MODELLING MORTALITY DEPENDENCE WITH REGIME-SWITCHING COPULAS

BY

RUI ZHOU

ABSTRACT

We propose a two-regime Markov switching copula to depict the evolution of mortality dependence. One regime represents periods of high dependence and the other regime represents periods of low dependence. Each regime features a regular vine (R-vine) copula that, built on bivariate copulas, provides great flexibility for modelling complex high-dimensional dependence. Our estimated model indicates that the years of recovery from extreme mortality deterioration and the years of health care reform more likely fall into the low regime, while the years in which extreme mortality deteriorating events break out and the peaceful years without major mortality-impacting events more likely fall into the high regime. We use a case study to illustrate how the regime-switching copula can be applied to assess the effectiveness of longevity risk hedge with different beliefs about future mortality dependence evolution incorporated.

KEYWORDS

Mortality dependence, R-vine copula, regime-switching, longevity risk, S-forward.

1. INTRODUCTION

With increasing global attention on longevity risk, that is, the risk of higher than expected mortality improvement, capital market solutions for hedging/transferring longevity risk are developed and promoted. Standardized longevity hedge instrument, whose payoff is linked to standard mortality indexes, is among the most discussed due to its potential liquidity. The q-forward agreed upon by J.P. Morgan and the UK pension fund buy-out company Lucida in 2008 marks the first capital market longevity transaction.

An inevitable problem of standardized longevity hedge is population basis risk, arising from the difference between the population underlying the hedger's liability and that underlying the hedge instrument. To assess the impact of

Astin Bulletin 49(2), 373-407. doi:10.1017/asb.2019.10 © 2019 by Astin Bulletin. All rights reserved.

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

basis risk on hedge effectiveness, it is necessary to model mortality dependence among multiple populations. Furthermore, standard mortality indexes can be a weighted average of mortality rates from several populations in order to attract wider interest from hedgers and investors. Pricing hedge instruments written on such a weighted index requires the modelling of multi-population mortality dependence as well. Besides longevity risk hedge, a multi-population mortality model is also important for the risk management of international insurance and annuity providers. The mortality dependence of different policyholder groups determines the extent of natural hedge between insurance and annuity liabilities. This dependence also plays an important role in risk measures, such as VaR, thereby affecting the capital requirement.

A number of multi-population mortality models have been proposed by researchers, including Li and Lee (2005), Cairns et al. (2011), Dowd et al. (2011), Jarner and Kryger (2011), Li and Hardy (2011), Yang and Wang (2013), Zhou et al. (2013), Zhou et al. (2014) and Li et al. (2015). Most of these studies considered linear relationship between time series and assumed multivariate normal residuals. For example, Yang and Wang (2013) and Zhou et al. (2014) applied vector error correction model (VECM), a special case of vector autoregressive (VAR) model for multi-population mortality. VECM not only considers the autocorrelation and cross-correlation of multivariate time series but also incorporates the long-term relation or common trends between series. However, VECM only captures linear and symmetric dependence in time and between series, and therefore can be restrictive in a high-dimensional case where dependence structure is complex.

Wang et al. (2015), Chen et al. (2015), Chuliá et al. (2016) and Chen et al. (2017) proposed the application of multivariate copulas to capture a wider range of mortality dependence. More specifically, to model mortality dependence Chen et al. (2015) used a static one-factor copula in which the factor is common among all population groups. Wang et al. (2015) considered elliptical and Archimedean copulas with time-varying parameters. Chuliá et al. (2016) fitted generalized dynamic factor models to the differences of the log-mortality rates. Chen et al. (2017) improved the work of Chen et al. (2015) by adding a country-specific factor and allowing copula parameters to vary with time. Copula is a flexible tool for modelling high-dimensional dependence because it allows for the construction of a joint distribution by estimating marginal distributions and copula separately. Copula approach can capture nonlinear dependence, and thus it is more suitable for modelling complex relations. Since mortality data are nonstationary due to mortality improvement over time, differenced time series are often used when modelling mortality dependence with copulas. However, using solely differenced series may lead to the ignorance of important information contained in the original series, such as the common trend identified by VECM.

