MODELING LONGEVITY RISK WITH GENERALIZED DYNAMIC FACTOR MODELS AND VINE-COPULAE

BY

HELENA CHULIÁ, MONTSERRAT GUILLÉN AND JORGE M. URIBE

ABSTRACT

We present a methodology to forecast mortality rates and estimate longevity and mortality risks. The methodology uses generalized dynamic factor models fitted to the differences in the log-mortality rates. We compare their prediction performance with that of models previously described in the literature, including the traditional static factor model fitted to log-mortality rates. We also construct risk measures using vine-copula simulations, which take into account the dependence between the idiosyncratic components of the mortality rates. The methodology is applied to forecast mortality rates for a population portfolio for the UK and to estimate longevity and mortality risks.

KEYWORDS

Longevity, mortality forecasting, factor models, vine-copulae, value-at-risk.

1. INTRODUCTION AND BACKGROUND

Longevity risk (LR), understood as the downward deviation in population mortality rates from the mortality trend forecast, has become a topic of great academic interest in recent times, especially in the actuarial literature. This interest is completely justified, given the fact that LR affects different products and contracts involving public and private sector agents alike. For instance, LR has an impact on immediate and deferred annuities, enhanced and impaired annuities, guaranteed annuity options, lifetime mortgages and, more importantly, defined benefit pension schemes (Richards and Jones, 2004).

Among these pension schemes, LR is relevant not only for insurance companies or private employers, who promise a pension on retirement based on the employee's final salary, but also for institutions within the public sector, which typically offer generous final salary benefits, albeit largely unfunded. MacMinn *et al*. (2006) and Dushi *et al*. (2010) provide further insights into the question of LR in the context of defined benefit pension schemes. LR is of particular

importance in Europe where marked reductions in mortality rates have been recorded over the last century, much of these trends affecting age groups above 65 years.

The study of LR has been undertaken from various perspectives. On the one hand, many authors have documented the importance of the risk in terms of its expected impact on the solvency of those insurance companies particularly vulnerable to it (Hári *et al.*, 2008; Hanewald *et al.*, 2011; Olivieri, 2011). On the other hand, many others have developed analytical frameworks aimed at developing hedging strategies for use by practitioners at insurance companies and pension funds facing LR. These strategies are based on the use of various financial instruments ranging from longevity bonds to survivor swaps (Dahl *et al.*, 2008; Cairns, 2011; Ngai and Sherris, 2011; Blake *et al.*, 2014; Lorson and Wagner, 2014; Wong *et al.*, 2014; among others).

The key to understanding LR lies in the stochastic nature of mortality rates, which means their forecast is subject to uncertainty and judgments concerning statistical confidence. The stochastic modeling of mortality rates has been well documented in the literature (Continuous Mortality Investigation, 2004, 2005, 2013; Cairns *et al.*, 2011) and is recommended by the regulators (Hollmann *et al.*, 2000).

Among the alternatives available, factor models are an attractive approach, due to the low frequency (i.e. annual) of the mortality data and the high relative number of specific mortality rates to be forecast. Specific mortality rates, which discriminate between age and sex, are thought to be more appropriate for dealing with mortality projections than aggregate rates, due to the heterogeneity of a population's mortality. Factor models allow the researcher to reduce the dimensionality problem and to construct more accurate forecasts. The strategy consists in making the mortality rates dependent on just a few unobserved stochastic factors, for example, stochastic trends. By so doing, the number of estimated parameters in the model is significantly reduced and optimal first-efficient forecasts are possible.

More recently, the literature has begun to recognize the importance of dependence relationships in different spectral frequencies of mortality rates, in order to make in- and out-of-sample estimations more accurately. For instance, some studies highlight the importance of cointegration among the variables and the need to pre-test for unit roots and multivariate cointegration before using traditional factor models or vector autoregression (VAR) models (Njenga and Sherris, 2011; Torri, 2011). These relations can be thought of as arising in the low frequency domain of the multivariate spectral density of the specific mortality rates.

However, the dependence patterns that emerge in the high and medium frequency domains of the spectra have been largely unexplored, although some studies have documented the importance of the dependence relationships between contiguous categories of mortality rates or their improvements (Denton *et al.*, 2005; Wills and Sherris, 2008; D'Amato *et al.*, 2012; Lin *et al.*, 2015).

In this article, we propose a methodology that makes advances on two different fronts: point estimation and forecasts based on generalized dynamic factor models (GDFM), and the construction of alternative confidence scenarios for these forecasts. The task we report in the second part of this article involves the estimation of the multivariate probability density function of the forecasted mortality rates and not only their first moment. By so doing, we are able to calculate risk measures that take into account linear and non-linear dependences among mortality rate variations by means of vine-copulae. In this way, we provide a robust alternative for measuring longevity (and mortality) risk using distorted risk measures, such as value-at-risk (VaR), or tail value-at-risk (TVaR).

The main novelty in the approach we report lies in the fact that we outline the necessary steps for blending GDFMs (hitherto unknown in the field of actuarial sciences) with vine-copula models (used here for the first time in addressing this specific task), so that the construction of risk measures is fully intuitive and rigorous, in the sense that the dependence of the contiguous categories of the mortality rates is fully incorporated into the construction of an insurance company's performance scenarios.

The methodology is applied to forecast mortality rates and to estimate risk measures for the UK, using data for 1950–2011 drawn from the Human Mortality Database. We expand the current literature in several directions. First, we compare the performance of four factor models in forecasting mortality rates: the first is a traditional (static) factor model that assumes the presence of one common stochastic trend in the data (better known in the actuarial literature as the Lee–Carter model); the second works with non-common trends but with stationary and common dynamic factors fitted to the differences of the log-mortality rates; the third is a mix of these first two and is a dynamic factor model (DFM) of the log-mortality rates, and, finally, a GDFM is constructed of the differences of the log-mortality rates. GDFMs are widely employed in the econometrics literature, but are largely unknown in the actuarial literature, in spite of their good properties for estimating unobservable factors (Forni *et al.*, 2000, 2004) and for forecasting (Forni *et al.*, 2005).

Second, we incorporate the explicit modeling of the dependence structure in the construction of scenarios for mortality rate projections. We also propose VaRs and TVaRs as a way to measure LR, and we describe how they should be estimated in this context. They should prove useful when assessing suitable capital requirements for the operation of firms exposed to such risk.

Finally, we seek to link the econometrics literature studying DFM, unit roots and cointegration with the actuarial literature examining mortality forecasting. Thus, we reference crucial findings in this branch of econometrics that should enrich the ongoing analysis of LR within actuarial science.

Our study is related to Plat (2011), in which the author estimates the VaR of a portfolio comprising 45,000 male and 36,000 female policyholders aged 65 and older. His approach relies on a two-step algorithm: first, he estimates the common trends on a dynamic basis (using blocks of 30 years' length); and, second, he estimates the stochastic variation around the trend projection. In

this way, he is able to simulate different scenarios of operation for the insurance company and to calculate the net asset value (NAV) of the company in each case. This allows him to estimate the VaR of the company at different levels of confidence. However, here, unlike Plat (2011), we do not assume a cointegration relationship operating on the 220 mortality rates of our empirical estimations. Therefore, our forecasting exercise only includes the modeling of the stochastic volatility around the trend by way of an alternative to the second step in Plat's algorithm.

The rest of this article is organized as follows. First, we present the main theoretical points regarding our methodological approach and we discuss the models and the estimation strategy. In Section 3, we explore the relationship between the models explained in Section 2 and some of the most popular factor models for the forecasting of mortality rates in the actuarial literature. In Section 4, we describe our data in terms of their stochastic time-series properties, such as the presence of unit roots and variability explained by the first principal components (PC). In Section 5, we present our main results and discuss our principal findings. Lastly, we conclude.

2. Methodology

In this section, we present the forecasting methodologies employed in the empirical section. We also explain how to estimate LR so as to incorporate the information about dependence in mortality series in the estimations. In most of Sections 2.1 and 2.2 below, we follow the notation and presentation in Bai and Ng (2008).

2.1. Factor models

Let N be the number of cross-sectional units and T be the number of time series observations. In our case, we have 220 cross-sectional units (mortality rates for ages from 0 to 109+ years, for males and females). If we consider males and females separately then, $N = 110$. For $i = 1, \ldots, N$ and $t = 1, \ldots, T$. The static factor model (SFM) is defined as follows:

$$\begin{aligned} x_{\text{it}} &= \lambda_i F_t + e_{\text{it}}, \\ x_{\text{it}} &= C_{\text{it}} + e_{\text{it}}, \end{aligned} \tag{2.1}$$

where e_{it} is referred to as the idiosyncratic error and λ_i is referred to as the factor loading. This is a vector of weights that unit i places on the corresponding r static common factors F_t. $C_{\text{it}} = \lambda_i F_t$ refers to the common component of the model. If we define $X_t = (x_{1t}, x_{2t}, \ldots, x_{Nt})'$ and $\Lambda = (\lambda_1, \ldots, \lambda_N)'$, in vector form, for each period, we have

$$\underset{(N \times 1)}{X_t} = \underset{(N \times r)\,(r \times 1)}{\Lambda F_t} + \underset{(N \times 1)}{e_t}, \tag{2.2}$$

where $e_t = (e_{1t}, e_{2t}, \ldots, e_{Nt})'$. Notice that although the model specifies a static relationship between x_{it} and F_t, F_t itself can be a dynamic vector process. In the case that F_t and X_t are jointly stationary (i.e. either each series is stationary or *all* of them are cointegrated), F_t can be thought to evolve according to a vector autoregressive (VAR) process:

$$A(L) F_t = u_t, \tag{2.3}$$

where $A(L)$ is a polynomial of the lag operator. The SFM has been implemented in several studies to forecast mortality rates (see, for example, Alonso (2008)). Note that if we set $L = 1$ and $A(1) = I_r$, where I_r is the identity of order r, the model in (2.3) becomes a multivariate random walk (MRW). Should we also require that $r = 1$, then we are in the presence of the model proposed by Lee and Carter (1992).

