



HAL
open science

Forecasting Mortality Rate Improvements with a High-Dimensional VAR

Quentin Guibert, Olivier Lopez, Pierrick Piette

► **To cite this version:**

Quentin Guibert, Olivier Lopez, Pierrick Piette. Forecasting Mortality Rate Improvements with a High-Dimensional VAR. 2017. hal-01613050

HAL Id: hal-01613050

<https://hal.science/hal-01613050>

Preprint submitted on 9 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Forecasting Mortality Rate Improvements with a High-Dimensional VAR

Quentin Guibert^{*1,2}, Olivier Lopez^{†3}, and Pierrick Piette^{‡1,3,4}

¹Univ Lyon, Université Claude Bernard Lyon 1, Institut de Science Financière et d'Assurances (ISFA),
Laboratoire SAF EA2429, F-69366, Lyon, France.

²Prim'Act, 42 avenue de la Grande Armée, 75017 Paris, France

³Univ. Pierre et Marie Curie Paris VI, 4 place Jussieu, 75005 Paris, France.

⁴Sinalys, 6 rue de Téhéran, 75008 Paris, France.

September 29, 2017

Abstract

Forecasting mortality rates is a problem which involves the analysis of high-dimensional time series, especially in multi-populations modeling. Most of usual mortality models propose to decompose the mortality rates into several latent factors to reduce this complexity. These approaches, in particular those using cohort factors, have a good fit, but they are less reliable for forecasting purposes. One of the major challenges is to determine the spatial-temporal dependence structure between mortality rates given a relatively moderate sample size. This paper proposes a large vector autoregressive (VAR) model fitted on the differences in the log-mortality rates, ensuring the existence of long-run relationships between mortality rate improvements. Our contribution is threefold. First, sparsity, when fitting the model, is ensured by using high-dimensional variable selection techniques without imposing arbitrary constraints on the dependence structure. The main interest is that the structure of the model is directly driven by the data, in contrast to the main mortality forecasting models. Hence, this approach is more versatile and would provide good forecasting performance for any considered population. Additionally, our estimation allows a one-step procedure, as we do not need to estimate hyper-parameters. The variance-covariance matrix of residuals is then estimated through a parametric form. Secondly, our approach can be used to detect nonintuitive age dependence in the data, beyond the cohort effect which is captured by our model. Third, our approach is natural to model the several populations in long run perspectives. Finally, in an out-of-sample forecasting study for mortality rates, we obtain rather good performances and more relevant forecasts compared to classical mortality models using the French, US and UK data. We also show that our results enlighten the so-called cohort effect for these populations.

Keywords: Mortality forecasting; High-dimensional time series; Vector Autoregression; Elastic-Net; Variance-covariance estimation; Modeling multiple populations.

*Email: q.guibert@hotmail.fr.

†Email: olivier.lopez0@upmc.fr.

‡Email: pierrick.piette@gmail.com.

1 Introduction

Identifying patterns in the mortality dynamics of a population is a hard task due to the complex underlying phenomena that impact the death rates. This problem is of crucial interest for government policies, pension funds and insurance companies. A wide range of models has been developed since the introduction of the famous model proposed by Lee and Carter (1992). Most of these approaches rely on time-series modeling with past data and forecast the main factors influencing the force of mortality, see among other Booth et al. (2002), Brouhns et al. (2002), Cairns et al. (2006), Cairns et al. (2009), Renshaw and Haberman (2008), Plat (2009), and Hunt and Blake (2014). Some reviews are available in the literature, see e.g. Booth and Tickle (2008), Cairns et al. (2008), and Barriau et al. (2012). The purpose of the present paper is to provide a new flexible modeling for the evolution of the mortality. Our high-dimensional vector autoregressive (VAR) approach, combined with an elastic-net penalty estimation method, aims to capture complex demographic effects without imposing a too restricting shape for the dynamics.

In recent years, some advances have been made to improve the forecast of mortality rates compared to the traditional factor-based models inspired by Lee and Carter (1992). This innovation has been provoked in particular by practitioners need for managing longevity risk and responding to the Solvency II requirements in insurance. Indeed, traditional models, even when a cohort effect is considered, have a reasonable fit, but poorer forecasts, indicating that these models may overfit the data. In such a context, one of the major concerns is to avoid the divergence of mortality rates between adjacent ages and different countries. Such inconsistency in the forecasting is pointed out for example by Börger et al. (2014), who explain that these low performances are due to the fact that traditional models mainly focus on the central trajectory projection. Several directions have been explored to overcome this issue. Li et al. (2013) develop an approach letting the age coefficients rotate over time, based on an expert judgment. Hunt and Villegas (2015) add an additional constraint on the cohort effect extensions of the Renshaw and Haberman (2006) model to overcome the convergence and robustness issues induced by the two-stage fitting algorithm for parameters. Regarding mortality trends of multiple populations, a relatively wide literature is organized around the idea of a biological convergence at a long horizon, see Dowd et al. (2011), Jarner and Kryger (2011), Li and Lee (2005), Enchev et al. (2016), and Cairns et al. (2016a) among others. These approaches estimate the mortality model by bringing together the data of several countries. Recently, Bohk-Ewald and Rau (2017) propose to approach the turning points of the mortality problem by combining trends of several countries.

Other approaches focusing on the age-period dependency have recently been proposed with the particularity of being more data-driven. Christiansen et al. (2015) use spatial statistics to forecast age-period mortality rate improvements using a kriging method. Their approach is parsimonious and provides good performances for short-term projection. However, it seems that their long-term results are more questionable. Doukhan et al. (2017) also focus on the surface of mortality improvements and model it parsimoniously with an AR-ARCH specification for a random field memory model. A valuable feature of their approach is that both dependencies between cohorts and the conditional heteroscedasticity of mortality are taken into account. Although they have good forecasting results, it is difficult to justify the size of the neighborhoods used to specify the memory process. Li and Lu (2017) choose a VAR process to consider the spatial dependence of mortality rates between neighboring ages adapted to short-term and long-term perspectives. These authors account for sparsity and stationarity in their VAR model by constraining the shape of the Granger causality matrix (Granger, 1969) as a lower triangular. Their model is also able to consider multiple populations.

In this paper, we propose an alternative approach that newly forecasts the age-period dependency using a large VAR specification on the log-mortality improvements. Although a VAR model is suitable for mortality time-series and is able to capture both long-term relationships and short-term shocks (see e.g. Salhi and Loisel, 2017), it is difficult to estimate accurately such models using an ordinary least square (OLS) technique, as these series are highly correlated and histories of data are relatively short. To avoid overparameterization, existing forecasting approaches impose an *a priori* spatio-temporal dependency structure between mortality rates or mortality improvements, which implies that only some selected series can interact. In contrast, our main contribution is the introduction of an estimation framework allowing for a large and flexible VAR structure without excluding potentially relevant relationships. A great feature of such a VAR specification is that all classical mortality models could naturally be included in our specification, especially the so-called cohort effect as noted by Li and Lu (2017).

Following recent developments in economics and finance (Fan et al., 2011; Furman, 2014), we develop a penalized VAR method based on the elastic-net (Zou and Hastie, 2005), which allows to take into account the sparsity correctly. Indeed, such a VAR model has a sparse structure in high dimension, which requires an accurate estimation method for shrinking zero coefficients in the Granger causality matrix. Compared to a classical maximum likelihood estimation approach, the key idea behind the elastic-net is to incorporate a penalty, which constrains the parameters. This penalty is a combination of an \mathcal{L}_2 term (as in a ridge regression) to avoid ill-conditioning matrices, and an \mathcal{L}_1 term (as in a LASSO regression) to produce a sparse model. By sparse model, we mean that our data-driven automatic selection produces a model with a relatively small number of non-zero parameters. As noted by Furman (2014), this is an attractive alternative to Bayesian VAR procedures usually considered in an econometrical framework and developed for example by Hahn (2014) for multiple populations modeling. Indeed, such approaches require to introduce relevant priors and do not address the sparsity's issue. The residuals are modeled as a Gaussian vector where the variance-covariance matrix is described using a parametric form for parsimony purposes.

Similarly to Doukhan et al. (2017), but contrary to Li and Lu (2017), our approach models the log of mortality improvements rather than the log of mortality rates. Several empirical elements have been advanced in the recent literature showing the interest of mortality improvements. Haberman and Renshaw (2012) show that a dual approach based on improvement rates can be followed for usual mortality models. They generally obtain quite comparable (but often better) forecasting results with this alternative route for the Lee-Carter model and its variants. As also noted by Bohk-Ewald and Rau (2017), mortality improvements seem to be easier to analyze, which facilitates the identification of divergences in mortality. As our approach is highly flexible, we expect that it can better capture complex patterns of mortality improvements. Another argument is that mortality improvements are generally stationary (see e.g. Chai et al., 2013), which is required for projection as our approach, contrary to Li and Lu (2017), does not impose constraints for guarantying stationarity.

We compare our high-dimensional VAR model to two standard factor-based mortality forecasting models: the usual Lee-Carter model (Lee and Carter, 1992) and the M7 model developed by Cairns et al. (2009). Using the root mean squared error measure, we show that our approach leads to better fitting (in-sample) and forecasting (out-sample) of the mortality rate time series from the three countries we have focused our analysis on. Moreover, our data-driven model implies more stable errors over different countries while the standard factors models tend to have more variable predictive power depending on the considered population.

The remainder of this paper is organized as follows. In Section 2 we describe the VAR model we retained. The high-dimensional estimation of this model is then developed in Section 3. We present the data we used, different results that we obtained and a comparison to other standard mortality

models in Section 4. Finally, Section 5 proposes an extension of the VAR model to multi-population modeling, and Section 6 considers some ways of improvement and concludes.

2 A Vector Autoregression approach for mortality rate improvements

In this section, we introduce an econometric model to describe the mortality improvement dynamics jointly. The mortality models we introduce in the literature in Section 1 are initially based on an analysis of the main factors explaining a common trend of mortality rates. Conversely, our approach only imposes an autoregressive structure, which encountered these classical models¹, as shown for example by Salhi and Loisel (2017) or Li and Lu (2017).

Throughout this paper, we focus on the time series $y_{i,t} = \ln(m_{i,t})$, where $m_{i,t}$ is the crude annual death rate at age i and at date t . These rates can be easily computed thanks to annual risk exposures and count of deaths for a country of interest. Those series are usually not stationary, as a trend can be observed in mortality rates and life expectancy. Since we want to apply our vector autoregressive model on stationary series, we compute the first difference of the log-mortality rate $\Delta y_{i,t} = y_{i,t} - y_{i,t-1}$ or, in other words, the log-mortality improvement rates. By working on these quantities, we remove a linear trend in the $y_{i,t}$ series.

With this notation, we specify the mortality rate improvement process by a stationary vector autoregressive model of temporal lag p or a VAR(p). For a minimum age i_{\min} and a maximum age i_{\max} , we define the d -dimensional vector of log-mortality rate improvements, with $d = i_{\max} - i_{\min} + 1$, as $\Delta \mathbf{Y}_t = (\Delta y_{i_{\min},t}, \Delta y_{i_{\min}+1,t}, \dots, \Delta y_{i_{\max},t})^\top$. Next, we assume the following dependence structure dynamic,

$$\Delta \mathbf{Y}_t = \mathbf{C} + \sum_{k=1}^p \mathbf{A}_k \Delta \mathbf{Y}_{t-k} + \mathbf{E}_t, \quad (2.1)$$

where, for $k = 1, \dots, p$, \mathbf{A}_k are $d \times d$ -autoregressive matrices, \mathbf{C} is a d -dimensional vector of constants (an intercept), and \mathbf{E}_t is a d -dimensional Gaussian white noise with mean 0 and Σ the related covariance matrix. We denote by $\epsilon_{i,t}$ its marginals. The matrices \mathbf{A}_k , $k \in \{1, \dots, p\}$, capture the relationship between current mortality improvements and the k^{th} lag of $\Delta \mathbf{Y}_t$. In other words, this corresponds to the Granger causality between different cohorts for the mortality improvement rates.

