in the next subsection.

5.3. Solvency II

Despite the fact that the implementation and applicability of the European Solvency II framework is the cause of much debate in the insurance industry and beyond, the general concepts which it puts forward for measuring risk are well-publicized and much studied (see, e.g., Börger, 2010; Richards et al., 2012). Therefore, the framework lends itself well to studying the interaction of risk and pricing from the perspective of an insurance or reinsurance company.

5.3.1. Solvency Capital Requirements (SCR)

Under Solvency II, an insurance company will have to demonstrate its solvency by showing that it possesses sufficient "own funds" to cover at least the Solvency Capital Re-

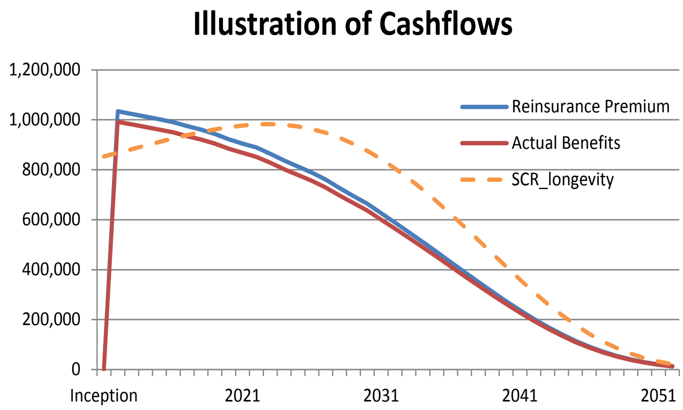

Figure 9: Illustration of a longevity swap reinsurance contract, males aged 65 years, annual benefits GBP 1 million.

quirements (SCR) which the company calculates for the various risks it has entered into.

Structurally, the total SCR is calculated as the diversified sum of all the SCRs calculated for each risk, including market risk, life risk and non-life risk. In this paper, we are interested only in the SCR for longevity risk, and therefore disregard all other risk modules, as well as operational risk.

The general framework for calculating the $SCR_{t=0}$ at time $t = 0$ for any risk category is to compare the expected position (net asset value) of the company at the end of the first year ($t = 1$) with its position under a stressed scenario. The stress is determined as the 99.5% Value-at-Risk, or 99.5th-percentile loss. The Solvency II framework gives a company the choice of either using the standard model, which is set out by the European Insurance and Occupational Pensions Authority (EIOPA), or applying for approval of a so-called internal model. For the standard model, the stress tests are simple and approximately calibrated to an average insurance company within the EEA.[8] Any internal or partial internal model will have to be calibrated on the company's own risk profile and will require a demonstration of effectiveness. In our case at hand, using a stochastic mortality model will facilitate calculating a 99.5th-percentile scenario.

We approximate the change in net asset value on the company's balance sheet as the difference between the value of liabilities under best estimate assumptions and the value of liabilities in the scenario corresponding to the 99.5th-percentile stress,

$$SCR_{t=0} = \Delta NAV_{t=1} = NAV_{t=1} - NAV_{t=1}^{stressed} \approx {}_{t=0}V^{stressed} - {}_{t=0}V^{best\,estimate},$$

where ${}_{t=0}V^{best\,estimate}$ is the present value of future liabilities calculated under best estimate assumptions, and ${}_{t=0}V^{stressed}$ is the present value of future liabilities under stressed assumptions. These present values should be discounted to time $t = 0$ using the risk-free interest rates mandated by EIOPA.[9] However, for simplicity's sake and in recognition of the current low interest environment, we have assumed a flat interest rate of 1.75% for all periods.

5.3.2. Risk Margin

The Solvency II risk margin[10] is a measure for the cost associated with an insurance

[8]European Economic Area, which incluces the member states of the EU as well as some additional countries.

[9]Here: apply GBP spot rates based on year-end 2009 (QIS5).

[10]Reference: Art. 77 of Directive 2009/138/EC

company having to hold at least the solvency capital requirements with respect to the business it underwrites. In this respect, the risk margin indirectly reflects the risk adversity of an insurance company, and the premium above best estimate future costs which it will require in order to take on additional risks. The Solvency II risk margin is defined as 6% of the present value of the SCRs for all future time periods:

$$Risk\,Margin = 6\% \times \sum_t SCR_t\, d_t,$$

where d_t is the risk-free discount factor for period t.[11] The cost-of-capital factor of 6% reflects a market average spread over risk-free which insurance companies typically are required to earn on their equity. In other words, the profit margin, which an insurance company taking on risk must include in its insurance premiums, should typically be no less than the risk margin.