In the general case F_t, could contain stationary and non-stationary factors, as in Peña and Poncela (2006) and Bai and Ng (2004). However, for empirical applications it is convenient to restrict our attention to cases in which *all* the factors in F_t are stationary or that in which *all* of them evolve following the same stochastic trends (by assumption). This is especially true for models with very large N, in which the implementation of traditional cointegration tests based on VAR representations of the original variables are not suitable.

The static model can be compared with the DFM, defined as

$$x_{it} = \lambda_i (L) f_t + e_{it}, \tag{2.4}$$

where $\lambda_i (L) = \left(1 - \lambda_{i1} L - \ldots - \lambda_{is} L^S\right)$ is a vector of dynamic factor loadings of order s. In the case when s is finite, we refer to it as a DFM, whereas a GDFM allows s to be infinite. Stock and Watson (2010) provide examples of the former and Forni *et al.* (2000) introduce the latter. In either case, the (dynamic) factors f_t evolve according to

$$f_t = C(L) \varepsilon_t, \tag{2.5}$$

where ε_t are *iid* errors. The dimension of f_t, denoted q, is the same as the dimension of ε_t.

An additional classification of the models in equations (2.2)–(2.3) and (2.3)–(2.4) concerns the fact as to whether the idiosyncratic disturbances in (2.2) or (2.4) are allowed to be weakly correlated. When they are not, it becomes an exact factor model; when they are, the model is an approximate factor model.

2.2. Identification

We can rewrite the model in (2.4) in static form, simply by redefining the vector of factors to contain the dynamic factors and their lags, and the matrix of loads accordingly. In this case both, SFM and DFM can be presented in matrix form as follows:

$$\underset{(N \times T)}{X} = \underset{(N \times r)(r \times T)}{\Lambda F} + \underset{(N \times T)}{e}, \tag{2.6}$$

where $X = (X_1, \ldots, X_N)$ and $F = (F_1, \ldots, F_T)$. Clearly, F and Λ are not separately identifiable. For any arbitrary $(r \times r)$ invertible matrix H, $F\Lambda' = FHH^{-1}\Lambda' = F^*\Lambda'^*$, where $F^* = F\Lambda$ and $\Lambda^* = \Lambda H^{-1}$, the factor model is observationally equivalent to $X = F^*\Lambda'^* + e$. Therefore, r^2 restrictions are required to uniquely fix F and Λ (Bai and Wang, 2012). Many alternatives are available in the literature to achieve this goal. For example, Harvey (1990), Zuur *et al.* (2003) and Holmes *et al.* (2014) propose the following procedure: In the first N-1 rows of Λ, the λ-value in the jth column and ith row is set to zero if $j > i$. The intercept is constrained so that each of the time series in F_t has a mean equal to zero over time (from $t = 0$ to $t = T$). The matrix of second moments $[e'e]$, is set equal to the identity matrix of order N, I_N.

Notice that the estimation of the factors using PC or singular value decomposition (SVD), by construction, imposes the normalization that $\frac{\Lambda'\Lambda}{N} = I_r$ and that $F'F$ is diagonal, which are enough to guarantee identification (up to a column sign rotation).

2.3. Factor estimation

There are a number of alternatives for estimating the models in equations (2.1)–(2.3) and (2.2)–(2.4). One of them involves using PC or equivalently SVD to estimate the factors and their loadings.

An alternative for estimating (2.2)–(2.3), under the assumption of Gaussian errors and possibly a MRW structure, arises from noticing that (2.2) and (2.3) can be thought of as a state-space representation, where the transition equation is a first-order Markov process. In this case, equation (2.3) is the hidden state vector, unobservable by definition, and (2.1) is the output, or measurement equation. Therefore, the model can be estimated by maximum likelihood, either, by using an expectation maximization algorithm, or other numeric-optimization algorithms (Hamilton, 1994; Holmes *et al.*, 2014). As is well known, in this context the Kalman filter is an optimal estimator of the parameters in the model, and the Kalman smoother can be used to estimate the unobservable factors. Here, we prefer to estimate the factors using PC, in the interests of parsimony and because of the documented advantages of this method in terms of model specification (Stock and Watson, 2002; Bates *et al.*, 2013).

The GDFM uses a two-step estimation strategy. First, the variance-covariance matrices of the common and the idiosyncratic components in equation (2.1) are estimated, by using the first q dynamic PC operating on the spectral density of x_{it}. Then, the information coming from the first step is used to extract linear combinations of the x's that are more efficient than standard PC. Particularly,

$$\hat{C}_t = \left[\Gamma_0^C \hat{Z}' \left(\hat{Z}\hat{\Gamma}_0 \hat{Z} \right)^{-1} \right] \left(\hat{Z} X_t \right), \tag{2.7}$$

where $\hat{C}_t$ is the estimation of the common component, Γ_0^C and $\hat{\Gamma}_0$ are contemporaneous-covariance matrices of the common components and the x's, respectively. The first matrix is estimated based on spectral density methods.

$\hat{Z}$ are generalized eigenvectors and therefore $\hat{Z}X_t$ are the generalized principal components (GPC).

2.4. Point forecasts

With the factors at hand, the forecasting of the mortality rates by linear regression techniques is straightforward. The factors estimated in a first regression stage by SVD, PC or GPC can be employed in a second regression. Consider forecasting x_{it+1} using all the data in X_t and treat F_t as observed. If e_{it} follows an autoregression and the errors are Gaussian, then

$$E[x_{it+1}| X_t, f_t, X_{t-1}, f_{t-1}, \ldots] = E[\lambda_i(L) f_{t+1} + e_{it+1}| X_t, f_t, X_{t-1}, f_{t-1}, \ldots]$$
$$= E[\lambda_i(L) f_{t+1}| X_t, f_t, X_{t-1}, f_{t-1}, \ldots] + [e_{it+1}| X_t, f_t, X_{t-1}, f_{t-1}, \ldots] \quad (2.8)$$
$$= \alpha(L) f_{t+1} + \delta(L) x_{it}.$$

In the SFM case L disappears. In the forecasting equation (2.8), we could include some other covariates. It would also be possible to index in time the matrix Λ, such that Λ_t contains the time-varying loads of the system. However, as noted by Bates *et al.* (2013), the original DFM or SFM seem to behave very well in the presence of parameter instability and, therefore, time-varying extensions do not provide further insights into the model's structure, but rather they consume degrees of freedom and demand additional restrictions be imposed on it.

Following Stock and Watson (2006), the h-step ahead forecast can be performed directly by projecting x_{it+h} onto the estimated factors lagged h periods, that is, by estimating β in the equation:

$$X_{t+h} = \beta F_t + e_{t+h}. \quad (2.9)$$

Unknown factors can be replaced by their (consistent) estimations $\hat{F}_t$ following Stock and Watson (2006). A direct forecast can be potentially less efficient than an iterated forecast (solving the full DFM or SFM forward using the KF), but it is also more robust to model misspecification. The importance of model misspecification in the particular context of the estimation and forecasting of mortality rates has been documented by Stallard (2006). Notice that β equals Λ when $h = 0$ in equation (2.9).

The GDFM was originally proposed by Forni and Reichlin (1998) and Forni *et al.* (2000). But it was not until Forni *et al.* (2005) that it could be used for forecasting purposes. It is a generalization of the DFM because it allows for a richer dynamic structure of the factors and it does not assume mutual orthogonality of the idiosyncratic components e_{it}. In this context,

$$X_{t+h} = \left[\Gamma_h^C \hat{Z} \left(\hat{Z} \hat{\Gamma}_0 \hat{Z} \right)^{-1} \right] \left(\hat{Z} X_t \right) + e_{t+h}. \quad (2.10)$$

In the empirical application, we opt to ignore the forecasting of the idiosyncratic component in equation (2.8), given that it would appear to be a white noise

process. Thus, we concentrate on the estimation of equations (2.9) and (2.10), based entirely on the information provided by the common factors, leaving the dependence relationship to affect only the "simulated scenarios" of the risk measures (as explained in Section 2.6.). The two-step strategy described, as we will see, allows a better forecasting performance in the majority of cases, but it makes the model weaker in the face of outliers or extreme variations in mortality rates. We discuss this point further in Section 5.