The VAR(p) model allows taking into account a more complex dependence structure than the usual mortality factor models. First, our model enables a larger flexibility in the long-term spatio-temporal dependence structure through the autoregressive matrices than the standard factor models. For a given square (i, t) in the Lexis diagram, we let the possibility for the improvement mortality rates $\Delta y_{i,t}$ to be dependent of all the ages among the d -dimensional space of ages, and through all the p temporal lags. In particular, we notice that this domain includes a cohort effect for these improvement rates. For each factor $\Delta y_{i-1,t-1}, \dots, \Delta y_{i-p,t-p}$, this effect is indeed captured by the loading coefficients positioned on the k^{th} -subdiagonal of the matrix \mathbf{A}_k for each $k \in \{1, \dots, p\}$. Hence, the VAR(p) structure permits for the shocks to propagate through different periods. Compared to the model proposed by Li and Lu (2017), the lag order p can take a value greater than 1, allowing to capture a more complex dependence structure.

¹More precisely, it is their alternatives using mortality improvements, as documented by Haberman and Renshaw (2012).

Furthermore, it has the ability to enlighten some effects that are not captured in the standard mortality literature, e.g. between neighboring cohorts, as we do not impose any constraint on the matrices \mathbf{A}_k , $k \in \{1, \dots, p\}$. Compared to most of factor models, the second improvement of our model on the dependence flexibility is that it captures the long-term co-movement by the autoregressive matrices and the short-term dependence through the covariance matrix at the same time.

Nevertheless, the major issue of our VAR(p) model is that it is a natural high-dimensional problem. The number of parameters for the Granger causality matrices is pd^2 , without considering the covariance matrix and the constant vector. In mortality modeling studies, it is common to focus on the age range from 0 to 100, that is to say $d = 101$, while the historical data for estimation rarely exceed 70 years. Given this, a VAR(3) implies 30,704 parameters estimated on only 7,070 observations, which makes the ordinary least-squares estimation not feasible. To avoid over-fitting, additional constraints have to be added. In this direction, Li and Lu (2017) impose that some parameters have to be nil for guarantying that the model is sparse and stationary. Conversely, we choose a less arbitrary high-dimensional selection variables technique, developed in the next section, to ensure sparsity.

Similarly, the covariance matrix estimation is also a high-dimensional problem with $\frac{d(d+1)}{2}$ parameters. In order to estimate prediction intervals, we consider an additional specification for the residuals. Although some high-dimensional techniques do exist for covariance estimation (see e.g. Schäfer and Strimmer, 2005; Opgen-Rhein and Strimmer, 2007; Bickel and Levina, 2008; Bien and Tibshirani, 2011), we rather choose a simple parametric form presented in the following part to reduce the number of parameters.

3 High-dimensional estimation of the VAR model

As highlighted in the previous section, the VAR(p) model estimation is a high-dimensional problem, especially with mortality data. The estimation can be decomposed into two parts: first, we estimate the pd^2 -dimensional autoregressive matrices, then the d^2 -dimensional covariance matrix. The dimension reduction in the autoregressive matrices is treated through an elastic penalization in Section 3.1. We tackle the problem of the covariance through the choice of a parametric form in Section 3.2.

3.1 Elastic-net

We now described the extension of the elastic-net regularization and variable selection method, proposed by Zou and Hastie (2005), for the high-dimensional estimation of our autoregressive matrices. This technique can be seen as the combination of the LASSO \mathcal{L}_1 -penalty, introduced by Tibshirani (1996), and the ridge \mathcal{L}_2 -penalty developed by Hoerl and Kennard (1970). Elastic-net has similar properties of variable selection as the LASSO. Moreover, it provides a grouping effect: highly correlated variables tend to be selected or dropped together. LASSO and elastic-net have already been extended to VAR model (Gefang, 2014; Basu et al., 2015), mostly with an economic application (see e.g. Song and Bickel, 2011; Furman, 2014).

Therefore, we estimate the VAR(p) model presented in Equation (2.1) with T observations of

the process $\Delta\mathbf{Y}_t$ for $t = t_{\min}, \dots, t_{\max}$ by minimizing the criterion

$$L(\mathbf{C}, \mathbf{A}_1, \dots, \mathbf{A}_p) = \frac{1}{T-p} \sum_{t=t_{\min}+p}^{t_{\max}} \|\Delta\mathbf{Y}_t - \mathbf{C} - \sum_{k=1}^p \mathbf{A}_k \Delta\mathbf{Y}_{t-k}\|_2^2 - \alpha\lambda \sum_{k=1}^p \|\mathbf{A}_k\|_1 - \frac{(1-\alpha)\lambda}{2} \sum_{k=1}^p \|\mathbf{A}_k\|_2^2, \quad (3.1)$$

where we define for a d -dimensional vector $\mathbf{b} = (b_i)_{1 \leq i \leq d}$

$$\|\mathbf{b}\|_2^2 = \sum_{i=1}^d |b_i|^2,$$

and for a $d \times d$ -dimensional matrix $\mathbf{B} = (b_{i,j})_{1 \leq i \leq d, 1 \leq j \leq d}$

$$\|\mathbf{B}\|_1 = \sum_{i=1}^d \sum_{j=1}^d |b_{i,j}| \text{ and } \|\mathbf{B}\|_2^2 = \sum_{i=1}^d \sum_{j=1}^d |b_{i,j}|^2.$$

The parameter $\alpha \in [0, 1]$ is a hyper-parameter which determines the mix between ridge and LASSO penalties. We use a 10-folds cross-validation method to choose the penalty coefficient λ . It determines the strength of the penalties, for example in the LASSO case, the higher λ gets, the fewer number of variables are selected. The algorithm we used is described in Friedman et al. (2010). In theory, the LASSO \mathcal{L}_1 -penalty part forces most of the coefficients to 0. Nevertheless, in a more practical approach, the algorithm employed does not lead to exact zeroes. Thus, the R-package `glmnet` (Friedman et al., 2010) applies a threshold on the coefficients. Furthermore, following Chatterjee and Lahiri (2011), the `sparsevar` R-package (Vazzoler et al., 2016) enables to apply a more tailor-made threshold for time-series estimation which equals to $\frac{1}{\sqrt{pd \ln T}}$, that we retain for our model.

The hyper-parameter α is determined through a grid search. For every value α_h of a pre-defined grid $\{\alpha_1, \dots, \alpha_H\}$, we estimate the parameters of the VAR model $\{\hat{\mathbf{C}}, \hat{\mathbf{A}}_1, \dots, \hat{\mathbf{A}}_p, \hat{\lambda}\}_{\alpha_h}$, as explained just before, and deduce the residuals $\hat{E}_{\alpha_h, t}$ for $t \in \{t_{\min} + p, \dots, t_{\max}\}$.

In the applications, we estimate the tuning parameters by minimizing the prediction error, that we obtain by computing the root-mean-square error

$$\text{RMSE}(\alpha_h) = \sqrt{\frac{1}{d(T-p)} \sum_{t=t_{\min}+p}^{t_{\max}} \|\hat{E}_{\alpha_h, t}\|_2^2}. \quad (3.2)$$

In our application, we considerate the grid $\{0.5, 0.6, 0.7, 0.8, 0.9, 1\}$ in order to impose a larger weight to the LASSO penalty for sparsity purposes.

The choice of the lag order p for our VAR elastic-net (VAR-ENET) model differs significantly from the usual lag order selection in the standard VAR models. The parameter p does not fully determine the number of parameters, since the LASSO penalty force the less significant coefficients to zero. By increasing the lag order, some non-null coefficients can be forced to zero in favor of other coefficients in autoregressive matrices of higher lag order. Moreover, if there is no significant coefficient above a certain lag order, all autoregressive matrices above the limit order are largely forced to zero. Thus, we chose a relatively large p to capture eventual high order lag effects, without being worried of over-fitting.

3.2 Variance-covariance estimation

The autoregressive matrices are not the only high-dimensional problem of the VAR(p) model, the variance-covariance matrix estimation has $\frac{d(d+1)}{2}$ parameters. This number can quickly get higher than the number of observations while dealing with mortality modeling, and then can cause overfitting, as noted e.g. by Li and Lu (2017).

To overcome this issue, we propose an approach to estimate the covariance matrix with a parametric covariance function, in a manner similar to Spodarev et al. (2013). Firstly, for each age i , we estimate the standard empirical variance $\hat{\sigma}_i^2$ of the residual, and for each couple of ages $(i, j) \in \{i_{\min}, \dots, i_{\max}\}^2$, we estimate the empirical correlation

$$\hat{r}_{i,j} = \frac{\hat{\sigma}_{i,j}}{\hat{\sigma}_i \hat{\sigma}_j},$$

where $\hat{\sigma}_{i,j}$ is the empirical covariance. Then, guided by the form of the empirical correlation matrices and by the approach of Christiansen et al. (2015), we use a parametric form close to the stable family of covariance functions

$$r_{i,j} = \beta e^{-(a_i + a_j) \times |i-j|} \times \mathbb{1}_{\{i \neq j\}} + \mathbb{1}_{\{i=j\}}, \quad (3.3)$$

with $\beta \geq 0$ and $a_i \geq 0$ for each age i . We fit the model based on the empirical correlation as the OLS solution. Thus, after determining $\hat{\beta}$ and $(\hat{a}_{i_{\min}}, \dots, \hat{a}_{i_{\max}})$, we compute our parametric correlation $\tilde{r}_{i,j}$ given by Equation (3.3). Finally, for each couple of ages (i, j) we estimate the covariance by

$$\tilde{\sigma}_{i,j} = \tilde{r}_{i,j} \times \hat{\sigma}_i \hat{\sigma}_j.$$

4 Empirical analysis

In this section, we apply our high-dimensional VAR-ENET model to real data and show its strengths in estimating and forecasting populations. Different populations are considered and we analyze both our in-sample and out-sample results compared to those obtained with standard factor models. In the following, the computations are carried out with the R software (R Core Team, 2017). Our scripts are available upon request.

4.1 Data

The datasets that we analyze comes from the Human Mortality Database (2016). We choose to illustrate our approach with historical mortality data from the England and Wales (UK), the United States (US) and France (FR), as these populations have been largely studied. At first, the overall population is considered, and then both males and females are segregated. We select the age-period observation $\{45, \dots, 99\} \times \{1950, \dots, 2012\}$ which was available for these 3 countries when the data was extracted. We begin our analysis by a visual inspection of our data. Figure 1 describes the shape of the period log-mortality improvements for populations on a Lexis diagram where the trajectory of one cohort follows a 45 degree line. Different cohort effects can be observed for these countries with pink (resp. green) shades for positive (resp. negative) improvements, indicating a lower (resp. higher) survival.

On the left side, French improvement rates do exhibit some significant diagonal patterns corresponding to the so-called cohort effects. Two diagonals stand out, more precisely for individuals

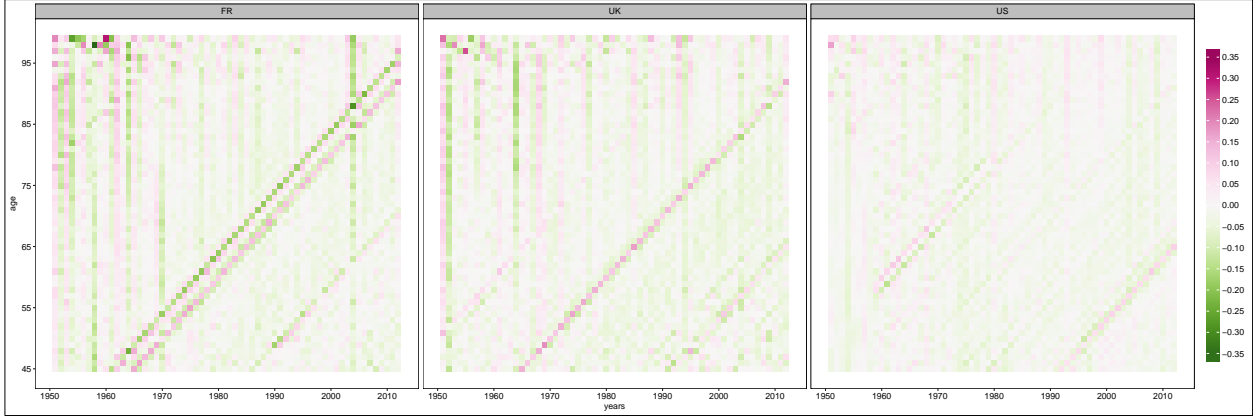


Figure 1: The period log-mortality improvements for England and Wales (UK), the United States (US) and the France (FR) on the age-period observation $\{45, \dots, 99\} \times \{1950, \dots, 2012\}$ both for males and females.

aged 45 in 1960: one of them has positive improvements and the other negative ones. We notice another improvement series for people aged 45 during the mid-80's. We note that these effects are strongly reduced with the correction developed by Boumezoued (2016b), thanks to fertility rates. Several vertical patterns corresponding to period effects are also observed, especially for the older ages. Vertical structures are much more evident in the English data. Moreover, cohort effects are visible for this country. We note a highly positive tendency for individuals born in 1920 in particular. Diagonal and vertical structures associated with cohort and period effects clearly stand out in the American data. Contrary to French and English mortality, the intensity of increasing at higher ages is less marked, and the observed effects seem to be more homogeneous. Female and male data are displayed in Appendix 1.