The theoretically correct method of calculating the risk margin requires a re-calculation of the SCR for all future years. We have applied an approximation, which is prevalent among practitioners. This assumes that the SCR for longevity risk runs off proportionately to the present value of liabilities.

5.4. Calculation

For each of the proposed multi-population models, we generate 10,000 stochastic scenarios for a single model point, that is, a 65-year-old male annuitant without dependents, projecting this group's mortality forward starting in 2006, the year after the end of our data period. We also use the models to generate the best estimate period mortality for both the insured populations in 2006.

Note that our data only includes the age range 65 to 84. Therefore, it is necessary to extrapolate the death rates, which we do by applying the simple Gompertz law of mortality.

We rank the scenarios based on the annuity value for a 65-year-old, calculated based on the extrapolation and discounting mentioned above. The scenario with the 50th highest annuity value is chosen as the 99.5th-percentile scenario, and the median scenario as approximation for the best estimate.

Using these scenarios, we are able to project the cash-flows associated with a longevity swap, in particular, we can calculate the present value of future annuity payments (benefits), the corresponding SCRs and the risk margin as defined above, for each model variant.

Since our results are to be compared to a base scenario, which artificially includes longevity basis risk, we generate a single population projection for population 1 (England and Wales), and apply the resulting mortality improvement trend to the 2006 period mortality of population 2 (UK insured lives).[12] We denote this the single population model for distinction from the multi-population models.

5.5. Results

From Table 4, which shows the results of our analysis, we observe that switching from a single-population projection to a multi-population model has only a very limited impact on

[11] According to Solvency II rules, discounting should be carried out based on a current risk-free yield curve. We have simplified this assumption to a flat 1.75% for ease of comparison.

[12] The model used is a single-population Lee-Carter model, with a random walk with drift for the dynamics of the time-varying factor.

Model	Annuity Value	Percentage Increase	Risk Margin	Percentage Increase
Single Population	16,791,056		1,208,110	
RWAR	16,830,293	0.23%	1,777,254	47.15%
VAR	16,933,509	0.85%	1,314,367	8.80%
VECM	17,321,705	3.16%	1,528,809	26.55%

Table 4: Annuity values and risk margins calculated by the four models under consideration.

the best estimate annuity value for a 65-year-old. The largest increase seen for the VECM model ties in with our previous observations that this model projected lower mortality rates, and was thus less likely to overestimate mortality than the other models.

In contrast, there is a substantial change in risk margin, when one "upgrades" from a single-population model to a multi-population model. Note that the risk margin reflects the amount of uncertainty associated with the future annuity cash flows, as represented by the 99.5th-percentile stressed scenario. Using the single population variance in stochastic projections clearly underestimates the uncertainty, which we expect to be higher in the case of a smaller sub-population.

We note that our stochastic scenarios do not include parameter uncertainty or model risk, and therefore likely understate the risk margin. Nevertheless, the relative increase in risk margin already reflects the need to model the smaller population trend explicitly and not rely on its presupposed similarity with the general population trend.

From Table 4 we also observe that the risk margins implied by RWAR, VAR and VECM are somewhat different. This observation suggests that the choice of the bivariate time-series process for $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ does have an impact on the pricing of longevity swap insurance. Practitioners are therefore recommended to choose a process that is the most appropriate, using validation methods we presented in Section 4.

Another interesting observation from Table 4 is that the VAR process does not require such a high mark up from the single population approach. This is because for this data set the volatility of $\kappa_t^{(2)}$ implied by the VAR process is the lowest among the three bivariate processes. For instance, when $t = 2025$, that is, 20 years after the end of the data period, the standard deviations of $\kappa_t^{(2)}$ generated from RWAR, VAR and VECM are, 2.95, 2.16 and 2.70, respectively.

We warn the reader that the stochastic scenarios discussed here included neither parameter uncertainty nor recalibration risk (see Cairns, 2011). The mark-up required by Solvency II for operational risk has also been disregarded. Actual reinsurance prices should therefore be substantially higher, even if reinsurers will be able to take into account some diversification benefits with their mortality risk portfolios as well as more realistic interest rate assumptions.

6. Conclusion

In this paper, we proposed two alternative approaches, VAR and VECM, for modeling the time-varying factors in a two-population stochastic mortality model. An important advantage of VAR and VECM over the existing RWAR approach is that they are based on a symmetric model structure, which waives the need of a subjective assumption about which of the two populations is dominant.

When applied to the data set we consider, VECM gives a better goodness-of-fit and forecasting performance than VAR and RWAR. In particular the VECM model seems to overcome in part the known deficiencies of the basic Lee-Carter model in its application to UK data showing cohort effects, and is not quite as likely to require future reserve strengthening as the RWAR for instance, due to its better robustness. Given the empirical results, VECM appears to be a reasonable approach for modeling the time-varying factors in a two-population stochastic mortality model.