2.5. Dependence between the idiosyncratic components

In spite of the preferred forecasting method, up to this juncture, we have only considered the dependence in the common factors and have said nothing about the possible dependence between the idiosyncratic components. The dependence arising in the noisy high frequency of the spectra is key to the estimation of "unexpected" movements in the time series and, as such, is crucial for the estimation of risk. In this article, we approach this dependence by means of copula functions, and so we are able to construct confidence intervals for our point forecasts, that is, as risk measures based on the percentiles of the simulated density (i.e. VaR and TVaR).

2.5.1. *Copula functions.* Formally, a copula is a multivariate probability distribution such that $C : [0, 1]^N \rightarrow [0, 1]$ where C is the copula and N is the number of mortality rates in our application. Sklar's Theorem (Sklar, 1959) can be used to construct the copula function. The theorem establishes that if F is a joint distribution function with marginals $F_1, \ldots, F_N$, there exists a copula, such that: $F(\underline{e}_1, \ldots, \underline{e}_N) = C(F_1(\underline{e}_1), \ldots, F_N(\underline{e}_N))$, for $\underline{e}_i \in \mathbb{R}$. If the marginals are continuous, C is unique. Otherwise, it is uniquely determined by rank $(F_1) \times$ rank $(F_2) \times \ldots \times$ rank (F_N), Rango$(F_1)\times$, where $rank(F_i) = F_i(\underline{e}_i)$ denotes the rank operator (McNeil *et al.*, 2005).

Before proceeding to the parameter estimation, we need to construct a pseudo-sample defined as

$$F_i(\underline{e}_i) = u_i \forall i = 1, .., N, \qquad (2.11)$$

where $\underline{e}_i = (e_1, \ldots, e_T)$ is a $(T \times 1)$ vector that contains the estimated idiosyncratic components for each individual i. We make use of the empirical cumulative distribution function (CDF) of the idiosyncratic terms, estimated as the residuals of the GDFM, as an approximation to $F_i(\cdot)$ in (2.11). Once the pseudo-sample has been constructed, we need to check its accuracy in describing the data using the Kolmogorov–Smirnov (KS) statistic.

One additional consideration must be highlighted at this stage. When there are several dimensions involved, as in our case, in which we have 110 mortality rates for each sex, the direct estimation of a N-dimensional copula is not recommended. Instead, the literature has developed an alternative estimation and

simulation procedure, based on bivariate conditional copulae (i.e. pair copulae) as described by Aas *et al.* (2009).

2.5.2. *Pair copulae.* Building on the work of Joe (1996) and Bedford and Cooke (2001, 2002), Aas *et al.* (2009) show that multivariate data, which exhibit complex patterns of dependence in the tails, can be modeled using a cascade of pair-copulae, acting on two variables at a time. This approach is particularly attractive in the present context, in which a very large cross-sectional dimension makes traditional high-dimensional copula methods unfeasible. The model construction is hierarchical and the various levels in the model correspond to the incorporation of more variables in the conditioning sets, using pair-copulae as simple building blocks.

Following Aas *et al.* (2009), consider a vector $\underline{e} = \left(\underline{e}_1, \ldots, \underline{e}_N \right)'$ of random variables with a joint density function $f \left(\underline{e}_1, \ldots, \underline{e}_N \right)$. This density can be factorized as follows:

$$
\begin{aligned}
f \left(\underline{e}_1, \ldots, \underline{e}_N \right) \\
= f_n \left(\underline{e}_N \right) \cdot f \left(\underline{e}_{N-1} \middle| \underline{e}_N \right) \cdot f \left(\underline{e}_{N-2} \middle| \underline{e}_{N-1}, \underline{e}_N \right) \cdots f \left(\underline{e}_1 \middle| \underline{e}_2, \ldots, \underline{e}_N \right),
\end{aligned}
\tag{2.12}
$$

and each term in (2.12) can be decomposed into the appropriate pair-copula times a conditional marginal density, using the general formula

$$
f \left(\underline{e} \middle| v \right) = c_{e v_j | v_{-j}} \left\{ F \left(\underline{e} \middle| v_{-j} \right), F \left(v_j \middle| v_{-j} \right) \right\} \cdot f \left(\underline{e} \middle| v_{-j} \right),
\tag{2.13}
$$

for an N-dimensional vector v. Here each v_j is an arbitrarily chosen component of v and v_{-j} denotes the vector v excluding the jth component. Then, $c_{e v_j}$ denotes a pair-copula between $\underline{e}$ and v_j. As noted by Aas *et al.* (2009), under appropriate regularity conditions a multivariate density can be expressed as the product of pair-copulae, acting on several different conditional probability distributions. From their work, it is also clear that the construction is iterative, and that given a specific factorization, there are still many different reparameterizations, which in principle could be used to perform the estimation.

The different constructs of pair-copulae available in the literature can be described, following Bedford and Cooke (2001, 2002), by a graphical model known as regular vines. The vine class is very large and it houses the graphical model used in this study, known as *d-vine* (Kurowicka and Cooke, 2005). A d-vine is a specific way to factorize the multivariate density and it implies the estimation of $N \left(N - 1 \right) / 2$ bivariate copulae in ascending hierarchical order as shown in Figure 1.

Figure 1 shows the specification corresponding to a five-dimensional d-vine. It consists of four trees T_j, $j = 1, \ldots, 4$. Tree T_j has $6 - j$ nodes and $5 - j$ edges. Each edge corresponds to a pair-copula density and the edge label corresponds to the subscript of the pair-copula density, e.g. edge 25| 34 corresponds to the
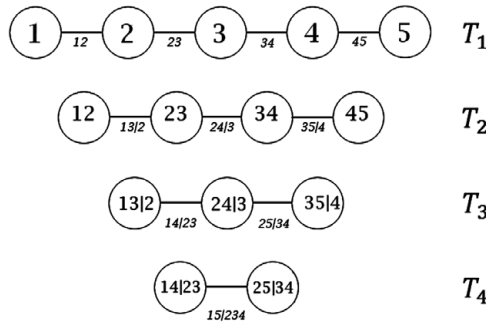
FIGURE 1: d-vine copula tree, taken from Aas *et al.* (2009), p. 184.

copula $c_{25|34}(\cdot)$. In a d-vine, no node in any tree T_j is connected to more than two edges. This graphical model is well suited for the case in which no particular variable is known to be a key variable governing all the interactions in the data set, as is the case of the idiosyncratic components of mortality rates.

Bedford and Cooke (2001) provide the density of an N-dimensional distribution in terms of a regular vine, which Aas *et al.* (2009) further specialize to a d-vine. The density becomes in this case

$$\prod_{k=1}^{N} f(\underline{e}_k) \prod_{j=1}^{N-1} \prod_{i=1}^{N-j} c_{i,i+j|i+1,\cdots,i+j-1}$$

$$\times \left\{ F\left(\underline{e}_i \middle| \underline{e}_{i+1}, \ldots, \underline{e}_{i+j-1}\right), F\left(\underline{e}_{i+j} \middle| \underline{e}_{i+1}, \ldots, \underline{e}_{i+j-1}\right) \right\}. \quad (2.14)$$

Here the index j identifies the trees, while i runs over the edges in each tree. Given the specific d-vine decomposition it is possible to approach the problem of estimation and posterior simulation by a maximum pseudo-likelihood approach.

Aas *et al.* (2009) provide the necessary steps to perform simulations using the estimated cascade of copulae in the d-vine construct. This method allows us to estimate percentiles of the multivariate density of the idiosyncratic components in the model, and therefore, to construct risk measures based on specific percentiles of the forecasted error.

2.6. Risk measures

Thanks to the copula simulation of different scenarios for the idiosyncratic terms and the point forecasts, we are able to choose a distorted risk measure from the family of measures that can estimate LR. Several alternatives are available within this family and include VaR an d TVaR. VaR at a level α is the α-quantile of a random variable x, that is, $VaR_\alpha(x) = \inf\{x| F_x(x) \geq \alpha\}$, where F_x is the distribution function of x and α is the tolerance level $\alpha \in (0, 1)$. A complementary measure, the Tail-VaR, or TVaR, corresponds to the mathematical

expectation beyond VaR, and it is defined as $TVaR_\alpha(x) = \frac{1}{1-\alpha}\int_\alpha^1 VaR_\lambda(x)\, d\lambda$. Additionally, GlueVaR measures, as proposed by Belles-Sampera *et al.* (2014), which are a combination of VaR and TVaR, may also be implemented.

We propose to estimate longevity (or conversely mortality) risk as follows:

$$LR = TVaR_\alpha(\hat{p}_{t+h}),\tag{2.15}$$

where α is associated with a suitable confidence level and $\hat{p}_{t+h}$ is the h-step ahead forecast of the death events. After projecting N mortality rates, we need to aggregate these projections in order to obtain the total number of "expected deaths" in a given portfolio. The number of expected deaths is a function of the exposed population by age and sex in the portfolio and the projection associated with each of the mortality rates. In this case, α is a very low percentile (i.e. 0.5%) if we want to measure LR, and a very high percentile (i.e. 99.5%) if we want to measure mortality risk. We provide both quantities in our empirical implementation.

Given that we are interested in constructing a methodology that might prove useful for any firm exposed to LR, we provide estimations of LR for different configurations of the exposed population, as explained in section 4.