We have chosen to apply the VAR-ENET to the first difference of log-mortality rates series because they are known in the literature as stationary time series. In order to verify this point, we perform a Phillips-Perron test (Perron, 1988) and an augmented Dickey-Fuller test (Said and Dickey, 1984) on every age mortality series for each of the nine populations of interest. All of these time series satisfy the Phillips-Perron test at a confidence level of 1%, and 93% of them are considered as stationary by the augmented Dickey-Fuller test at a level of 5%. These results strengthen our choice to focus on the first difference of time-series.

4.2 In-sample analysis

In this section, we present the results of our empirical estimations with the VAR-ENET model for each population. The goodness of fit is analyzed by comparing the in-sample results with:

- the usual Lee-Carter model (Lee and Carter, 1992), estimated with the approach of Brouhns et al. (2002) and given by

$$y_{i,t} = \alpha_i + \beta_i \kappa_t, \quad (4.1)$$

with the hyper-parameters α_i and β_i , and the mortality trend κ_t ;

- the M7 model developed by Cairns et al. (2009), which considers a quadratic and a cohort

effect, i.e.

$$y_{i,t} = \kappa_t^{(1)} + (i - \bar{i}) \kappa_t^{(2)} + \kappa_t^{(3)} \left((i - \bar{i})^2 - \hat{\sigma}_i^2 \right) + \gamma_{t-i}, \quad (4.2)$$

where $\kappa_t^{(j)}$, $j = 1, 2, 3$, are period effects, γ_{t-i} is a cohort effect, \bar{i} is the average age in the data, and $\hat{\sigma}_i^2$ is the average value of $(i - \bar{i})^2$.

4.2.1 Parameters estimation

Let us present our estimated results on the period 1950 – 2012. The parameters are estimated as described in Section 3. For the lag order p , we choose the value 7, which represents between 10% and 15% of the observation, depending if we analyze the in-sample or, as in the latter sections, out-sample. Table 1 reports the list of the estimated hyper-parameters for each population of interest. We note that, for some populations, we retain the value 1 for α , i.e. we estimate the model with the LASSO constraint only.

Table 1: The estimated VAR-ENET hyper-parameters.

| Country | Population | α | λ |
|---------|------------|----------|-----------|
| FR | Female | 0.6 | 0.0003 |
| FR | Male | 0.9 | 0.0002 |
| FR | Total | 0.7 | 0.0003 |
| US | Female | 1.0 | 0.0004 |
| US | Male | 0.6 | 0.001 |
| US | Total | 1.0 | 0.0003 |
| UK | Female | 0.8 | 0.0004 |
| UK | Male | 1.0 | 0.0001 |
| UK | Total | 0.5 | 0.0003 |

Note: This table displays the estimated hyper-parameters α and λ in Equation (3.1) for the VAR(7) models. We consider males, females and the overall populations for FR, UK and US.

The first Granger causality matrix \mathbf{A}_1 for each population is displayed in Figure 2. These estimated matrices are sparse and capture two main structures. First, an expected cohort effect, induced by individuals belonging to the same generation, is highlighted by allocated coefficients on the k^{th} subdiagonal for \mathbf{A}_k for $k \in \{1, \dots, p\}$ (not shown for $k \in \{2, \dots, 7\}$). Indeed, those coefficients in the VAR model describe the Granger causality of $\Delta y_{i-k,t-k}$ on $\Delta y_{i,t}$. This effect appears positively mainly for the younger ages of our different samples. It is more diffuse between the ages of 65 and 85 years old, and seems to become negative after. On the presented figure, the difference between countries appears clearly and the so-called cohort effect is relatively strong for the US population, compared to the FR and the UK. The cohort effect is still clearly visible for $k \in \{2, \dots, 7\}$.

Then, we notice some age-specific effects corresponding to vertical and, to a lesser extent, horizontal patterns, especially in the UK data. These second types of patterns reveal non-trivial

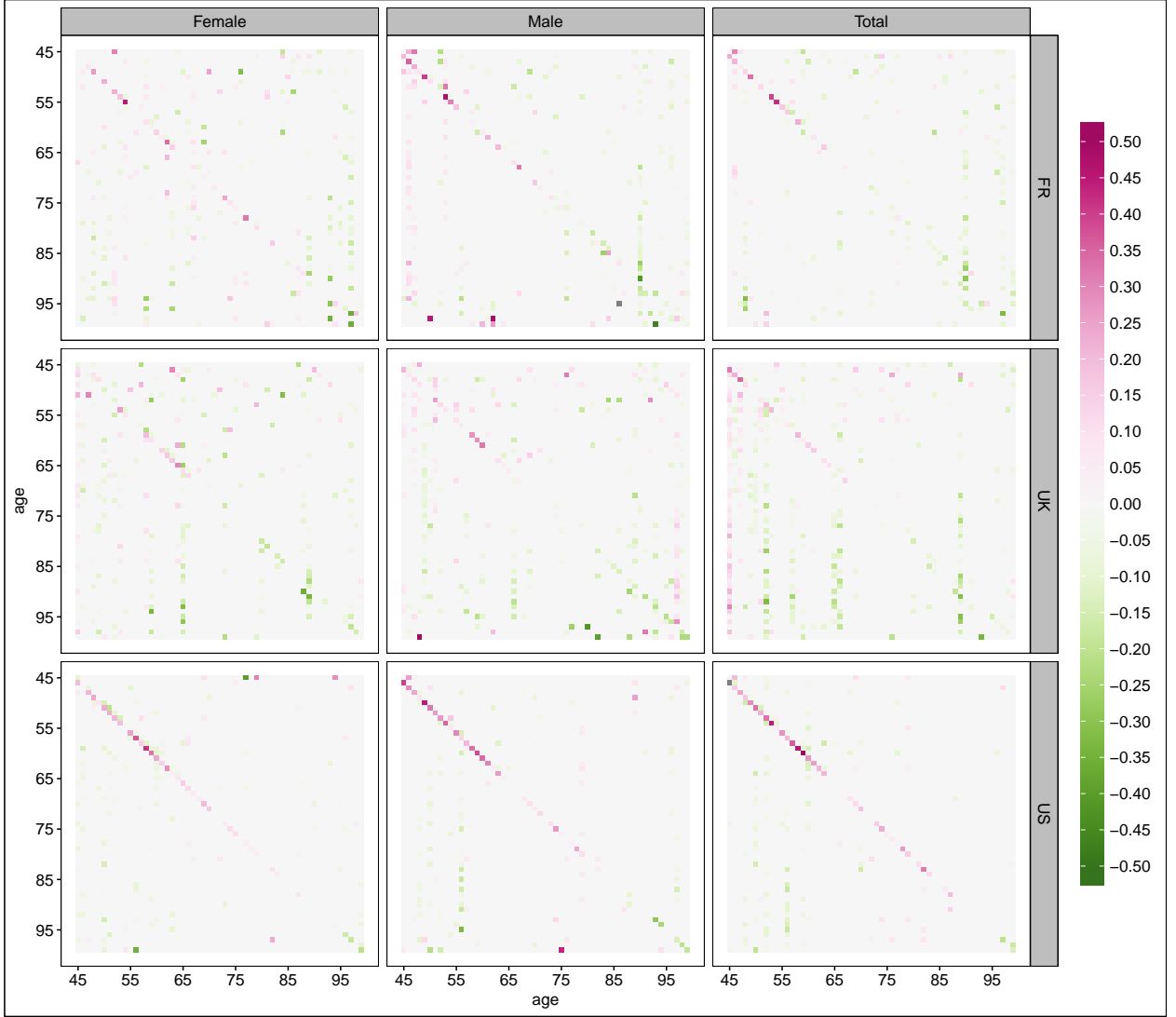


Figure 2: The Granger causality matrix \mathbf{A}_1 for England and Wales (UK), the United States (US) and the France (FR) on the age range 45-99 for females, males and the overall population.

interactions between different cohorts which are non-necessary within a close neighboring or localized under the diagonal of the Granger causality matrices as it is proposed by Li and Lu (2017) or Doukhan et al. (2017). This latter effect, underlined by our data-driven approach, has not been well documented in the literature yet to our knowledge. More concretely, a vertical pattern on the i^{th} column of \mathbf{A}_1 reveals a persistent effect from the term $\Delta y_{i,t-1}$ on $\Delta y_{j,t}$, $j \in \{45, \dots, 99\}$, meaning an effect on the older cohorts for $j < i$ and on the younger cohorts for $j > i$. Similar structures can be observed on the matrices \mathbf{A}_k for $k \in \{2, \dots, 7\}$. On these datasets, these effects are quite difficult to interpret, but one can note that the transfers from the older to the younger cohorts are scattered, whereas some vertical lines are generally captured for columns corresponding to the ages between 45 and 65. Disaggregated data would be very useful to explore these effects further. Indeed, these patterns could result from biological, environmental or societal causes, unless it is due to some anomalies in the HMD (Cairns et al., 2016b; Boumezoued, 2016b).

4.2.2 In-sample model comparison

We now study how our approach fits well and captures better the mortality pattern on these subsets compared to the classical Lee-Carter (LC) model (4.1) and the M7 model (4.2). These last models are estimated using the R-package `StMoMo` (Villegas et al., 2017) following their usual two-stage fitting procedure: first, we estimate the factor coefficients of each model, and then we forecast it using univariate ARIMA processes, automatically selected by the R-package using an AIC criterion. To be comparable with our one-stage fitting approach, the results for these models are those obtained after fitting the time-series parameters.

Table 2 contains the values of the RMSE, as defined in Equation (3.2) for each model and each population. The LC and M7 models have quite comparable results, whereas the VAR-ENET model has the lowest RMSE value in each country. These results underline the high degree of versatility of our approach.

Table 2: The RSME of the LC, the M7 and the VAR models.

| Country | Model | RMSE Female | RMSE Male | RMSE Overall |
|---------|-------|-------------|-----------|--------------|
| FR | LC | 0.056 | 0.053 | 0.045 |
| FR | M7 | 0.093 | 0.078 | 0.078 |
| FR | VAR | 0.011 | 0.014 | 0.013 |
| UK | LC | 0.052 | 0.056 | 0.048 |
| UK | M7 | 0.059 | 0.050 | 0.044 |
| UK | VAR | 0.015 | 0.010 | 0.009 |
| US | LC | 0.045 | 0.045 | 0.042 |
| US | M7 | 0.052 | 0.053 | 0.048 |
| US | VAR | 0.016 | 0.015 | 0.012 |

Note: This table reports the RSME values obtained after fitting the LC, the M7 and the VAR-ENET models. We compare this indicator for males, females and the overall populations for FR, UK and US.

The RMSE value is also computed for each age and each year, and the results are displayed respectively on Figures 3 and 4 for all of the overall populations. First, we note that, on top of having the lowest RMSE, the VAR-ENET model leads to a lower and a more stable error over the age. For example, while the RMSE of the LC and M7 tends to increase in higher ages on the FR data, this error measure stays relatively at the same level for the VAR-ENET model. For the UK data, the LC model has a relatively high error compared to the M7. Although this last model is known to offer a good fit for this population, the VAR-ENET model offers an significant gain in terms of goodness of fit.

The RMSE patterns across periods are more erratic. The LC and M7 have effectively the same level. Some peaks are observed, especially for the FR data where high correlation for the RMSE series is visible. In contrast, for the VAR-ENET model, the errors remain rather low and the peaks are diminished in absolute value. The results for females and males are given in Appendix 2.