We illustrated our idea with the Lee-Carter structure, because it is relatively simple and well-known to the insurance industry. Nevertheless, the two proposed modeling approaches can be applied readily to structures other than the Lee-Carter. For instance, we may also use VAR and VECM to model the time-varying factors in the model structures that are documented in Cairns et al. (2009) and Dowd et al. (2011).

We used RWAR, VAR and VECM to price a simple longevity swap reinsurance contract. To evaluate longevity basis risk, we compared the prices calculated from these models with the price that is based on the mortality improvement trend projected by a single-population model that is fitted to data from the general population. We demonstrated that when comparing only best estimate benefit projections, longevity basis risk may seem of little relevance. However, longevity basis risk is shown as having a substantial impact on the risk-adjusted price, which we modeled here as the risk margin according to European Solvency II guidelines. We also found that RWAR, VAR and VECM yield quite different risk margins. This result suggests that practitioners should be careful in choosing a process of the time-varying factors in a two-population mortality model.

In using RWAR, VAR or VECM, it is assumed implicitly that the difference in mortality between two related populations will begin reverting to its long-term equilibrium immediately at the forecast origin. This assumption is quite debatable, as a divergence in mortality between different socio-economic groups has been observed in recent years (see, e.g., Mackenbach, 2003; Waldron, 2007). It would be interesting to study in future research how this assumption may be avoided. This might be achieved by permitting mean-reversion to begin at random future time.

Mortality dynamics are subject to jumps, which arise from, for example, a breakthrough in treating a certain disease. These jumps affect pension payouts as well as the pricing of mortality-linked securities. Another possible avenue for future research is the incorporation of jumps into VAR and VECM. The extension can be made flexible so that it can capture the possibility that an interruptive event may affect different populations at different time points.

References

[1] Börger, M. (2010). Deterministic Shock vs. Stochastic Value-at-Risk – An Analysis of the Solvency II Standard Model Approach to Longevity Risk. Blätter der DGVFM 31(2): 225-259.

[2] Cairns, A.J.G. (2011). Robust Hedging of Longevity Risk. Paper presented at the Seventh International Longevity Risk and Capital Markets Solutions Conference, Frankfurt, Germany.

[3] Cairns, A.J.G., 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.

[4] Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D. and Khalaf-Allah, M. (2011). Bayesian Stochastic Mortality Modelling for Two Populations. ASTIN Bulletin 41: 29-55.

[5] Continuous Mortality Investigation Bureau (2009). A Prototype Mortality Projections Model: Part Two – Detailed Analysis. Continuous Mortality Investigation Working Paper 39.

[6] Dowd, K., Cairns, A.J.G., Blake, D., Coughlan, G.D., Epstein, D. and Khalaf-Allah, M. (2011). A Gravity Model of Mortality Rates for Two Related Populations. North American Actuarial Journal 15: 334-356.
[7] Engle, R. F. and Granger, C. W. J. (1987). Co-integration and Error Correction: Representation, Estimation and Testing. Econometrica 55: 251-276.
[8] Human Mortality Database, University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany), 2012 (www.mortality.org or www.humanmortality.de).
[9] Jarner, S.F. and Kryger, E.M. (2011). Modelling Adult Mortality in Small Populations: The SAINT Model. ASTIN Bulletin 41: 377-418.
[10] Lee, R. and Carter, L. (1992). Modeling and Forecasting U.S. Mortality. Journal of the American Statistical Association 87: 659-671.
[11] Lee, R. and Miller, T. (2001). Evaluating the Performance of the Lee-Carter Method for Forecasting Mortality. Demography 38: 537-549.
[12] Li, J.S.H. and Hardy, M.R. (2011). Measuring Basis Risk in Longevity Hedges. North American Actuarial Journal 15: 177-200.
[13] Li, J.S.H., Hardy, M.R. and Tan, K.S. (2010). Developing Mortality Improvement Formulae: The Canadian Insured Lives Case Study. North American Actuarial Journal 14: 381-399.
[14] Li, J.S.H., Hardy M.R. and Tan, K.S. (2009). Uncertainty in Mortality Forecasting: An Extension to the Classical Lee-Carter Approach. ASTIN Bulletin 39: 137-164.
[15] Li, N. and Lee, R. (2005). Coherent Mortality Forecasts for a Group of Population: An Extension of the Lee-Carter Method. Demography 42: 575-594.
[16] Mackenbach, J.P., Bos, V., Andersen, O., Cardano, M., Costa, G., Harding, S., Reid, A., Hemstrom, O., Valkonen, T. and Kunst, A.E. (2003). Widening Socioeconomic Inequalities in Mortality in Six Western European Countries. International Journal of Epidemiology 32: 830-836.
[17] Murphy, K.M. and Topel, R.H. (2002). Estimation and Inference in Two-Step Econometric Models. Journal of Business and Economic Statistics 20: 88-97.
[18] Richards, S.J., Currie, I.D. and Ritchie, G.P. (2012). A Value-at-Risk Framework for Longevity Trend Risk. A discussion paper presented to The Institute and Faculty of Actuaries.
[19] Schwarz, G. (1978). Estimating the Dimension of a Model. Annals of Statistics 6: 461-464.
[20] Waldron, H. (2007). Trends in Mortality Differentials and Life Expectancy for Male Social Security-Covered Workers, by Average Relative Earnings. ORES Working Paper Series Number 108.
[21] Wilmoth, J.R. (1993). Computational Methods for Fitting and Extrapolating the Lee-Carter Model of Mortality Change. Technical Report. Department of Demography. University of California, Berkeley.
[22] Zhou, R., Li, J.S.H. and Tan, K.S. (2011). Pricing Standardized Mortality Securitizations: A Two-Population Model with Transitory Jump Effects. Paper presented at the Seventh International Longevity Risk and Capital Markets Solutions Conference, Frankfurt, Germany.