3. RELATED LITERATURE

Factor models have been used extensively for forecasting mortality rates, after first being proposed for this purpose by Lee and Carter (1992). By the time these authors introduced the model to the actuarial and demographic literature, factor models were well established in the fields of psychology and econometrics. The Lee–Carter model is a single-factor model, where the factor is a stochastic trend shared by all the specific mortality rates. The model has been subject to much criticism (see, for example, Dushi *et al.*, 2010; Mitchell *et al.*, 2013), but it remains a plausible alternative within academia (see Bisetti and Favero, 2014, for a recent implementation of the model). Indeed, it has become something of a "workhorse" within the actuarial field, as has its extension to Poisson log-bilinear projections proposed by Brouhns and Denuit (2002) (See for example the works by Delwarde *et al.*, 2007; Lemoine, 2014, who add a switching component to the model).

The model is criticized primarily on the grounds that one single factor is deemed incapable of capturing all the common components underpinning the dynamics of the mortality rates. As a result, the model has been expanded to include more factors. These are generally additional stochastic trends, as in Yang *et al.* (2010), Alai and Sherris (2014), Jevtić et al. (2013) and Alonso (2008); but they might also be stationary factors subjacent to the differences (or the derivatives) of the log mortality rates, as opposed to the log-level mortality rates (Cossette *et al.*, 2007; Haberman and Renshaw, 2012, 2013; Mitchell *et al.*, 2013).

This last extension is particularly interesting from both a practical and theoretical perspective. In pragmatic terms, Mitchell *et al.* (2013) show that a stationary factor model can outperform a non-stationary model (containing possibly more factors). Thus, working with the log-mortality rate differences, as opposed to working with the rates in log-levels (as in the Lee–Carter model), may well increase the forecaster's ability. From a theoretical perspective this is important, especially since the econometrics literature shows that linear regressions between integrated series of order greater than zero (for example I(1) series) can easily derive in spurious estimations (Granger and Newbold, 1974). To avoid such spurious regressions, the series must be differentiated as many times as needed to ensure that all become stationary. After differentiation, traditional regressions can be performed. Indeed, this is the intuition behind Mitchell *et al.*'s (2013) main findings; namely, that using the series in differences appears to be a better empirical strategy than using the series in levels.

In other words, if log-mortality rates could be described by independent random walk trends and the Lee–Carter model imposes a common trend, this could lead to the imposition of a spurious cointegration relationship between all the log-mortality rates.

A potentially better alternative is available when the series at hand are indeed cointegrated and they all share the same stochastic trend or trends. In this case, performing a regression analysis in levels (for example, when forecasting mortality rates using their non-stationary first singular values) is fully justified and it might even be necessary so as to preserve the correct specification of the model. A more efficient alternative, in this case, would be to use an error correction model (ECM) (Engle and Granger, 1987) or its multivariate version, the vector error correction (VEC) model. In both cases, the estimation is super-consistent, as shown, for example, in Stock and Watson (1988). In other words, under cointegration, the differentiation of the series is not recommended because such a strategy might lead to biases in the estimated parameters and the forecasted quantities.

Thus, it is clear why pre-testing for unit roots and cointegration is a necessity of the first-order in order to avoid under- or over-differentiation of the series. Cointegration in multiple times series is generally based on tests constructed on the VAR representation of the system (Johansen, 1988; Stock and Watson, 1988). Indeed, this econometric machinery of unit roots pretesting procedures, cointegration tests, VAR and VEC models has recently been put to use in the actuarial literature, specifically to perform the task of estimating and forecasting mortality rates (Gaille and Sherris, 2011; Njenga and Sherris, 2011; Torri, 2011; D'Amato *et al.*, 2014). The aforementioned studies present evidence in favor of cointegration, using few variables. This finding seems to justify the use of factor models fitted to the levels of log-mortality rates (without differentiating the series). Indeed, the relationship between factor models and cointegration was explored more than two decades ago in Escribano and Peña (1994).

Unfortunately, the extension of this toolbox to the forecasting of mortality rates by age cohorts and sex is not straightforward. For example, when $N = 110$,

which is easily the case for mortality rates discriminated by age and sex, traditional cointegration tests based on the VAR representation of the system are not well suited, given the extremely large number of parameters to be estimated in the reduced-form VAR. For example, in a VAR comprising four lags and 110 mortality rates, $(101 \wedge 2) \, 4 = 48,400$ coefficients must be estimated, plus 110 variances and 5,995 covariances.

At this point, researchers are faced with a dichotomy: either they assume the cointegration of the mortality rates and estimate a factor model in levels, or they differentiate the series and fit a stationary factor model on the differences. We explore the empirical consequences of these two alternatives and perform several comparisons, particularly in terms of the mean-squared forecasting error, for different horizons and different modeling strategies.

3.1. Models M1–M8

In this section, we undertake a theoretical comparison of the GDFM and DFM using the mortality models studied by Cairns *et al.* (2009) and we expand their typology to include a number of recently proposed models. We also switch our notation in this section to mimic that employed in Cairns *et al.* (2009) with the aim of providing greater clarification to the experienced reader.

In broad terms, the main models used in the actuarial sciences to forecast mortality rates can be summed up in Table 1. As is evident, the models proposed here (M9, M11 and M12) belong to the family of log-mortality rate models, and they differ from the logit-models in that the latter use logit transformations of the mortality rates in the estimation. However, in line with Mitchell *et al.* (2013), the logit-models do not seem to perform any better or any worse than the log models, so we concentrate our attention here on factor models of the log-rates or the differentiated log-rates.

Notice that M1 is a special case of M9, when $L = 0$ and $r = 1$. Likewise, M10 is a special case of M11, when $L = 0$ and $r = 1$, so M1 and M10 can be regarded as SFMs. The GDFM fitted over the differences of log-mortality rates (M12) differs from all the other models in that it does not use PC to estimate the factors, but rather it uses GPC.

In our empirical section, we do not consider any model with cohort effects, but Mitchell *et al.* (2013) provide comparisons of their model with M2 and it seems to perform better in most cases. The extension of M11 or M12 to incorporate cohort effects is straightforward and could be explored in future research, in keeping with the proposals in Haberman and Renshaw (2012, 2013).

Here, we have opted instead to model contemporaneous dependence through copula functions. The copula approach to the estimation of longevity or mortality risks is novel (see, for example, Lin *et al.*, 2015, for a related implementation) and is concerned with the dependence structure of the idiosyncratic components of the model (which we obviated in Table 1). This modeling strategy only makes sense within the context of an approximate factor

TABLE 1

MORTALITY MODELS.

Model	Formula
M1: Lee and Carter (1992)	$\log m(t, x) = \beta_x^{(1)} + \beta_x^{(2)} k_t$
M2: Renshaw and Haberman (2006)	$\log m(t, x) = \beta_x^{(1)} + \beta_x^{(2)} k_t + \beta^{(3)} \gamma_{t-x}^{(3)}$
M3: Currie (2006)	$\log m(t, x) = \beta_x^{(1)} + k_t + \gamma_{t-x}^{(3)}$
M4: Currie et al. (2004)	$\sum_{i,j} \theta_{ij} B_{ij}^{ay}(x, t)$
M5: Cairns et al. (2006)	$\text{logit } q(t, x) = k_t^{(1)} + k_t^{(2)}(x - \bar{x})$
M6: Cairns et al. (2009)	$\text{logit } q(t, x) = k_t^{(1)} + k_t^{(2)}(x - \bar{x}) + \gamma_{t-x}^{(3)}$
M7: Cairns et al. (2009)	$\text{logit } q(t, x) = k_t^{(1)} + k_t^{(2)}(x - \bar{x})$ $+ k_t^{(3)}\left((x - \bar{x})^2 - \hat{\sigma}_x^2\right) + \gamma_{t-x}^{(4)}$
M8: Cairns et al. (2009)	$\text{logit } q(t, x) = k_t^{(1)} + k_t^{(2)}(x - \bar{x}) + \gamma_{t-x}^{(3)}(x_c - x)$
M9: Dynamic Factor Model in Levels, DFM	$\log m(t, x) = \beta_x^{(1)} + \sum_{i=2}^{r+1} \beta_x^{(i)}(L) k_t^{(i-1)}$
M10: Mitchell et al. (2013)	$\Delta \log m(t, x) = \alpha_x^{(1)} + \beta_x^{(2)} \Delta k_t$
M11: Dynamic Factor Model in Differences, DDFM	$\Delta \log m(t, x) = \alpha_x^{(1)} + \sum_{i=2}^{r+1} \beta_x^{(i)}(L) \Delta k_t^{(i-1)}$
M12: Generalized Dynamic Factor Model in Differences, DGDFM	$\Delta \log m(t, x) = \alpha_x^{(1)} + \sum_{i=2}^{r+1} \beta_x^{(i)}(L) \Delta q_t^{(i-1)}$

NOTE: Mortality Models: from M1 to M8 are taken from Cairns et al. (2009). Following these authors, $\beta_x^{(i)}$, $k_t^{(i)}$, and $\gamma_{t-x}^{(i)}$ are age, period and cohort effects respectively. The $B_{ij}^{ay}(x, t)$ are B-spline basis functions and the θ_{ij} are weights attached to each basis function. $\bar{x}$ is the mean age over the range of ages being used in the analysis. $\hat{\sigma}_x^2$ is the mean value of $(x - \bar{x})^2$. L is the lag-operator and $q_t^{(i)}$ are generalized principal components.

model, such as M11 and M12. Otherwise, identification issues might arise in the estimation process given that a SFM, such as M1 or M10, generally uses the orthogonality condition between the idiosyncratic models to identify the factors.