Some of the peaks may be explained by specific events that have an unexpected impact on the mortality rates, such as an influenza epidemic or a heat wave (Huynen et al., 2001). Indeed, this type of exogenous stresses is difficult to predict with only mortality rate series, which explains why the peaks are observable for the three models. For example, we try to explain the relatively high RMSE

in France in 2004. In 2003 a heat wave led to one of the hottest summer ever recorded in France and, as a direct consequence, to higher mortality rates during this year, especially for the elderly. Then, the mortality was much lower in 2004 due to the so-called harvesting effect (Toulemon and Barbieri, 2008; Izraelewicz, 2012). On the contrary, in the calculation of the RMSE, the mortality rates of the year 2004 are forecasted from the observation of the 2003 rates in accordance with the temporal dynamics we have imposed; in this way, the 2004 mortality was expected to be relatively high. This must explain why we observe a RMSE peak in 2004 for the French population.

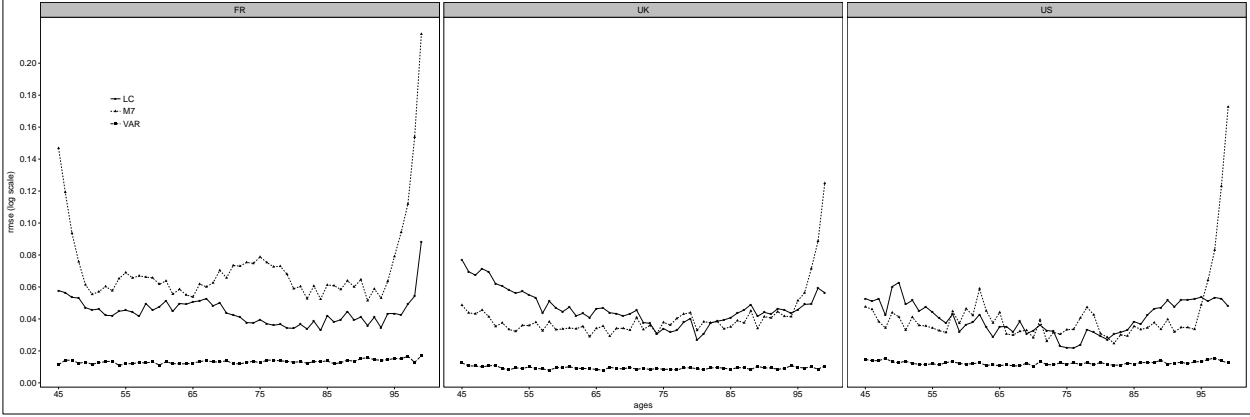


Figure 3: The RMSE for England and Wales (UK), the United States (US) and the France (FR) grouped by age for the overall populations.

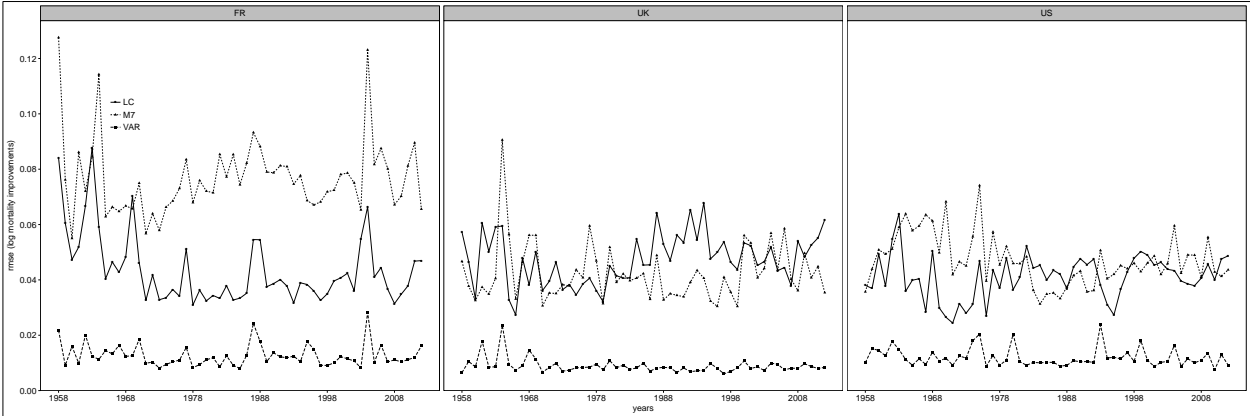


Figure 4: The RMSE for England and Wales (UK), the United States (US) and the France (FR) grouped by period for the overall populations.

4.3 Out-of-sample performance

For risk management in insurance or more generally for demographers or public policy purposes, mortality rates require being predicted based on the past information. A quite usual test for accuracy is to analyze how the model is able to reproduce the mortality rates correctly. Note that this objective is more demanding than measuring the prediction power on the residual life expectancy. A reasonable model should be able to predict a kind of convergence for mortality at a

similar level.

We focus on the prediction power of the VAR-ENET model compared to the LC and the M7 through an analysis of the out-sample forecasting performance on the same age-period space. To this end, we first estimate each model based on the observations from 1950 to 2000, then we forecast the mortality rates for the period 2001-2012. Note of course that the mortality rates can be easily calculated using the VAR model, based on the predicted improvement rates and the initial values known for the last year of the training sample. We choose a similar measure to the one taken for the in-sample analysis, the root mean squared forecast error (RMSFE) that we define for a projection horizon h as

$$\text{RMSFE} = \sqrt{\frac{1}{dh} \sum_{i=i_{\min}}^{i_{\max}} \sum_{t=t_0+1}^{t_0+h} (y_{i,t} - \hat{y}_{i,t})^2}, \quad (4.3)$$

where t_0 is the year 2000 in our study.

We compare the predictive power of the different models on the period 2001 – 2012 for the 3 populations (overall, female, male) of the 3 countries of interest. The results are displayed in Table 3 (we also display the results for different estimation years but with the same forecasting period in Appendix 3). We note that the RMSFE is smaller for every population with the VAR-ENET than with the LC or M7 models, indicating that the former one has a higher predictive power. Furthermore, we notice that, in the VAR-ENET applications, all the forecasting errors are of the same order of magnitude, no matter the selected population. On the contrary, the M7 and, to a lesser extent, the Lee-Carter model tend to have a more variable forecasting error, depending on the considered population. These results highlight the stability of the prediction error of the VAR-ENET over different populations compared to the other two benchmark factor models due to the more data-driven approach of the first one, allowing it to better capture the specificities of each populations’ mortality dynamic.

Table 3: The RSMFE of the LC, the M7 and the VAR models estimated.

| Country | Model | RMSFE Female | RMSFE Male | RMSFE Overall |
|---------|-------|--------------|------------|---------------|
| FR | LC | 0.103 | 0.098 | 0.066 |
| FR | M7 | 0.500 | 0.159 | 0.201 |
| FR | VAR | 0.076 | 0.079 | 0.063 |
| UK | LC | 0.133 | 0.122 | 0.128 |
| UK | M7 | 0.171 | 0.098 | 0.126 |
| UK | VAR | 0.091 | 0.084 | 0.076 |
| US | LC | 0.085 | 0.104 | 0.087 |
| US | M7 | 0.172 | 0.109 | 0.103 |
| US | VAR | 0.066 | 0.088 | 0.078 |

Note: This table reports the out-of-sample performance *via* the RSMFE values for the LC, the M7 and the VAR-ENET models estimated on the period 1950 – 2000. We compare this indicator for males, females and the overall populations for FR, UK and US.

We plot the RMSFE in Figure 5 for the overall population of our three countries of interest. The results for female and male populations are postponed in Appendix 3. First, we see that for all projection horizons and all considered populations, our VAR-ENET model implies a lower level of

RMSFE, that is to say, it provides more accurate forecasts than the two benchmarks. We note that on the French population, our model and the LC one have a very similar profile of prediction power over the different periods. On the contrary, the M7 model tends to produce much larger predictive errors when applied to the French data, suggesting that this latter model does not capture well the specificities of this country's mortality dynamics. While the M7 seems to diverge on the French population, we note that it has a higher predictive power on the US and UK populations.

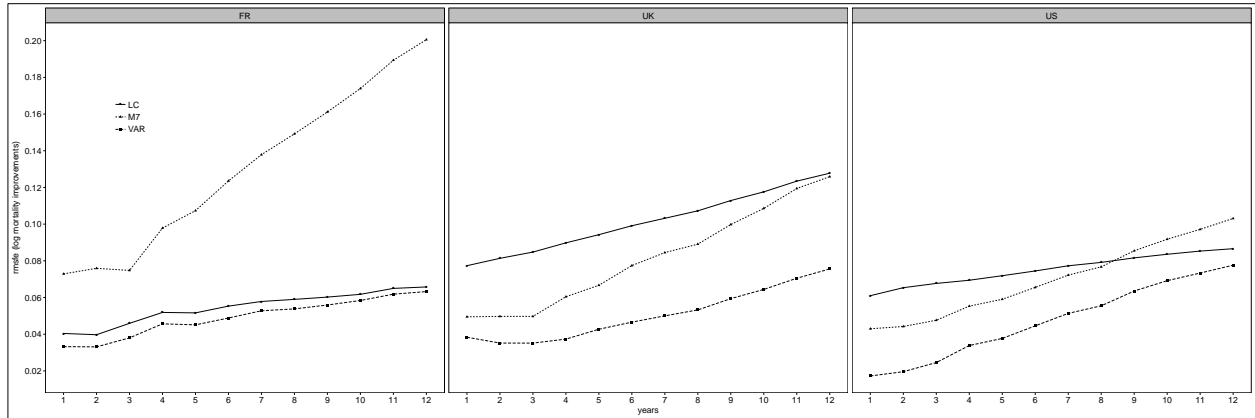


Figure 5: The RMSFE for England and Wales (UK), the United States (US) and France (FR) grouped by period for the overall populations.

We now focus on the forecasting error by age groups. We choose to separate the age dimension into 5 classes and compute the RMSFE at a projection horizon of 10 years. By doing so, we compare the predictive power over the different ages. Indeed, depending on the purpose of the mortality forecasting application, one could be more interested in producing accurate predictions for some specific ages. We show the results of the three models in Figure 6. We see that, in general, our VAR-ENET model has a smaller forecasting error for all age groups than the M7 model. Considering the Lee-Carter, it outperforms our model for higher ages, nevertheless we note that the RMSFE is more volatile among the age groups, especially on the UK population.

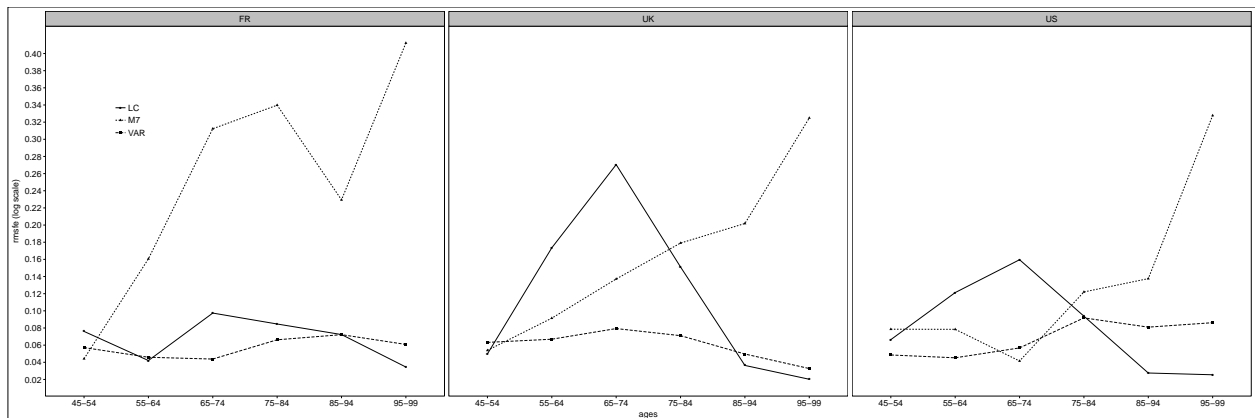


Figure 6: The RMSFE for England and Wales (UK), the United States (US) and France (FR) grouped by age for the overall populations.

The results presented in this section suggest that our model slightly outperforms the two bench-

marks for these populations. In addition, it seems to be more stable. However, the improvement between the VAR-ENET model and the LC model is not as important as in our in-sample study. This finding is actually not so surprising. Indeed, mortality forecasting based on aggregated data such as HMD is complicated since the underlying mechanisms regarding the population dynamics at micro level are unavailable (Bensusan et al., 2010-2017; Boumezoued, 2016a). For instance, several latent and heterogeneous factors may cause the observable trend of mortality rates, and their contributions to this overall trend may vary over time. Thus, performances of all forecasting models based on aggregated data seem to be intrinsically limited. In particular, we note that our data-driven model is subject to the same issue. Furthermore, although the VAR-ENET model tends to better capture the spatio-temporal mortality dependence structure given the past aggregated data without other information, if some external hints about the future trend are available, a standard factor model may provide more accurate forecasts.