In this paper, we consider a special class of copula—vine copula for modelling mortality dependence. The idea of vine copula is to use bivariate copulas as the building blocks of multivariate copulas. Standard multivariate copula,

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

such as Gaussian copula and t copula, can be restrictive for high-dimensional data in the sense that it imposes the same dependence structure between pairs of variables. In contrast, vine copula allows for an arbitrary choice of dependence structure between pairs of variables and hence can be used to construct more complex high-dimensional dependence. Vine copula was first proposed by Joe (1996) to extend parametric bivariate extreme value copula families to higher dimensions. Bedford and Cooke (2001, 2002) formally introduced regular vine as a graphical model to specify multivariate distributions using bivariate copula densities. Aas et al. (2009) generalized the construction of vines by using arbitrary bivariate copulas.

Vine copula has been applied to the analyses of financial data by several researchers (Brechmann et al., 2012; Nikoloulopoulos et al., 2012; Brechmann and Czado, 2013; Dißmann et al., 2013). Chollete et al. (2009) first studied regime-switching vine copulas for equity returns with one regime restricted to a canonical vine (C-vine) copula, a special case of R-vine copula, and the other regime restricted to a multivariate Gaussian copula. Stöber and Czado (2014) extended Chollete et al. (2009) by considering a general R-vine copula for each regime. Fink et al. (2017) analysed the dependence of global equity and implied volatility indexes using regime-switching R-vine copulas.

The only application of vine copula in mortality modelling was conducted by Chuliá et al. (2016), in which a dynamic factor copula is fitted to mortality improvement rates and then a vine copula, nested under the factor copula, is used to model the dependence between idiosyncratic factors. In this paper, we use vine copula as the main dependence structure. More specifically, our approach is composed of three steps. In the first step, we apply an age-period-cohort (APC) model to decompose mortality rates to age, period and cohort effects. In the second step, we fit an ARIMA-GARCH model to each series of period effects in order to capture autocorrelation and conditional heteroscedasticity. Mortality heteroskedasticity has been detected and subsequently captured by GARCH model by several researchers, including Gao and Hu (2009), Chai et al. (2013), Chen et al. (2015), Lin et al. (2015), Wang and Li (2016) and Chen et al. (2017). In the second step, we model the dependence between the residuals series using copulas. In the final step, we construct a copula to model the dependence between the standardized residuals from the ARIMA-GARCH models.

We further extend our approach using a regime-switching copula to depict the evolution of mortality dependence over time with one high-dependence regime and one low-dependence regime. Wang et al. (2015), Chuliá et al. (2016), and Chen et al. (2017) proposed the capturing of dependence evolution by copulas with time-varying parameters, which are further modelled as functions of lagged variables and/or some score functions. Although these models are shown to yield better goodness of fit than static copulas, no model interpretation regarding how mortality dependence actually evolved historically has been provided. Interestingly, we find a claim in Wang et al. (2015) which states that there is lower dependence in a peaceful time period and higher dependence in

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

a time period of extreme mortality deterioration, but we could not verify this claim based on the results of Wang et al. (2015). If this statement is true, we should at times observe a change between high and low dependence. Based on this conjecture, we formed the idea of using two-regime Markov switching model to describe such transitions. An advantage of two-regime Markov switching model is its easy interpretation. By examining the probability of each year being in a high or low-dependence regime, we can obtain a clear picture of the historical evolution of mortality dependence.

Surprisingly, our approach brings us to a conclusion that is different from the statement made in Wang et al. (2015). We find that the period following an extreme mortality deterioration and the period of health care reforms are more likely to be in the low-dependence regime, suggesting that recovery from extreme mortality deterioration and health care reforms in different populations may take very different paths. However, the periods that an extreme mortality-deteriorating event breaks out and the peaceful periods with no major mortality-impacting events show a high probability in the high-dependence regime, indicating that the simultaneous occurrence of extreme mortality deterioration in various populations could be a result of the inherent dependence between these populations.