    Our empirical illustration for the UK compares M1, M9, M11 and M12. And M10 is included as a special case of M11.

4. DATA AND PRELIMINARY ANALYSIS

Our data consist of annual mortality rates for males and females in the UK from 1950 to 2011. The data for 0–101 years were taken as they appear on the webpage of the Human Mortality Database. The data for 102–109+ years were extrapolated to produce a mortality rate equal to 1 at 110 years. By so doing, we are able to prevent implausible variability being registered at older ages. We pretest for unit roots in the log-mortality rates and for the differences of the

log-mortality rates. The series show evidence of unit root behavior in the great majority of cases in logs, and they seem stationary when we differentiate them.

In order to determine the number of static and dynamic factors, (i.e. the number of PC and their lags), we follow two statistics proposed by Bai and Ng (2002). Specifically, to determine the number of static factors we use the IC information criterion given by

$$IC(k) = \ln(S(k)) + kg(N, T), \qquad (4.1)$$

where $S(k) = (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \hat{\lambda}_i^{k'} \hat{F}_t^k)^2$ is the mean-squared error divided by NT. k. is the number of static factors and $g(N, T)$ is a penalty function, such that $g(N, T) = \frac{N+T}{NT} ln(\frac{NT}{N+T})$. The number of static factors, $\hat{k}_{IC}$ is such that

$$\hat{k}_{IC} = \mathop{\mathrm{argmin}}_{0 \leq k \leq k_{\max}} IC(k), \qquad (4.2)$$

where $k_{\max}$ is then the maximum possible number of factors, $k_{\max} = 15$ in our case. Once the number of static factors is selected, the number of dynamic factors is determined using the methodology proposed by Bai and Ng (2007). Broadly speaking, this methodology deals with the rank of the vector space spanned by the original dynamic factors, which is expressed in the vectors containing the static factors. In general $q \leq r$, where q are the original, primitive, dynamic factors, and r the number of static factors spanned by q.

We estimate the number of static and dynamic factors following (4.1) and (4.2) for different portfolio populations. We consider six cases: An exposed population of 0–109+ years of males and females; another product designed for people aged between 18 and 64 years (male and females) and an exposed population ranging from 65 to 109+ years. The population was set at 30,000 people (either males or females), which is approximately 0.1% of the total population of the UK for the year 2011. The participation of each age cohort in the total was set according to its participation in the population of the UK for 2011.

We also present the variability explained by the first 15 PC for the series in differences in Table 2. The percentage explained by the PC selected for modeling is higher in the case of males, ranging from 72.63% to 93.77% and lower for females between 18 and 64 years, for which the first principal component only explains 11.46% of the total variation. In this case, the copula function plays a central role in determining the risk profile of the portfolio.

Notice that in all the cases, we identify one dynamic factor but, by contrast, several static factors (as many as seven in the case of males aged between 65 and 109+ years old). This indicates the presence of high dynamics in the system, which traditional SFMs ignore. Thus, the assumption implicit in the literature, initiated by Lee and Carter, that one single "mortality course" explains the dynamic of the system as a whole seems plausible, but it seems convenient to incorporate dynamics in modeling and forecasting by including "lags" of that

TABLE 2

PERCENTAGE OF THE VARIABILITY EXPLAINED BY THE FIRST 10 PRINCIPAL COMPONENTS IN THE UK'S
MORTALITY RATES AND OPTIMAL NUMBER OF FACTORS.

	r	1	2	3	4	5
	0–109+	63.99%	69.34%	**72.63%**	75.03%	77.06%
	18–64	63.78%	69.75%	**73.51%**	76.34%	79.07%
	65–109+	75.91%	86.31%	88.92%	90.45%	91.77%
Males	**6**	**7**	**8**	**9**	**10**
	0–109+	79.01%	80.84%	82.38%	83.73%	84.98%
	18–64	81.45%	83.48%	85.25%	86.97%	88.49%
	65–109+	92.84%	**93.77%**	94.43%	95.00%	95.52%
	r	**1**	**2**	**3**	**4**	**5**
	0–109+	23.50%	**30.88%**	35.97%	40.47%	44.71%
	18–64	**11.47%**	19.83%	26.94%	33.54%	39.40%
	65–109+	48.79%	**64.62%**	67.91%	71.09%	73.97%
Females	**6**	**7**	**8**	**9**	**10**
	0–109+	48.72%	52.38%	55.60%	58.69%	61.56%
	18–64	44.77%	49.50%	53.89%	58.02%	61.92%
	65–109+	76.32%	78.45%	80.51%	82.28%	84.00%

NOTE: series of the differences of log-mortality rates. Data for the UK from 1950 to 2011, taken from the Human Mortality Database. They correspond to males and females in different age populations. The highlighted numbers correspond to the number of factors, r, identified following the optimality criterion proposed by Bai and Ng (2002).

single factor. In all circumstances, this strategy appears to be more appropriate in the case of males, as we only identify up to two static factors for the female populations.

5. RESULTS AND DISCUSSION

Our empirical section uses data for the UK. It compares M1, M9, M11 and M12 from Table 1 (M10 is included as a special case of M11). All factor models presented here were estimated using MATLAB. In the estimations, we used some routines available on the webpage of Serena Ng (http://www.columbia.edu/ sn2294/) to estimate the DFM, and to select the optimal number of static and dynamic factors. To estimate the GDFM, for both one- and two-sided filters, we used codes available on webpage of Mario Forni. (http://morgana.unimore.it/forni_mario/matlab.htm). To estimate the copula-functions we used R, specifically employing the package CDVine.

5.1. Forecasting

We define the h-step ahead forecast for mortality rate i, and its associated mean-squared forecasting error as follows:

$$\hat{x}_{iT+h|T} = x_{iT} + \Delta\hat{x}_{iT+1|T} + \ldots + \Delta\hat{x}_{iT+h|T}, \qquad (5.1)$$

$$MSFE_i^h = \frac{1}{T_1 - T_0 - h + 1} \sum_{T=T_0}^{T_1-h} \left(\hat{x}_{iT+h|T} - x_{T+h}\right), \qquad (5.2)$$

where T_0 is the last year within the sample (in our case 2001), T_1 is the last year outside the sample (in our case 2011). Thus, we use years from 1950 to 2001 to estimate the models and the last ten years to measure their relative performance. The mean-squared forecasting error (MSFE) for several forecasting horizons is presented for males and females in Figure 2.

Given that we are primarily concerned with risk measures calculated in short periods of time (for example one year ahead, when a firm has to set a new capital buffer), we confined our forecasting analysis from 1 to 10 years. The MSFE, presented for each model, is a weighted average of the individual forecasting errors, for the individual changes in log-motility rates, the weights being the population in each category for the cases considered in the exercise.

In general, the GDFM performs better than the other models in forecasting, especially for medium- and long-term horizons. The worst model is, in general, the Lee–Carter model, followed by the dynamic model in log-rates. We confirmed the finding reported by Mitchell *et al.* (2013) that the models "in differences" outperform the models in "levels". Our intuition for this finding is that the log-rate models impose a cointegration relationship between the mortality rates that can be false for the 110 rates as a whole. This is the case, even if some subsets of the variables are effectively cointegrated; i.e., while they shared the same common trend, some others certainly did not and, therefore, imposing a common trend can result in spurious estimations of the factor loads and forecasting projections. This in turn could deteriorate the forecasting estimation, especially for medium- and long-term horizons.

Additionally and as expected, we find that the adjustment, through the spectral density matrices applied in the GDFM, improves the forecast over the standard DFM, but that the improvement is small in magnitude.