4.4 Forecasting application

Figure 7 displays the median forecasts of the log of death rates for ages in $\{45, 65, 85, 95\}$ using the VAR-ENET model from 2013 to 2062. We note that the trends seem rather realistic, even though we remark that the forecasted mortality rate at age 65 series for the UK male and female cross around the year 2040. More generally, the male and female mortality rates tend to converge rapidly for the UK population. This forecasting result has already been observed in the literature with other models with a similar estimation period (see e.g. Bohk-Ewald and Rau (2017)). We note on the forecasted series that there are some limited shocks during the first projection years, followed by a linear trend. This effect is characteristic of the VAR model and shows how it can propagate innovation shocks among a cohort for example.

Figure 8 compares the median forecasts of the log of death rates of the LC model, the M7 model and the VAR-ENET model from 2013 to 2062. First we note that for many of the forecasted series, the LC and the VAR-ENET produce similar trends, while the M7 produce rather unrealistic forecasts for most of the considered ages, especially for the French female population. On the considered populations, the VAR-ENET model tends to produce higher mortality rates at younger ages compared to the Lee-Carter, whereas at older ages we remark the opposite effect. For example, one can observe that the projected rates of mortality for British men obtained with the Lee-Carter model decrease faster than those generated with the VAR-ENET model at age 45. Their trend is rather close to those of the M7 model, but it is less credible than the forecasts of the VAR-ENET model given the past trends. At age 95, we remark that the mortality rates projected by the Lee-Carter model decrease slowly. Our forecasts for this age have a steeper slope, which is closer to the trend at age 85. An explanation for that, as noted by Li and Lu (2017), is the co-integration effect offered by our approach, which permitted to obtain a more consistent behavior for neighbored age in the long run.

5 A multi-population extension

Some of the standard mortality forecasting models can be extended to multi-population. However, many of these extensions suffer from limits. One of the recurrent limits in multi-population mortality modeling is the restriction of the extension to only 2 populations. For example, we can note the GRAVITY model of Dowd et al. (2011) or the Bayesian model of Cairns et al. (2011). Another

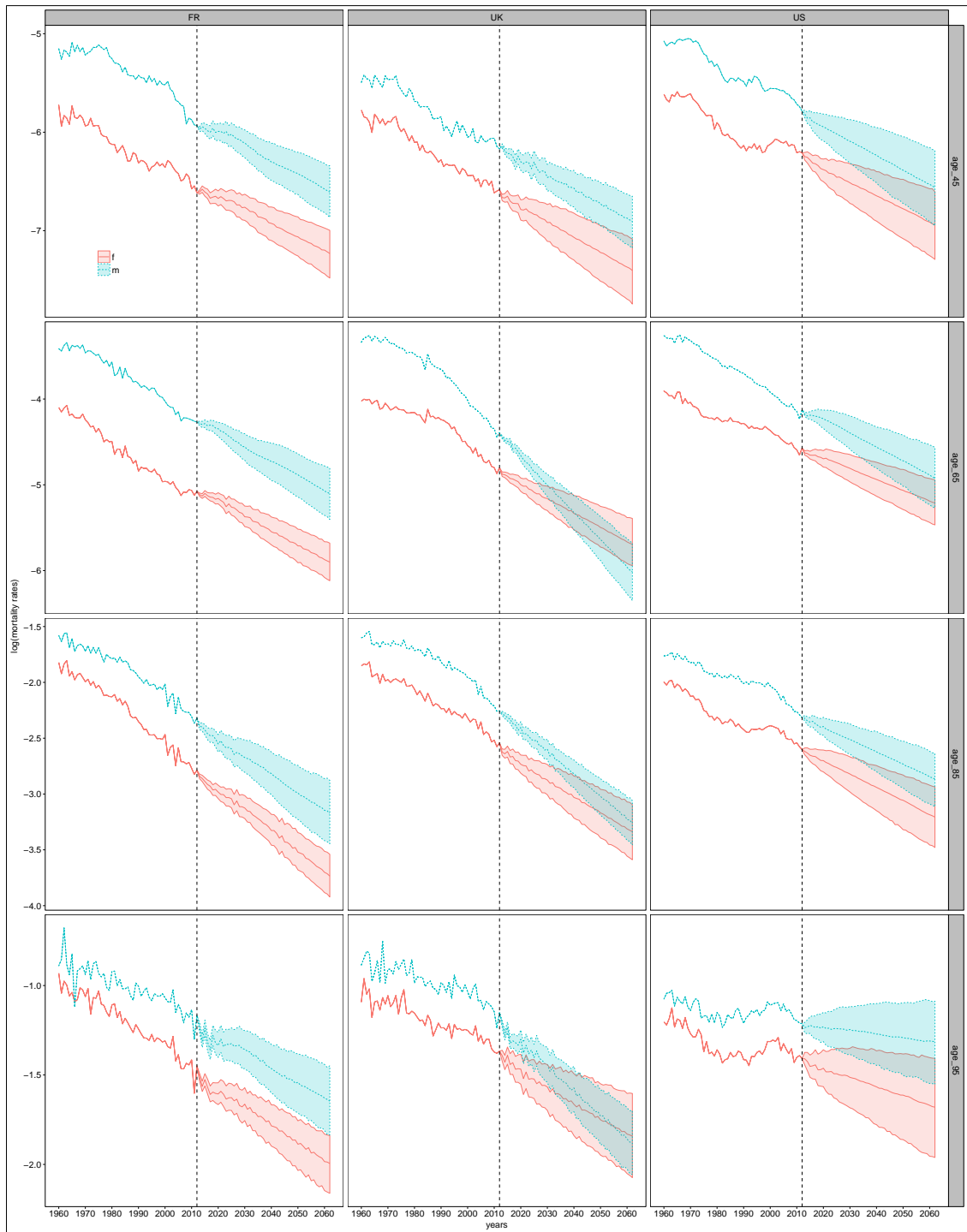


Figure 7: The observed and the projected log of death rates for British (UK), American (US) and French(FR) females and males with the 97.5% prediction intervals, obtained from the VAR-ENET model.

restriction imposed by some existing multi-population models is the necessity to determine a dominant population and sub-populations, see e.g. the SAINT model of Jarner and Kryger (2011), or a common trend for the different populations like in Li and Lee (2005).

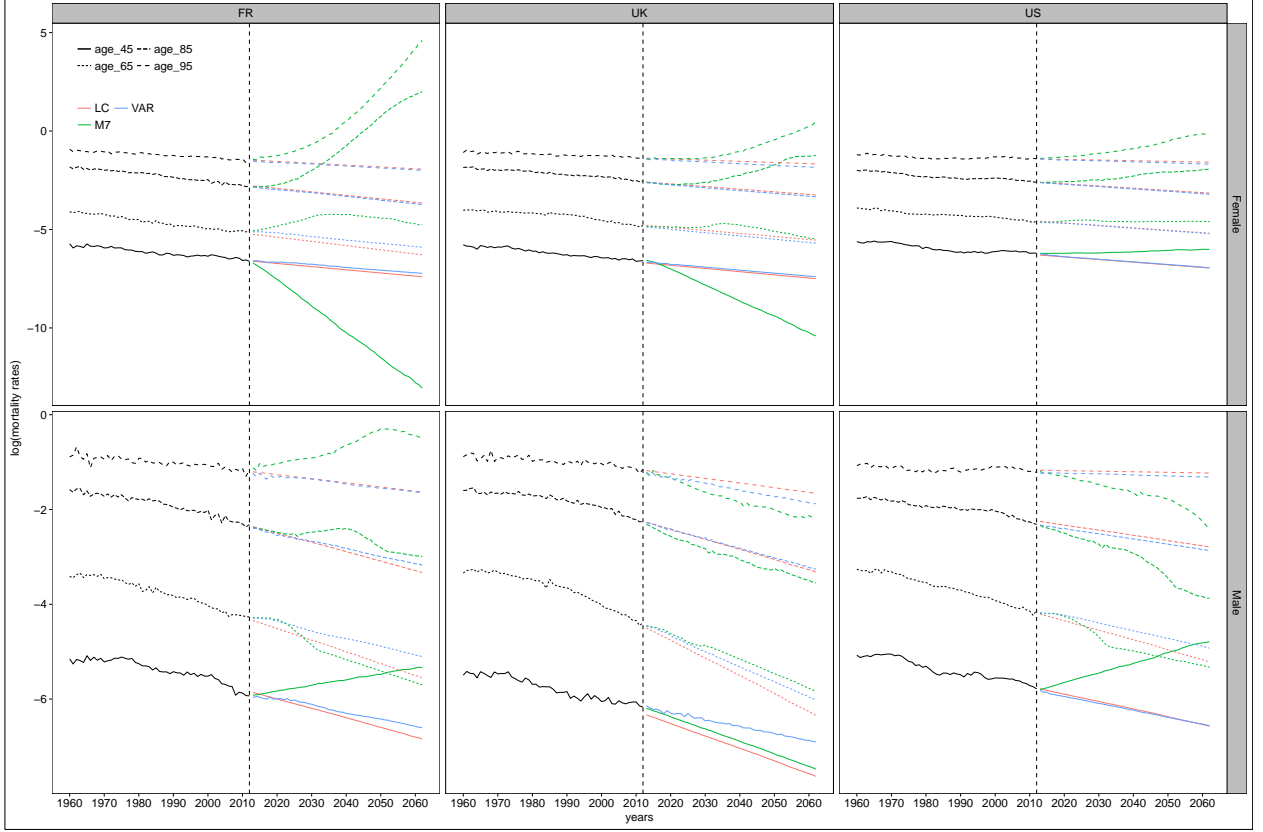


Figure 8: The observed and the projected log of death rates for British (UK), American (US) and French (FR) females and males. This figure compares trends obtained with the LC, the M7 and the VAR-ENET models.

In this section, we propose an extension of our model for multi-population mortality forecasting. We denote M the number of selected populations and $y_{m,i,t}$ the log of mortality rates for the m^{th} population. We suppose that the pair (i_{\min}, i_{\max}) is the same for all the populations to avoid exaggerated notations, although we could have chosen M different pairs of age limits. Thus, we define M different d -dimensional vectors $\Delta \mathbf{Y}_{m,t}$ that we concatenate into a single Md -dimensional vector $\Delta \mathbf{Y}_t = (\Delta y_{1,i_{\min},t}, \dots, \Delta y_{1,i_{\max},t}, \Delta y_{2,i_{\min},t}, \dots, \Delta y_{M,i_{\max},t})^\top$. We then apply the same model as in (2.1) except that the dimension equals now to Md .

The $(Md) \times (Md)$ -dimensional autoregressive matrices and the Md -dimensional vector of constants are estimated through the same elastic-net methodology as in the single population problem. However, the covariance matrix estimation needs to be extended since its structure may change significantly compared to the single population case.

In the multi-population context, the covariance matrix Σ is $(Md) \times (Md)$ -dimensional. Firstly, we propose to consider this matrix as a block matrix broken into M^2 different $d \times d$ -dimensional submatrices noted $\Sigma_{m,n}$ for each population couple $(m, n) \in \{1, \dots, M\}^2$. Then, for each population m , we estimate the diagonal submatrix $\Sigma_{m,m}$ through the same methodology as in the single population, obtaining in this way $\tilde{\Sigma}_{m,m}$. Since $\tilde{\Sigma}_{m,m}$ is a positive-definite matrix, we define its Cholesky decomposition

$$\tilde{\Sigma}_{m,m} = \tilde{R}_m^\top \tilde{R}_m,$$

where \tilde{R}_m is an upper triangular matrix. Finally, for a couple of populations (m, n) we estimate the covariance submatrix as

$$\tilde{\Sigma}_{m,n} = \rho_{m,n} \times \tilde{R}_m^\top \tilde{R}_n,$$

where $\rho_{m,n} \in [-1, 1]$ is the empirical Pearson's correlation coefficient between the observations of residuals $(\epsilon_{m,i,t})_{(i,t)}$ and $(\epsilon_{n,i,t})_{(i,t)}$ for $(i, t) \in \{i_{\min}, \dots, i_{\max}\} \times \{t_{\min} + p, \dots, t_{\max}\}$.