Appendix

In this appendix, we present the log-likelihood functions for RWAR, VAR and VECM. Let us define the following notation:

- Θ: the parameter vector of a model
- bvnpdf(s, u, v): the probability density function, evaluated at s, for a bivariate normal random vector with mean vector μ and covariance matrix v

- $f(\cdot|\cdot)$: a generic conditional density function, where the conditioning variables follow the vertical bar
- V: the covariance matrix of the innovations $(\epsilon_t^{(1)}, \epsilon_t^{(2)})$
- $S_t = \begin{pmatrix} \kappa_t^{(1)} \\ \kappa_t^{(1)} - \kappa_t^{(2)} \end{pmatrix}$ and $\Delta_t = \begin{pmatrix} \Delta\kappa_t^{(1)} \\ \Delta\kappa_t^{(2)} \end{pmatrix}$

For RWAR, the conditional log-likelihood function is given by

$$\begin{aligned}
& \ln f(S_{t_0+1}, S_{t_0+2}, \ldots, S_{t_1}|S_{t_0}, \Theta) \\
= & \sum_{i=t_0}^{t_1-1} \ln f(S_{i+1}|S_i, \Theta) \\
= & \sum_{i=t_0}^{t_1-1} \ln \text{bvnpdf}\left(S_{i+1}, \begin{pmatrix} \mu \\ \mu_\Delta \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & \phi \end{pmatrix} S_i, V_1 \right).
\end{aligned}$$

For VAR, the conditional log-likelihood function is given by

$$\begin{aligned}
& \ln f(\Delta_{t_0+2}, \ldots, \Delta_{t_1}|\Delta_{t_0+1}, \Theta) \\
= & \sum_{i=t_0+1}^{t_1-1} \ln f(\Delta_{i+1}|\Delta_i, \Theta) \\
= & \sum_{i=t_0+1}^{t_1-1} \ln \text{bvnpdf}\left(\Delta_{i+1}, \begin{pmatrix} \phi_0 \\ \theta_0 \end{pmatrix} + \begin{pmatrix} \phi_1 & \phi_2 \\ \theta_1 & \theta_2 \end{pmatrix} \Delta_i, V_2 \right).
\end{aligned}$$

For VECM, the conditional log-likelihood function is given by

$$\begin{aligned}
& \ln f(\Delta_{t_0+2}, \ldots, \Delta_{t_1}|\Delta_{t_0+1}, \kappa_{t_0}^{(1)}, \kappa_{t_0}^{(2)}, \Theta) \\
= & \sum_{i=t_0+1}^{t_1-1} \ln f(\Delta_{i+1}|\Delta_i, \kappa_{t_0}^{(1)}, \kappa_{t_0}^{(2)}, \Theta) \\
= & \sum_{i=t_0+1}^{t_1-1} \ln \text{bvnpdf}\left(\Delta_{i+1}, \begin{pmatrix} \phi_0 \\ \theta_0 \end{pmatrix} + \begin{pmatrix} \rho^{(1)} \\ \rho^{(2)} \end{pmatrix} (\kappa_{t-1}^{(1)} - \kappa_{t-1}^{(2)}) + \begin{pmatrix} \phi_1 & \phi_2 \\ \theta_1 & \theta_2 \end{pmatrix} \Delta_i, V_3 \right).
\end{aligned}$$