The above findings are confirmed when we use the AIC criterion to compare the models. In Table A1 of the Appendix, we present the AIC criterion for different portfolio populations and for models M1 (LC), M9 (DFM), M11 (DDFM) and M12 (DGDFM). Once again, the performance of the models in differences is clearly superior to that of the models in levels. The worst performance in each case is given by the Lee–Carter model, followed by the DFM in levels. Nevertheless, the two models in differences provide remarkably similar performances. Hence, the previously reported gains in terms of out-of-sample forecasting of the generalized version of the model could, to a certain extent,
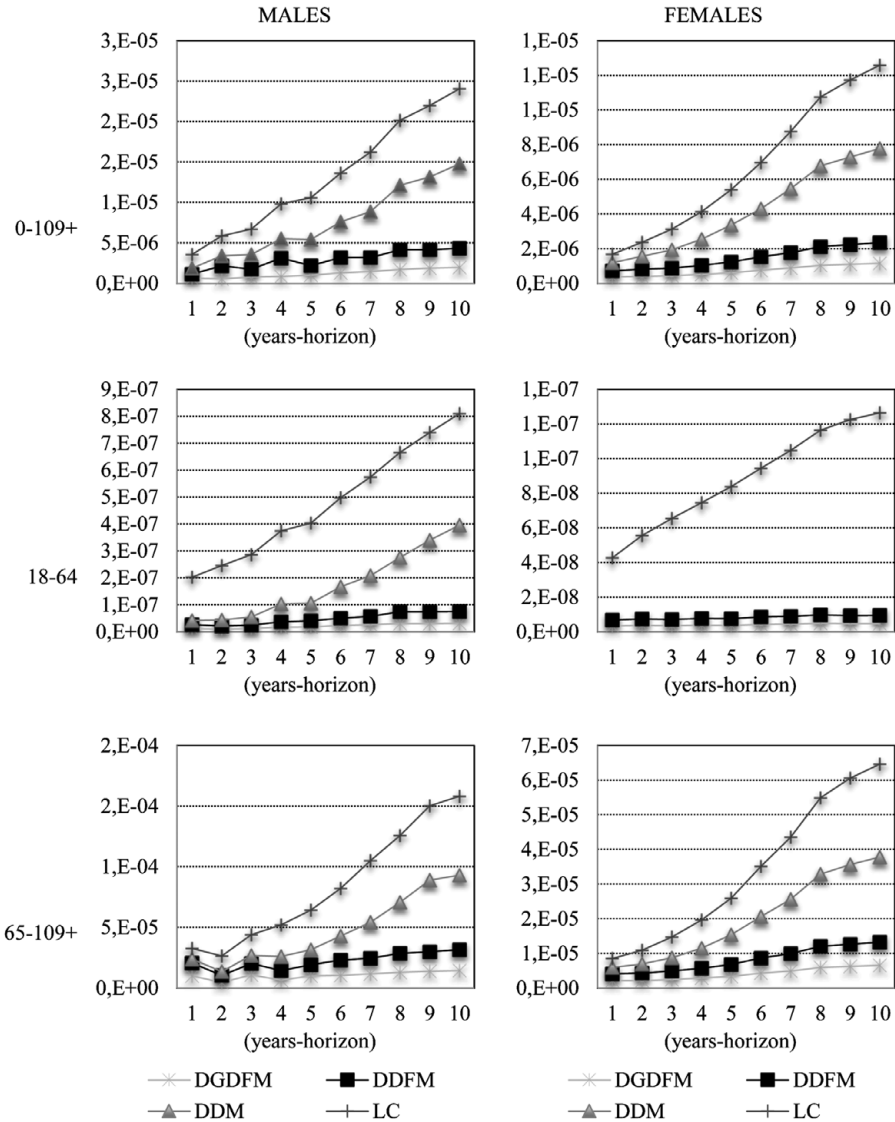
FIGURE 2: Mean-squared forecasting error (MSFE) by forecasting horizon. The models were estimated using the UK mortality data from 0 to 109+ years. The estimation period runs from 1950 to 2001, and the forecasting period comprises 10 years, from 2002 to 2011. The participation in populations of 30,000 individuals was set according to the UK population discriminated by ages in the year 2011.

be due to over-fitting issues related to model M11. The strategy performed by the GDFM, which uses an interactive weighting scheme, seems particularly appropriate for forecasting periods of more than two years, and to some degree prevents over-fitting.

We also analyzed information about the models' forecasting performance by age ranges. We find that all the models perform better for ages between 0 and 95 than for ages above 95. This poor performance is natural, given that higher ages are associated with greater year-on-year variability. Interestingly, the models in log levels seem to perform better than the models in differences at these ages. This could be related to the information loss in the process of differentiation and issues concerning the over-fitting of the models in differences at higher ages, which in turn makes them particularly weak when dealing with outliers or considerable variability. Thus, for the last ranges we conclude that the loss of information generated from the differentiation of the series, and because of possible over-fitting, is greater than the gain derived from avoiding the imposition of a cointegration relationship, at least for short-term horizons. However, this "outperformance" disappears as the forecast is forwarded in time. For 10 years ahead, the models in differences outperform the models in log-levels again, even at ages above 95. This last result can be explained in line with section 2, because the models in differences do not impose an implausible cointegration relation among all the series in the system.

5.2. Longevity and mortality risks

We calculate TVaR and VaR measures one year ahead using copulae. The procedure is as follows:

- In order to estimate the copula functions, we first transformed the original data (i.e. the idiosyncratic components of the GDFM) using the empirical distribution function. This step provides us with a pseudo-sample as described in (2.11), which must be distributed as a standard uniform random variable, if the empirical distribution is a good approximation of the marginal density. In order to check the accuracy of the procedure, we simulated a uniform random variable in the interval [0,1] and then we compared the simulated variable with each series in the pseudo-sample, by means of a Kolmogorov–Smirnov statistic. In each case, the null hypothesis is not rejected and, therefore, the empirical distribution provides a good approximation of the marginal, as is required for this empirical exercise.
- The selection of each conditional copula in the estimation is an empirical concern. Therefore, we considered different alternatives, which summarized different possible dependence structures in the data in a flexible way. We specifically compared the following copulae: Gaussian, Clayton, BB6, Survival-Joe, and Rotated-Clayton (180 degrees and 270 degrees). We tested more than 40 copulae in a preliminary exercise and these six copulae invariably performed much better, so we concentrated our search for the best copula on these six functions. We selected the best copula among the candidates, using the AIC criterion, and used this to construct the "multivariate dependence tree" shown in Figure 1.

TABLE 3

LONGEVITY AND MORTALITY RISKS FOR THREE PORTFOLIOS OF
SIZE 30,000 (ONE-YEAR AHEAD FORECAST).

	Level	TVaR 0.5	VaR 0.5	Expected Deaths	VaR 99.5	TVaR 99.5
Males	0–109+	238	239	252	261	262
	18–64	78	78	80	82	82
	65–109+	1,283	1,288	1,346	1,394	1,399
Females	0–109+	254	255	259	263	263
	18–64	52	52	54	56	56
	65–109+	1,242	1,244	1,257	1,269	1,270

NOTE: The statistics were calculated using data for the UK from 1950 to 2011, as taken from the Human Mortality Database. The forecasting horizon was set at one year. The second column shows the different population portfolios considered in the exercise. The hypothetical populations mimic the age composition of the UK in 2011.

- Before calculating the risk measures, we constructed the AIC criteria for the models fitted to each population portfolio. We did so in order to compare the fitted model with that of the independence assumption among the idiosyncratic components of the data. In each case, using the vine-copula model, we obtained a reduction in the AIC criterion compared to the independence assumption. For instance, for a portfolio of males aged 0–109+ years, the reduction in the AIC was −4013.72; for those aged 18–64 years the reduction was −808.23; and, for those aged 65–109+ years it was −1864.65. In the case of females, the reduction was −3184.57, −732.55 and −1138.18, respectively. We found very similar results when using the in-sample residuals between 1992 and 2001 and between 2002 and 2011, instead of the AIC criteria. These are not shown for reasons of space.
- Finally, we report the TVaR and VaR at 0.5% (left tail) and the same statistics for the 99.5 percentiles (right tail) in Table 3. This level of confidence is standard practice in the insurance market. Recall, 0.5% of VaR or TVaR is concerned with LR, while the right tail illustrates mortality risk. In both cases, risk should be understood as a significant dispersion from the expected number of deaths, forecasted with the GDFM.
- Confidence intervals for the measures proposed in Table 3 can be constructed using bootstrapping techniques. Given that these statistics are estimated using the simulated observations at very high and very low percentiles (99.5 and 0.5) of the 10,000 scenarios, it is possible to simulate these 10,000 scenarios 100 times and to calculate the statistics in each case.

The results of the forecasting exercise in Table 3 can be read as the expected number of events (deaths) in one year. In the case of the 30,000 males aged between 0 and 109+ years (in a portfolio that mimics the population structure of UK males in 2011), we can expect to see 252 deaths, while we would expect 259

events in a portfolio composed solely by woman (again preserving the female population structure in the UK for these ages).

In contrast, in a population of 30,000 males aged between 18 and 64 years, we can expect 80 deaths in one year, and 54 in the case of females of the same ages. Finally, and more importantly, if we were to construct a population portfolio comprising males aged between 65 and 109+ years, we would expect to observe 1,346 deaths in one year; while for women between these same ages we would expect to observe 1,257 events.

In terms of LR, as expected, we found that this risk increases with advancing age. Logically, not only is it the oldest part of the population that presents the highest mortality rates, it also presents greater variability and, therefore, forecasting at ages above 65 years is subject to considerably greater uncertainty, which can be understood as greater risk. For example, the VaR (99.5%) for males between 0 and 109+ years is 261 and the VaR (0.5%) is 239. For woman these values are 263 and 255, respectively. These estimations provide useful insights for any insurance or pension company. Specifically, they tell us that it is possible to assert with a 99.5% level of statistical confidence that no more than 261 persons will die in one year, or conversely that no fewer than 239 will die, even when the expected number of deaths is 252.

The calculations increase significantly for the upper ages, for males and females, but in a larger proportion for the former. In a portfolio of males aged between 65 and 109+ years (the portfolio with the greatest longevity (and mortality) risk), we can expect with a 99.5% level of confidence to observe 1,346–1,288(58 persons surviving in one year above the projections provided by the best available model (the GDFM) or 1,394–1,346(48 people dying above this projection. Alternatively, if we use the TVaR to make the same calculations, we find that 63 people will survive above our expectations and 53 will die above our expectations.