Thus, our extended covariance model has only $M(2d + 1) + \frac{M(M-1)}{2}$ parameters. The number of ages, d , being generally much larger than the number of populations in mortality modeling, we note that $M(2d + 1) \gg \frac{M(M-1)}{2}$, meaning that we do not add many parameters for covariance estimation while modeling M populations together compared to fitting M single models.

6 Conclusions

In this paper, we have proposed a vector-autoregression elastic-net (VAR-ENET) model on the differentiated log-mortality, leading to three key results. First, this new high-dimensional time series analysis outperforms in fitting the mortality rate series of each of the nine populations we considered, compared to the benchmark models (LC and M7). Our forecasting performances are at least comparable to the best model among the LC and the M7 models. Furthermore, thanks to its data-driven approach, the VAR-ENET leads to more stable errors than the standard factor models over populations, showing its power of adaptability to the specific mortality dynamics of different populations. Compared to the usual approach which requires to compare a variety of possible models and then select the best for a particular age-period population, our approach gives directly and with little effort a serious candidate for a consistent modeling of the mortality, regardless of the population features. The second key result is that, although we let a large freedom in the spatio-temporal dependence structure without imposing a priori constraints, the VAR-ENET model enlightens two main effects: the so-called cohort effect and a specific age effect. While the first one has already been well studied in many papers on mortality modeling, the latter is less known in the literature. Finally, contrary to the standard factor models, the extension of the VAR-ENET to multi-population mortality modeling is straightforward, and does not need any constraint on the number of populations or on the hierarchy between them.

Some points should however be improved and need further researches. The first one concerns the interpretation of the results given by our VAR-ENET model. Although it seems to have a better forecasting and adaptability power than the standard factor-based models, the last ones do benefit from a greater interpretability. Indeed, even if most of the coefficients in the autoregressive matrices are estimated to zero in the VAR-ENET and that the non-null coefficients seems to form specific patterns, the comprehension of the underlying dynamics remains complex. On the contrary, it is much easier to understand the mortality dynamics in terms of period, age and cohort effects, which are directly visible through the use of the classical factor-based models.

Second, we are also aware that some of the hyper-parameter selection techniques we applied can be improved. Firstly, we imposed the lag order p equals to 7 for all the population. In a sensibility analysis, we have noted that according to the considered population, the highest predictive power of the VAR-ENET(p) model is not reached at the same lag order p . These results suggest that an optimization on the hyper-parameter p could be developed. The second hyper-parameter to be improved is the mixing weight parameter α between the LASSO and ridge penalties. In the general

context of elastic-net regression, it is usually selected with a grid search. However, Friedman et al. (2010) propose to optimize it through a cross-validation by following the same methodology as the λ selection.

Finally, the log-mortality rates series $y_{i,t}$ are known to not be stationary, but also to be cointegrated (see e.g. Chen et al., 2015; Salhi and Loisel, 2017; Li and Lu, 2017). In our paper, we choose to study the first difference in the log mortality-rates and, by doing so, we lose some information about the long-term co-movement. Another way that we can deal with the non-stationarity and cointegration is to rather select the Vector Error Correcting Model (VECM). Nevertheless, although high-dimensional VAR model has been relatively well studied and recently documented especially in financial econometrics, VECM sparse estimation with elastic-net or other techniques seems to be a new field (see e.g. Wilms and Croux, 2016), and could be developed further for the mortality projection. A major improvement of our model would be to implement the elastic-net procedure to VECM estimation and apply it to the log-mortality series.

Acknowledgments

This work benefited from the financial support of the ANR project "LoLitA" (ANR-13-BS01-0011).

References

- Barrieu, P., Bensusan, H., El Karoui, N., Hillairet, C., Loisel, S., Ravanelli, C., and Salhi, Y. (2012). Understanding, modelling and managing longevity risk: key issues and main challenges. *Scandinavian Actuarial Journal* 2012.3, pp. 203–231. DOI: [10.1080/03461238.2010.511034](https://doi.org/10.1080/03461238.2010.511034).
- Basu, S., Michailidis, G., and others (2015). Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics* 43.4, pp. 1535–1567.
- Bensusan, H., Boumezoued, A., El Karoui, N., and Loisel, S. (2010-2017). Bridging the gap from microsimulation practice to population models: a survey. Work in progress.
- Bickel, P. J. and Levina, E. (2008). Covariance regularization by thresholding. *The Annals of Statistics*, pp. 2577–2604.
- Bien, J. and Tibshirani, R. J. (2011). Sparse estimation of a covariance matrix. *Biometrika* 98.4, pp. 807–820.
- Bohk-Ewald, C. and Rau, R. (2017). Probabilistic mortality forecasting with varying age-specific survival improvements. *Genus* 73.1, p. 1. DOI: [10.1186/s41118-016-0017-8](https://doi.org/10.1186/s41118-016-0017-8).
- Booth, H. and Tickle, L. (2008). Mortality Modelling and Forecasting: a Review of Methods. *Annals of Actuarial Science* 3.1-2, pp. 3–43. DOI: [10.1017/S1748499500000440](https://doi.org/10.1017/S1748499500000440).
- Booth, H., Maindonald, J., and Smith, L. (2002). Applying Lee-Carter under conditions of variable mortality decline. *Population studies* 56.3. 00355, pp. 325–336.
- Boumezoued, A. (2016a). Approches micro-macro des dynamiques de populations hétérogènes structurées par âge. Application aux processus auto-excitants et à la démographie. PhD thesis. Université Pierre et Marie Curie.
- (2016b). Improving HMD mortality estimates with HFD fertility data. Working paper.
- Börger, M., Fleischer, D., and Kuksin, N. (2014). Modeling the Mortality Trend under Modern Solvency Regimes. *ASTIN Bulletin* 44.01, pp. 1–38. DOI: [10.1017/asb.2013.24](https://doi.org/10.1017/asb.2013.24).
- Brouhns, N., Denuit, M., and Vermunt, J. K. (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics* 31.3, pp. 373–393.

- Cairns, A. J. G., 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, pp. 687–718. DOI: [10.1111/j.1539-6975.2006.00195.x](https://doi.org/10.1111/j.1539-6975.2006.00195.x).
- (2008). Modelling and management of mortality risk: a review. *Scandinavian Actuarial Journal* 2008.2-3, pp. 79–113. DOI: [10.1080/03461230802173608](https://doi.org/10.1080/03461230802173608).
- Cairns, A. J. G., Kallestrup-Lamb, M., Rosenskjold, C. P., Blake, D., Dowd, K., and others (2016a). Modelling Socio-Economic Differences in the Mortality of Danish Males Using a New Affluence Index. Tech. rep. Working paper. Department of Economics and Business Economics, Aarhus University.
- Cairns, A. J. G., Blake, D., Dowd, K., and Kessler, A. R. (2016b). Phantoms never die: living with unreliable population data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 179.4, pp. 975–1005. DOI: [10.1111/rssa.12159](https://doi.org/10.1111/rssa.12159).
- 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, pp. 1–35.
- Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., and Khalaf-Allah, M. (2011). Bayesian Stochastic Mortality Modelling for Two Populations. *ASTIN Bulletin: The Journal of the International Actuarial Association* 41.01, pp. 29–59. DOI: [10.2143/AST.41.1.2084385](https://doi.org/10.2143/AST.41.1.2084385).
- Chai, C. M. H., Siu, T. K., and Zhou, X. (2013). A double-exponential GARCH model for stochastic mortality. *European Actuarial Journal* 3.2. 00003, pp. 385–406. DOI: [10.1007/s13385-013-0077-5](https://doi.org/10.1007/s13385-013-0077-5).
- Chatterjee, A. and Lahiri, S. N. (2011). Bootstrapping lasso estimators. *Journal of the American Statistical Association* 106.494, pp. 608–625.
- Chen, H., MacMinn, R., and Sun, T. (2015). Multi-population mortality models: A factor copula approach. *Insurance: Mathematics and Economics*. Special Issue: Longevity Nine - the Ninth International Longevity Risk and Capital Markets Solutions Conference 63, pp. 135–146. DOI: [10.1016/j.insmatheco.2015.03.022](https://doi.org/10.1016/j.insmatheco.2015.03.022).
- Christiansen, M. C., Spodarev, E., and Unseld, V. (2015). Differences in European Mortality Rates: A Geometric Approach on the Age–Period Plane. *ASTIN Bulletin: The Journal of the International Actuarial Association* 45.03, pp. 477–502. DOI: [10.1017/asb.2015.13](https://doi.org/10.1017/asb.2015.13).
- Doukhan, P., Pommeret, D., Rynkiewicz, J., and Salhi, Y. (2017). A class of random field memory models for mortality forecasting. *Insurance: Mathematics and Economics*. In Press. DOI: [10.1016/j.insmatheco.2017.08.010](https://doi.org/10.1016/j.insmatheco.2017.08.010).
- Dowd, K., Cairns, A. J. G., Blake, D., Coughlan, G. D., and Khalaf-Allah, M. (2011). A Gravity Model of Mortality Rates for Two Related Populations. *North American Actuarial Journal* 15.2, pp. 334–356. DOI: [10.1080/10920277.2011.10597624](https://doi.org/10.1080/10920277.2011.10597624).
- Enchev, V., Kleinow, T., and Cairns, A. J. G. (2016). Multi-population mortality models: fitting, forecasting and comparisons. *Scandinavian Actuarial Journal* 0.0, pp. 1–24. DOI: [10.1080/03461238.2015.1133450](https://doi.org/10.1080/03461238.2015.1133450).
- Fan, J., Lv, J., and Qi, L. (2011). Sparse High Dimensional Models in Economics. *Annual review of economics* 3, pp. 291–317. DOI: [10.1146/annurev-economics-061109-080451](https://doi.org/10.1146/annurev-economics-061109-080451).
- Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software* 33.1, pp. 1–22.
- Furman, Y. (2014). VAR Estimation with the Adaptive Elastic Net. SSRN Scholarly Paper ID 2456510. Rochester, NY: Social Science Research Network.
- Gefang, D. (2014). Bayesian doubly adaptive elastic-net Lasso for VAR shrinkage. *International Journal of Forecasting* 30.1, pp. 1–11.

- Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica* 37.3, pp. 424–438. DOI: [10.2307/1912791](https://doi.org/10.2307/1912791).
- Haberman, S. and Renshaw, A. (2012). Parametric mortality improvement rate modelling and projecting. *Insurance: Mathematics and Economics* 50.3, pp. 309–333. DOI: [10.1016/j.insmatheco.2011.11.005](https://doi.org/10.1016/j.insmatheco.2011.11.005).
- Hahn, L. J. (2014). A Bayesian Multi-Population Mortality Projection Model. Master thesis. Ulm, Germany: Universität Ulm.
- Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12.1, pp. 55–67.
- Human Mortality Database (2016). University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany).
- Hunt, A. and Blake, D. (2014). A General Procedure for Constructing Mortality Models. *North American Actuarial Journal* 18.1, pp. 116–138. DOI: [10.1080/10920277.2013.852963](https://doi.org/10.1080/10920277.2013.852963).
- Hunt, A. and Villegas, A. M. (2015). Robustness and convergence in the Lee–Carter model with cohort effects. *Insurance: Mathematics and Economics* 64, pp. 186–202.
- Huynen, M. M., Martens, P., Schram, D., Weijenberg, M. P., and Kunst, A. E. (2001). The impact of heat waves and cold spells on mortality rates in the Dutch population. *Environmental Health Perspectives* 109.5, pp. 463–470.
- Izraelewicz, E. (2012). L’effet moisson - l’impact des catastrophes vie sur la mortalité à long terme - Exemple de la canicule de l’été 2003. *Bulletin Français d’Actuariat* 12.24, pp. 113–159.
- Jarner, S. F. and Kryger, E. M. (2011). Modelling Adult Mortality in Small Populations: The Saint Model. *ASTIN Bulletin: The Journal of the International Actuarial Association* 41.02, pp. 377–418. DOI: [10.2143/AST.41.2.2136982](https://doi.org/10.2143/AST.41.2.2136982).
- Lee, R. D. and Carter, L. R. (1992). Modeling and Forecasting U. S. Mortality. *Journal of the American Statistical Association* 87.419, pp. 659–671. DOI: [10.2307/2290201](https://doi.org/10.2307/2290201).
- Li, H. and Lu, Y. (2017). Coherent Forecasting of Mortality Rates: A Sparse Vector-Autoregression Approach. *ASTIN Bulletin: The Journal of the IAA* 47.2, pp. 563–600. DOI: [10.1017/asb.2016.37](https://doi.org/10.1017/asb.2016.37).
- Li, N. and Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography* 42.3, pp. 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 projection. *Demography* 50.6, pp. 2037–2051. DOI: [10.1007/s13524-013-0232-2](https://doi.org/10.1007/s13524-013-0232-2).
- Opgen-Rhein, R. and Strimmer, K. (2007). From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC systems biology* 1.1, p. 37.
- Perron, P. (1988). Trends and random walks in macroeconomic time series: Further evidence from a new approach. *Journal of economic dynamics and control* 12.2, pp. 297–332.
- Plat, R. (2009). On stochastic mortality modeling. *Insurance: Mathematics and Economics* 45.3, pp. 393–404.
- R Core Team (2017). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria.
- Renshaw, A. E. and Haberman, S. (2006). A cohort-based extension to the Lee–Carter model for mortality reduction factors. *Insurance: Mathematics and Economics* 38.3, pp. 556–570. DOI: [10.1016/j.insmatheco.2005.12.001](https://doi.org/10.1016/j.insmatheco.2005.12.001).
- (2008). On simulation-based approaches to risk measurement in mortality with specific reference to Poisson Lee–Carter modelling. *Insurance: Mathematics and Economics* 42.2, pp. 797–816. DOI: [10.1016/j.insmatheco.2007.08.009](https://doi.org/10.1016/j.insmatheco.2007.08.009).