These results highlight an empirically important finding of our exercise: the asymmetric nature of mortality and LRs. By using the VaR at 99.5%, the LR is 20.8% higher than the mortality risk for the portfolio of males (8% in the case of females). Using the TVaR these numbers are 15.8 and 15.3%, respectively. In both cases the LR is higher. Moreover, the LR of the older population (let's say the portfolio comprising males aged between 65 and 109+ years) is considerably higher than the mortality risk of a younger population aged between 18 and 64 years. In fact, in order to make these risks equivalent, you would have to affiliate 29 times more people aged between 18 and 64 than those aged between 65 and 109+, in the expectation of some kind of "cancellation" between the longevity and mortality risks of the different population portfolios.

Finally, we compare the LR of males and females of the same ages (but with different population structures). We observe that the risk is higher for males than it is for females, which does not contradict the well-documented fact that woman tend to live longer than men. Indeed, it is precisely because of greater female longevity that the variability of mortality rates in males is greater in the upper ages and, therefore, more difficult to forecast.

6. CONCLUSIONS

The incorporation of dynamics by means of adding lags (new factors) when forecasting mortality rates generates a better fit of the models to the data, especially when modeling male populations; however, it is the differentiation of the series that increases the forecasting capability of the factor models. We provide some intuition for these findings and show that the gains in terms of forecasting increase as the forecasting horizon expands. The intuition for this finding is based on the fact that traditional factor models of the log-mortality rates, such as the Lee–Carter model, impose, by assumption, a cointegration relationship between all the series in the system that is very unlikely to be observed in the data, due to the great number of series in an exercise of this kind (more than 200 for males and females in our example).

Further improvements can be achieved by using (one-sided) GPC in the estimation stage. Generalized factor models allow for a richer dynamics in the data and use the information contained in the spectral density matrix to improve the model fit.

It should be noted, however, that models in differences perform worse than models in levels for short forecasting horizons at ages above 95 years. This finding is related to the fact that the better fit of the models in differences is achieved at the expense of some degree of over-fitting for the older population, especially for males. This over-performance of the models in levels disappears as the forecast horizon expands and, indeed, for 10 years ahead it is fully reversed. The intuition is that the imposition of the shared stochastic trend deteriorates forecasting capabilities in a cumulative fashion, and that this is worse than the over-fitting disadvantage suffered by the models in differences for ages above 95 in medium- and long-term forecasting horizons.

Finally, we found that LR is higher among older portfolios and markedly higher among men than among women. There is also an asymmetric relation between the longevity and mortality risks, which hinders any attempt at compensating one of the risks in one population with the other risk in a different exposed population.

REFERENCES

AAS, K., CZADO, C., FRIGESSI, A. and BAKKEN, H. (2009) Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, **44**(2), 182–198.

ALAI, D.H. and SHERRIS, M. (2014) Rethinking age-period-cohort mortality trend models. *Scandinavian Actuarial Journal*, 2014 (3), 208–227.

ALONSO, A.M. (2008) *Predicción de Tablas de Mortalidad Dinámicas Mediante un Procedimiento Bootstrap.* Madrid: Fundación MAPFRE.

BAI, J. and NG, S. (2002) Determining the number of factors in approximate factor models. *Econometrica*, **70**(1), 191–221.

BAI, J. and NG, S. (2004) A panic attack on unit roots and cointegration. *Econometrica*, **72**(4), 1127–1177.

BAI, J. and NG, S. (2007) Determining the number of primitive shocks in factor models. *Journal of Business and Economic Statistics*, **25**(1), 52–60.

BAI, J. and NG, S. (2008) Large dimensional factor analysis. *Foundations and Trends in Econometrics*, **3**(2), 89–163.

BAI, J. and WANG, P. (2012) *Identification and estimation of dynamic factor models.* MPRA Paper, 38434. Available at: https://mpra.ub.uni-muenchen.de/38434/2/MPRA_paper_38434.pdf

BATES, B.J., PLAGBORG-MØLLER, M., STOCK, J.H. and WATSON, M.W. (2013) Consistent factor estimation in dynamic factor models with structural instability. *Journal of Econometrics*, **177**(2), 289–304.

BEDFORD, T. and COOKE, R.M. (2001) Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, **32**(1–4), 245–268.

BEDFORD, T. and COOKE, R.M. (2002) Vines - A new graphical model for dependent random variables. *Annals of Statistics*, **30**(4), 1031–1068.

BELLES-SAMPERA, J., GUILLÉN, M. and SANTOLINO, M. (2014) Beyond value-at-risk: GlueVaR distortion risk measures. *Risk Analysis*, **34**(1), 121–134.

BISETTI, E. and FAVERO, C.A. (2014) Measuring the impact of longevity risk on pension systems: The case of Italy. *North American Actuarial Journal*, **18**(1), 87–103.

BLAKE, D., BOARDMAN, T. and CAIRNS, A. (2014) Sharing longevity risk: Why governments should issue longevity bonds. *North American Actuarial Journal*, **18**(1), 258–277.

BROUHNS, N. and DENUIT, M. (2002) Risque de longévité et rentes viagères II. Tables de mortalité prospectives pour la population belge. *Belgian Actuarial Bulletin*, **2**(1), 49–63.

CAIRNS, A.J. (2011) Modelling and management of longevity risk: Approximations to survivor functions and dynamic hedging. *Insurance: Mathematics and Economics*, **49**(3), 438–453.

CAIRNS, A.J.G., BLAKE, D. and DOWD, K. (2006) A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *The Journal of Risk and Insurance*, **73**(4), 687–718.

CAIRNS, A., BLAKE, D., DOWD, K., COUGHLAN, 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.

CAIRNS, A., BLAKE, D., DOWD, K., COUGHLAN, G.D., EPSTEIN, D. and KHALAF-ALLAH, M. (2011) Mortality density forecasts: An analysis of six stochastic mortality models. *Insurance: Mathematics and Economics*, **48**(3), 355–367.

CONTINUOUS MORTALITY INVESTIGATION. (2004) *Projecting future mortality: A discussion paper*. Working paper, 3. Available at: http://www.actuaries.org.uk/research-and-resources/documents/cmi-working-paper-3-projecting-future-mortality-discussion-paper

CONTINUOUS MORTALITY INVESTIGATION. (2005) *Projecting future mortality: Towards a proposal for a stochastic methodology*. Working paper, 15. Available at: http://www.actuaries.org.uk/research-and-resources/documents/cmi-working-paper-15-projecting-future-mortality-towards-proposal-s

CONTINUOUS MORTALITY INVESTIGATION. (2013) *The CMI mortality projections model*, CMI_2012. Working paper, 63. Available at: http://www.actuaries.org.uk/research-and-resources/pages/cmi-working-paper-63

COSSETTE, H., DELWARDE, A., DENUIT, M., GUILLOT, F. and MARCEAU, É. (2007) Pension plan valuation and mortality projection: A case study with mortality data. *North American Actuarial Journal*, **11**(2), 1–34.

CURRIE, I. (2006) *Smoothing and Forecasting Mortality Rates with P-splines.* Presented at the Institute of Actuaries, London. Available at: http://www.macs.hw.ac.uk/~iain/research/talks/Mortality.pdf

CURRIE, I.D., DURBAN, M. and EILERS, P.H. (2004) Smoothing and forecasting mortality rates. *Statistical Modelling*, **4**(4), 279–298.

D'AMATO, V., HABERMAN, S., PISCOPO, G. and RUSSOLILLO, M. (2012) Modelling dependent data for longevity projections. *Insurance: Mathematics and Economics*, **51**(3), 694–701.

D'AMATO, V., HABERMAN, S., PISCOPO, G., RUSSOLILLO, M. and TRAPANI, L. (2014) Detecting common longevity trends by a multiple population approach. *North American Actuarial Journal*, **18**(1), 139–149.

DAHL, M., MELCHIOR, M. and MØLLER, T. (2008) On systematic mortality risk and risk-minimization with survivor swaps. *Scandinavian Actuarial Journal*, 2008 (2–3), 114–146.

DELWARDE, A., DENUIT, M., GUILLÉN, M. and VIDIELLA-I-ANGUERA, A. (2007) Application of the poisson log-bilinear projection model to the G5 mortality experience. *Belgian Actuarial Bulletin*, **6**(1), 54–68.

DENTON, F.T., FEAVER, C.H. and SPENCER, B.G. (2005) Time series analysis and stochastic forecasting: An econometric study of mortality and life expectancy. *Journal of Population Economics*, **18**(2), 203–227.

DUSHI, I., FRIEDBERG, L. and WEBB, T. (2010) The impact of aggregate mortality risk on defined benefit pension plans. *Journal of Pension Economics and Finance*, **9**(4), 481–503.

ENGLE, R.F. and GRANGER, C.W.J. (1987) Co-Integration and error correction: Representation, estimation and testing. *Econometrica*, **55**(2), 251–276.

ESCRIBANO, Á. and PEÑA, D. (1994) Cointegration and common factors. *Journal of Time Series Analysis*, **15**(6), 577–586.

FORNI, M. and REICHLIN, L. (1998) Let's get real: Factor to analytical cycle approach dynamics disaggregated business. *The Review of Economic Studies*, **65**(3), 453–473.

FORNI, M., HALLIN, M., LIPPI, M. and REICHLIN, L. (2000) The generalized dynamic-factor model: Identification and estimation. *The Review of Economics and Statistics*, **82**(4), 540–554.

FORNI, M., HALLIN, M., LIPPI, M. and REICHLIN, L. (2004) The generalized dynamic factor model consistency and rates. *Journal of Econometrics*, **119**(2), 231–255.

FORNI, M., HALLIN, M., LIPPI, M. and REICHLIN, L. (2005) The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association*, **100**(471), 830–840.

GAILLE, S. and SHERRIS, M. (2011) Modelling mortality with common stochastic long-run trends. *The Geneva Papers on Risk and Insurance-Issues and Practice*, **36**(4), 595–621.

GRANGER, C.W.J. and NEWBOLD, P. (1974) Spurious regressions in econometrics. *Journal of Econometrics*, **2**(2), 111–120.

HABERMAN, S. and RENSHAW, A. (2012) Parametric mortality improvement rate modelling and projecting. *Insurance: Mathematics and Economics*, **50**(3), 309–333.

HABERMAN, S. and RENSHAW, A. (2013) Modelling and projecting mortality improvement rates using a cohort perspective. *Insurance: Mathematics and Economics*, **53**(1), 150–168.

HAMILTON, J.D. (1994) *Time Series Analysis*. Princeton: Princeton University Press.

HANEWALD, K., POST, T. and GRÜNDL, H. (2011) Stochastic mortality, macroeconomic risks and life insurer solvency. *The Geneva Papers on Risk and Insurance-Issues and Practice*, **36**(3), 458–475.

HÁRI, N., DE WAEGENAERE, A., MELENBERG, B. and NIJMAN, T.E. (2008) Longevity risk in portfolios of pension annuities. *Insurance: Mathematics and Economics*, **42**(2), 505–519.

HARVEY, A.C. (1990) *Forecasting, Structural Time Series Models and the Kalman Filter.* New York: Cambridge University Press.

HOLLMANN, F.W., MULDER, T.J. and KALLAN, J.E. (2000) Methodology and assumptions for the population projections of the United States: 1999 to 2100. *Population Division U.S. Census Bureau*. Working paper, 38. Available at: http://census.gov/population/www/documentation/twps0038/twps0038.html

HOLMES, E.E., WARD, E.J. and SCHEUERELL, M.D. (2014) *Analysis of Multivariate Time- Series using the MARSS Package*. Seattle: Northwest Fisheries Science Center, NOAA.

JEVTIĆ, P., LUCIANO, E. and VIGNA, E. (2013) Mortality surface by means of continuous time cohort models. *Insurance: Mathematics and Economics*, **53**(1), 122–133.

JOE, H. (1996) Families of m-variate distributions with given margins and m(m−1)/2 bivariate dependence parameters. *Lecture Notes-Monograph Series*, **28**, 120–141.

JOHANSEN, S. (1988) Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, **12**(2–3), 231–254.

KUROWICKA, D. and COOKE, R.M. (2005) Distribution-free continuous bayesian belief nets. In Wilson A., Keller-McNulty S., Armijo Y., Limnios N., *Modern Statistical and Mathematical Methods in Reliability*, pp. 309–323. World Scientific Publishing Company, Singapore.

LEE, R.D. and CARTER, L.R. (1992) Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, **87**(419), 659–671.

LEMOINE, K. (2014) Mortality regimes and longevity risk in a life annuity portfolio. *Scandinavian Actuarial Journal* (Published Online), 1–36. Available at: http://www.tandfonline/com/doi/abs/10.1080/03461238.2014.882860?journalCode=sact20

LIN, T., WANG, C.W. and TSAI, C.C.H. (2015) Age specific copula-AR-GARCH mortality models. *Insurance: Mathematics and Economics*, **61** (March 2015), 110–124.

LORSON, J. and WAGNER, J. (2014) The pricing of hedging longevity risk with the help of annuity securitization: An application to the German market. *The Journal of Risk Finance: The Convergence of Financial Products and Insurance*, **15**(4), 385–416.

MACMINN, R., BROCKETT, P. and BLAKE, D. (2006) Longevity risk and capital markets. *The Journal of Risk and Insurance*, **73**(4), 551–557.

MCNEIL, A.J., FREY, R. and EMBRECHTS, P. (2005) *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton: Princeton University Press.

MITCHELL, D., BROCKETT, P., MENDOZA-ARRIAGA, R. and MUTHURAMAN, K. (2013) Modeling and forecasting mortality rates. *Insurance: Mathematics and Economics*, **52**(2), 275–285.

NGAI, A. and SHERRIS, M. (2011) Longevity risk management for life and variable annuities: Effectiveness of static hedging using longevity bonds and derivatives. *Insurance: Mathematics and Economics*, **49**(1), 100–114.

NJENGA, C.N. and SHERRIS, M. (2011) Longevity risk and the econometric analysis of mortality trends and volatility. *Asia-Pacific Journal of Risk and Insurance*, **5**(2), 1–54.

OLIVIERI, A. (2011) Stochastic mortality: Experience-based modeling and application issues consistent with Solvency 2. *European Actuarial Journal*, **1**(1), 101–125.

PEÑA, D. and PONCELA, P. (2006) Nonstationary dynamic factor analysis. *Journal of Statistical Planning and Inference*, **136**(4), 1237–1257.

PLAT, R. (2011) One-year value-at-risk for longevity and mortality. *Insurance: Mathematics and Economics* **49**(3), 462–470.

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), 556–570.

RICHARDS, S. and JONES, G. (2004) *Financial Aspects of Longevity Risk*. Working paper presented in the Staple Inn Actuarial Society. Available at: http://sias.org.uk/resources/papers/f4c264d7c93007ff32ea515801f77973.pdf

SKLAR, A. (1959) Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris* VIII, 229–231.

STALLARD, E. (2006) Demographic issues in longevity risk analysis. *The Journal of Risk and Insurance*, **73**(4), 575–609.

STOCK, J.H. and WATSON, M.W. (1988) Testing for common trends. *Journal of the American Statistical Association*, **83**(404), 1097–1107.

STOCK, J.H. and WATSON, M.W. (2002) Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, **97**(460), 1167–1179.

STOCK, J.H. and WATSON, M.W. (2006) Forecasting with many predictors. In Elliot G., Granger C.W.J., Timmermann A., *Handbook of Economic Forecasting*, vol. 1, pp. 515–554. Elsevier, Tampa (FL).

STOCK, J.H. and WATSON, M.W. (2010) Dynamic factor models. In Clements M.P., Henry D.F., *Oxford Handbook of Economic Forecasting*. pp. 35–59. Oxford University Press, Oxford.

TORRI, T. (2011) Building blocks for a mortality index: An international context. *European Actuarial Journal*, **1**(1), 127–141.

WILLS, S. and SHERRIS, M. (2008) Integrating financial and demographic longevity risk models: An Australian model for financial applications. *Australian School of Business Research.* Working paper, 2008ACTL05. Available at: http://www/actuaries.asn.au/Library/Events/FSF/2008/FSF08_6d_Paper_Wills_Mortality.pdf

WONG, T.W., CHIU, M.C. and WONG, H.Y. (2014) Time-consistent mean-variance hedging of longevity risk: Effect of cointegration. *Insurance: Mathematics and Economics*, **56** (May 2014), 56–67.

YANG, S.S., YUE, J.C. and HUANG, H.-C. (2010) Modeling longevity risks using a principal component approach: A comparison with existing stochastic mortality models. *Insurance: Mathematics and Economics*, **46**(1), 254–270.

ZUUR, A.F., FRYER, R.J., JOLLIFFE, I.T., DEKKER, R. and BEUKEMA, J.J. (2003) Estimating common trends in multivariate time series using dynamic factor analysis. *Environmetrics,* **14**(7), 665–685.

HELENA CHULIÁ
Universitat de Barcelona, Department of Econometrics and Riskcenter-IREA, Spain
E-Mail: hchulia@ub.edu

MONTSERRAT GUILLÉN
Department of Econometrics and Riskcenter-IREA, Universitat de Barcelona, Spain
E-Mail: mgillen@ub.edu

JORGE M. URIBE (Corresponding author)
Department of Economics, Universidad del Valle, Colombia
E-Mail: jorge.uribe@correounivalle.edu.co

APPENDIX

TABLE A1

AIC CRITERION FOR MODEL COMPARISONS FOR DIFFERENT POPULATIONS.

		DGDFM	DDFM	DFM	LC
Males	0–109+	−6.5249	**−6.6173**	−5.0002	−4.3895
	18–64	−2.7058	**−2.7622**	−2.0039	−1.6170
	65–109+	−2.9117	**−3.0171**	−2.4105	−1.9965
Females	0–109+	**−6.2723**	−6.2706	−5.0471	−4.8098
	18–64	−2.5560	**−2.5600**	−1.7946	−1.7946
	65–109+	−2.9900	**−2.9973**	−2.0796	−1.9105

NOTE: The statistics were calculated using data for the United Kingdom from 1950 to 2011, taken from the Human Mortality Database. The shadowed numbers correspond to the best model.