- Said, S. E. and Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 71.3, 03347, pp. 599–607.
- Salhi, Y. and Loisel, S. (2017). Basis risk modelling: a cointegration-based approach. *Statistics* 51.1, pp. 205–221. DOI: [10.1080/02331888.2016.1259806](https://doi.org/10.1080/02331888.2016.1259806).
- Schäfer, J. and Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology* 4.1.
- Song, S. and Bickel, P. J. (2011). Large vector auto regressions. *arXiv preprint arXiv:1106.3915*.
- Spodarev, E., Shmileva, E., and Roth, S. (2013). Extrapolation of stationary random fields. *arXiv preprint arXiv:1306.6205*.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*. 20416, pp. 267–288.
- Toulemon, L. and Barbieri, M. (2008). The mortality impact of the August 2003 heat wave in France: Investigating the ‘harvesting’ effect and other long-term consequences. *Population Studies* 62.1, pp. 39–53. DOI: [10.1080/00324720701804249](https://doi.org/10.1080/00324720701804249).
- Vazzoler, S., Frattarolo, L., and Billio, M. (2016). sparsevar: A Package for Sparse VAR/VECM Estimation. R package version 0.0.10.
- Villegas, A. M., Kaishev, V., and Millossovich, P. (2017). StMoMo: An R Package for Stochastic Mortality Modelling. R package version 0.4.0.
- Wilms, I. and Croux, C. (2016). Forecasting using sparse cointegration. *International Journal of Forecasting* 32.4, pp. 1256–1267.
- Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67.2, pp. 301–320.

Appendix 1 Improvement rates data for females and males

Figures 9 and 10 describe the log-mortality improvements for females and males.

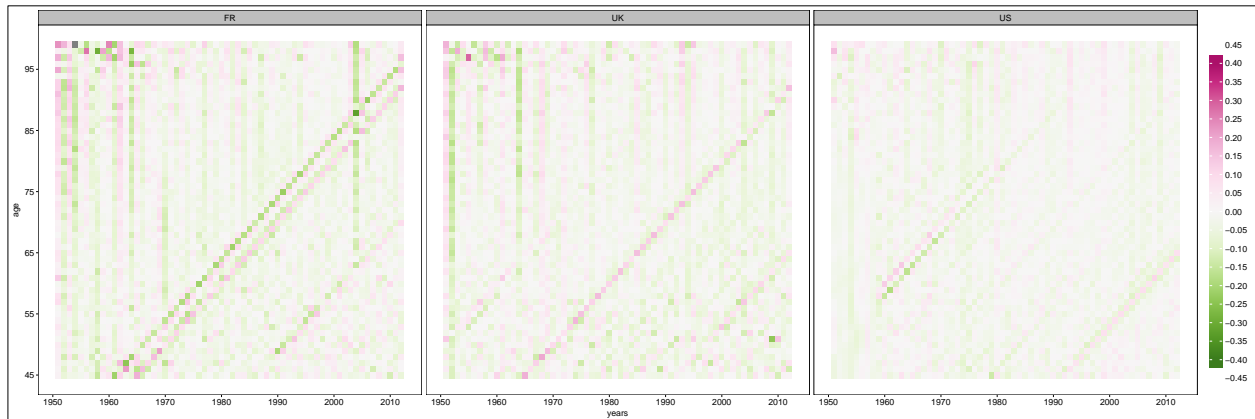


Figure 9: The period log-mortality improvements for England and Wales (UK), the United States (US) and the France (FR) on the age-period observation $\{45, \dots, 99\} \times \{1950, \dots, 2012\}$ for females.

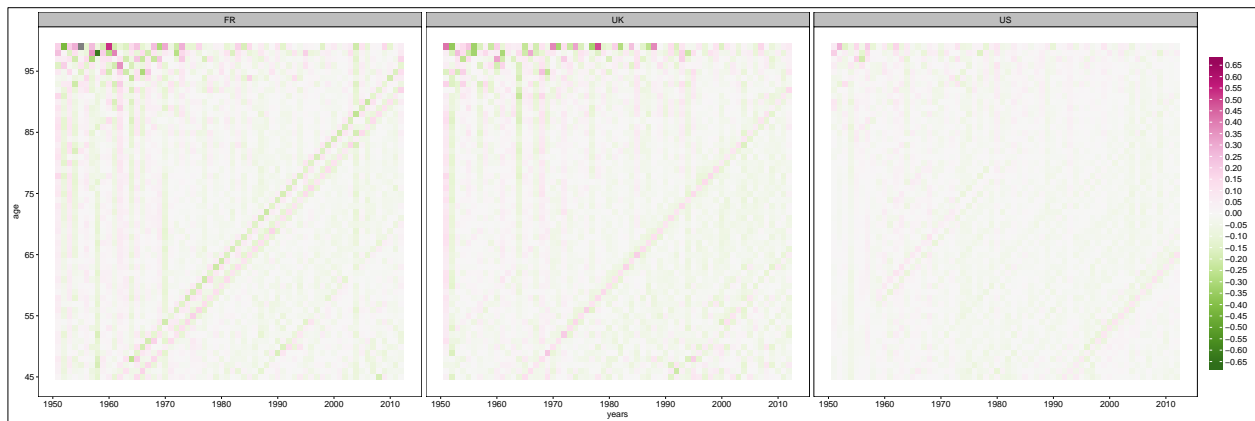


Figure 10: The period log-mortality improvements for England and Wales (UK), the United States (US) and the France (FR) on the age-period observation $\{45, \dots, 99\} \times \{1950, \dots, 2012\}$ for males.

Appendix 2 In-sample analysis for females and males

Figures 11 and 12 present the in-sample performance in terms of RSME for females. Figures 13 and 14 present the in-sample performance in terms of RSME for males.

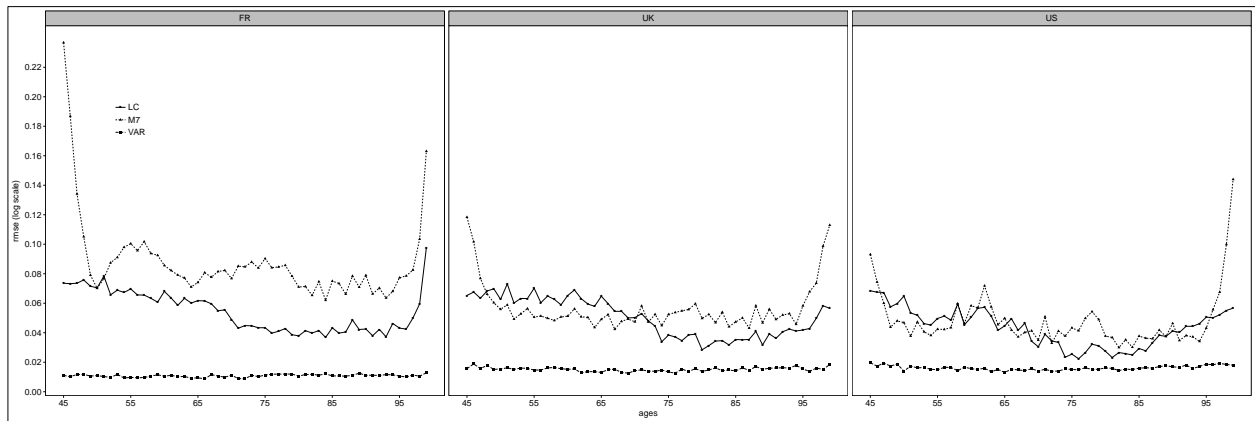


Figure 11: The RMSE for England and Wales (UK), the United States (US) and the France (FR) grouped by age for the female populations.

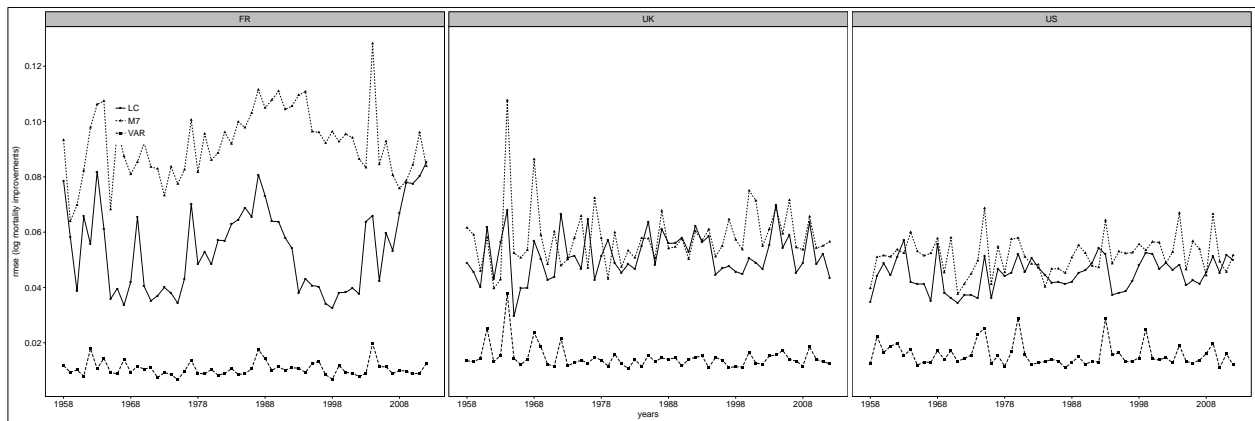


Figure 12: The RMSE for England and Wales (UK), the United States (US) and the France (FR) grouped by period for the female populations.

Appendix 3 Out-of-sample analysis for females and males

Tables 4 and 5 present the out-of-sample performance in terms of RMSFE for different model estimation period, respectively $\{1970, \dots, 2000\}$ and $\{1980, \dots, 2000\}$.

Figures 15 and 16 present the out-of-sample performance in terms of RSMFE for females. Figures 17 and 18 present the out-of-sample performance in terms of RSMFE for males.

Table 4: The RSMFE of the LC, the M7 and the VAR models estimated on the period 1970 – 2000.

| Country | Model | RSMFE Female | RSMFE Male | RSMFE Overall |
|---------|-------|--------------|------------|---------------|
| FR | LC | 0.077 | 0.079 | 0.056 |
| FR | M7 | 0.203 | 0.120 | 0.111 |
| FR | VAR | 0.068 | 0.072 | 0.058 |
| UK | LC | 0.124 | 0.125 | 0.118 |
| UK | M7 | 0.123 | 0.101 | 0.091 |
| UK | VAR | 0.094 | 0.081 | 0.069 |
| US | LC | 0.121 | 0.101 | 0.081 |
| US | M7 | 0.147 | 0.106 | 0.101 |
| US | VAR | 0.073 | 0.081 | 0.086 |

Note: This table reports the out-of-sample performance *via* the RSMFE values for the LC, the M7 and the VAR-ENET (4) models estimated on the period 1970 – 2000. We compare this indicator for males, females and the overall populations for FR, UK and US.

Table 5: The RSMFE of the LC, the M7 and the VAR models estimated on the period 1980 – 2000.

| Country | Model | RSMFE Female | RSMFE Male | RSMFE Overall |
|---------|-------|--------------|------------|---------------|
| FR | LC | 0.063 | 0.085 | 0.066 |
| FR | M7 | 0.117 | 0.101 | 0.096 |
| FR | VAR | 0.069 | 0.082 | 0.064 |
| UK | LC | 0.106 | 0.111 | 0.103 |
| UK | M7 | 0.098 | 0.092 | 0.088 |
| UK | VAR | 0.088 | 0.081 | 0.070 |
| US | LC | 0.093 | 0.096 | 0.080 |
| US | M7 | 0.129 | 0.103 | 0.107 |
| US | VAR | 0.097 | 0.093 | 0.082 |

Note: This table reports the out-of-sample performance *via* the RSMFE values for the LC, the M7 and the VAR-ENET (3) models estimated on the period 1980 – 2000. We compare this indicator for males, females and the overall populations for FR, UK and US.

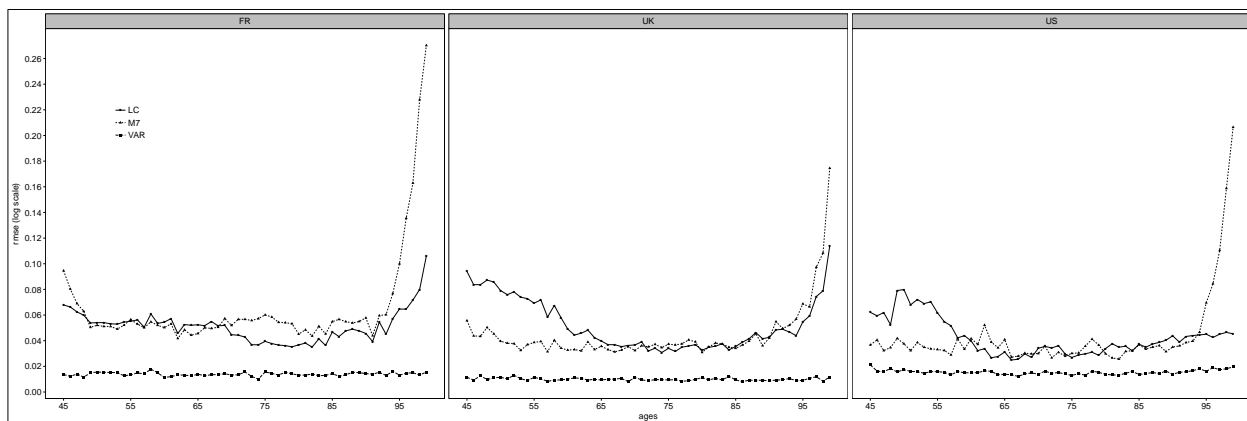


Figure 13: The RMSE for England and Wales (UK), the United States (US) and the France (FR) grouped by age for the male populations.

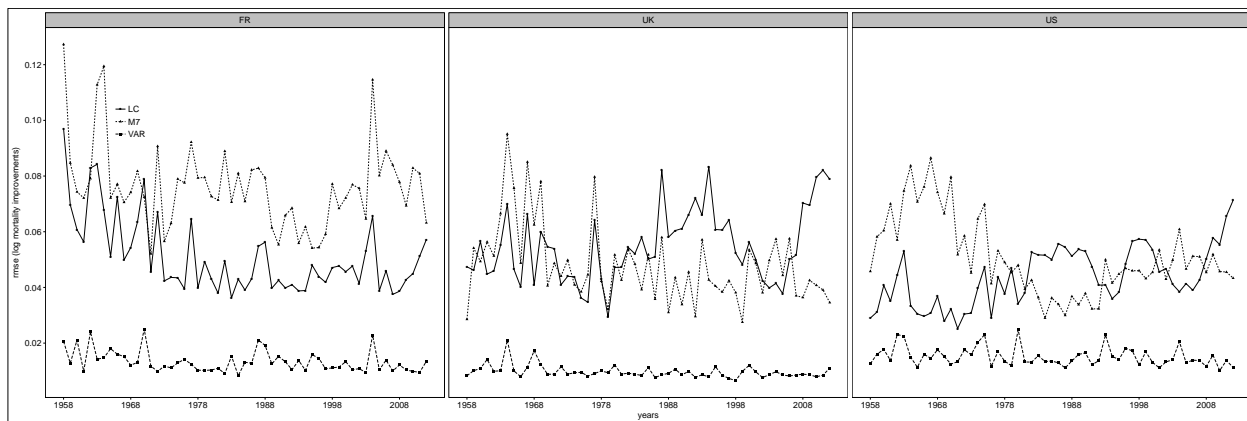


Figure 14: The RMSE for England and Wales (UK), the United States (US) and the France (FR) grouped by period for the male populations.

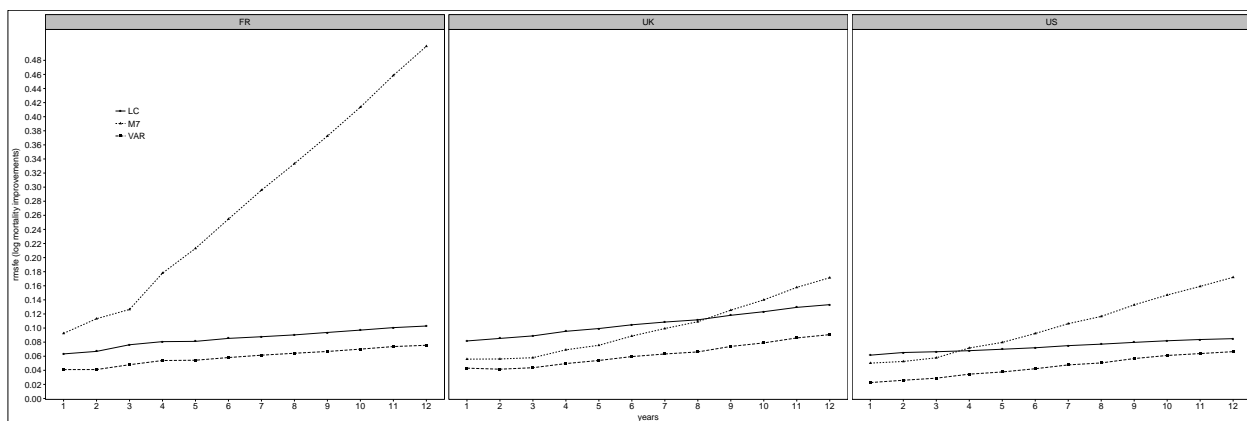


Figure 15: The RMSFE for England and Wales (UK), the United States (US) and France (FR) grouped by period for the female populations.

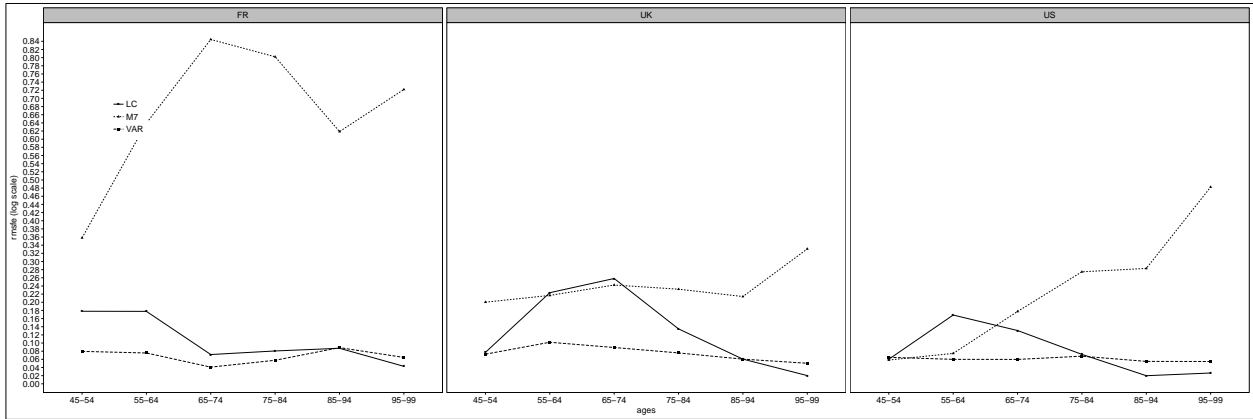


Figure 16: The RMSFE for England and Wales (UK), the United States (US) and France (FR) grouped by age for the female populations.

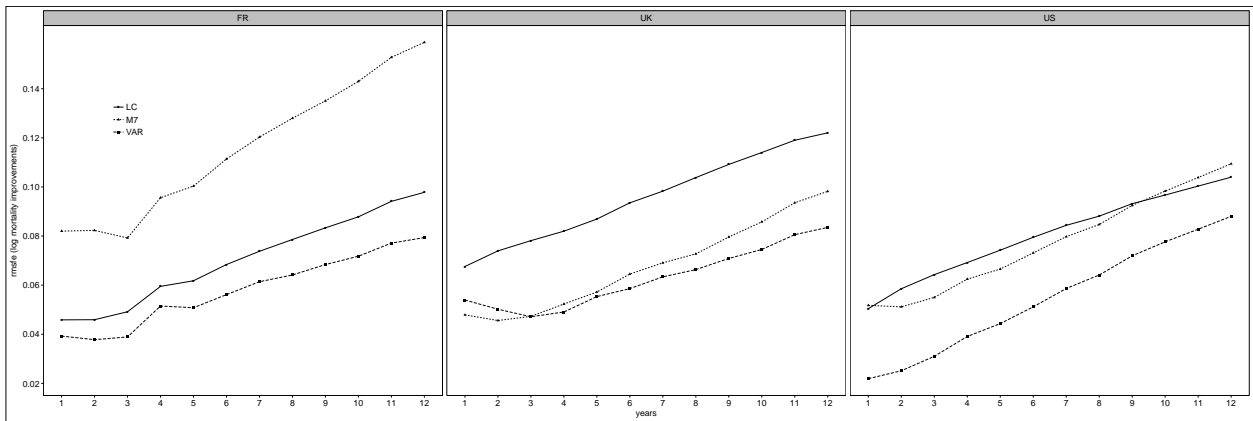


Figure 17: The RMSFE for England and Wales (UK), the United States (US) and France (FR) grouped by period for the male populations.

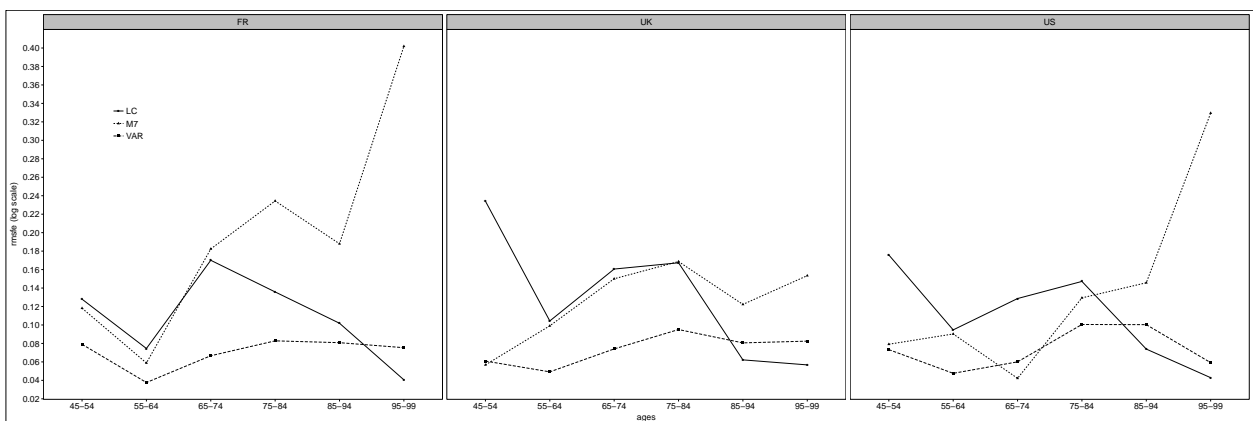


Figure 18: The RMSFE for England and Wales (UK), the United States (US) and France (FR) grouped by age for the male populations.