The two-regime Markov switching copula can be readily applied to longevity risk hedging and pricing. In this paper, we illustrate its use in the assessment of longevity hedge effectiveness. We assume that an annuity provider seeks to hedge the longevity risk in its annuity liability using S-forward. We consider three possible routes that the hedger believes mortality dependence would evolve in a 10-year hedge horizon: first, there is no major mortality-impacting event occurring in this period and hence mortality dependence stays in the high regime; second, mortality dependence switches between high and low regimes according to a Markov process; and third, there is no mortality dependence. We examine how the beliefs of future mortality dependence evolution affect the assessment of hedge effectiveness.

The remainder of this paper is organized as follows. Section 2 describes the mortality data and base mortality model. Section 3 discusses the marginal model for period components. Section 4 introduces R-vine copula and its estimation. Section 5 extends the single-regime R-vine copula to a two-regime Markov switching copula. Section 6 sets up a case study and illustrates how the regime-switching copula can be used to assess the effectiveness of longevity hedge. Section 7 demonstrates the selection of regime-switching R-vine copula for a dataset with higher dimension. Section 8 concludes the paper.

2. Mortality Data and Base Mortality Model

We obtain the number of deaths and the number of exposures-at-risk for populations of Italy, France, England and Wales, and Switzerland from Human Mortality Database.[1] The age range we use is 50–89 and the sample period

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

is 1900–2012. We intentionally use a long sample period in order to analyse potential changes in the mortality dependence amongst the four populations.

We employ the simple APC model (Osmond, 1985) as the base mortality structure. The mortality rates of each population are modelled by one individual APC model expressed as follows:

$$\ln(m_{x,t}^{(i)}) = a_x^{(i)} + k_t^{(i)} + g_{t-x}^{(i)},$$

where $i = 1, 2, 3, 4$ represents male populations of Italy, France, England and Wales, and Switzerland, respectively; $m_{x,t}^{(i)}$ is the central death rate of an individual aged x in year t from population (i); $a_x^{(i)}$ is the age-specific parameter for population (i), referred to as age effect; $k_t^{(i)}$ is the time-varying parameter for population (i), referred to as period effect; and $g_{t-x}^{(i)}$ is the cohort-specific parameter for population (i), referred to as cohort effect.

To estimate the APC model, we apply maximum likelihood estimation method. In particular, we assume that the number of deaths follows a Poisson distribution:

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

where $D_{x,t}^{(i)}$ and $E_{x,t}^{(i)}$ are the number of deaths and the number of exposures-at-risk at age x in year t for population (i). The loglikelihood function for population (i) is then calculated as

$$\sum_{x=50}^{89} \sum_{t=1900}^{2012} \left(D_{x,t}^{(i)} \ln\left(m_{x,t}^{(i)}\right) - m_{x,t}^{(i)} E_{x,t}^{(i)}\right) + \text{constant}.$$

To ensure parameter uniqueness, we impose three identification constraints when fitting each APC model: $\sum_{t=1900}^{2012} k_t^{(i)} = 0$, $\sum_{t-x=1811}^{1962} g_{t-x}^{(i)} = 0$, and $\sum_{t-x=1811}^{1962} (t-x) g_{t-x}^{(i)} = 0$.

Figure 1 shows the estimated $a_x^{(i)}$, $k_t^{(i)}$ and $g_{t-x}^{(i)}$ for the four populations. The age and period components of the four populations follow similar trends. As commonly found in existing literatures, mortality deteriorates as an individual ages and improves with time due to ever advancing health care and living conditions. The cohort components of England and Wales and Switzerland populations share a similar pattern, both with individuals born between 1860 and 1920 at the highest level. The Italian cohort effect is less pronounced than the other three populations. The French cohort effect varies significantly for individuals born between 1880 and 1962. More discussion on cohort effects of mortality rate can be found in Wilmoth (2005), Janssen (2005), and Beltrán-Sánchez et al. (2012).

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

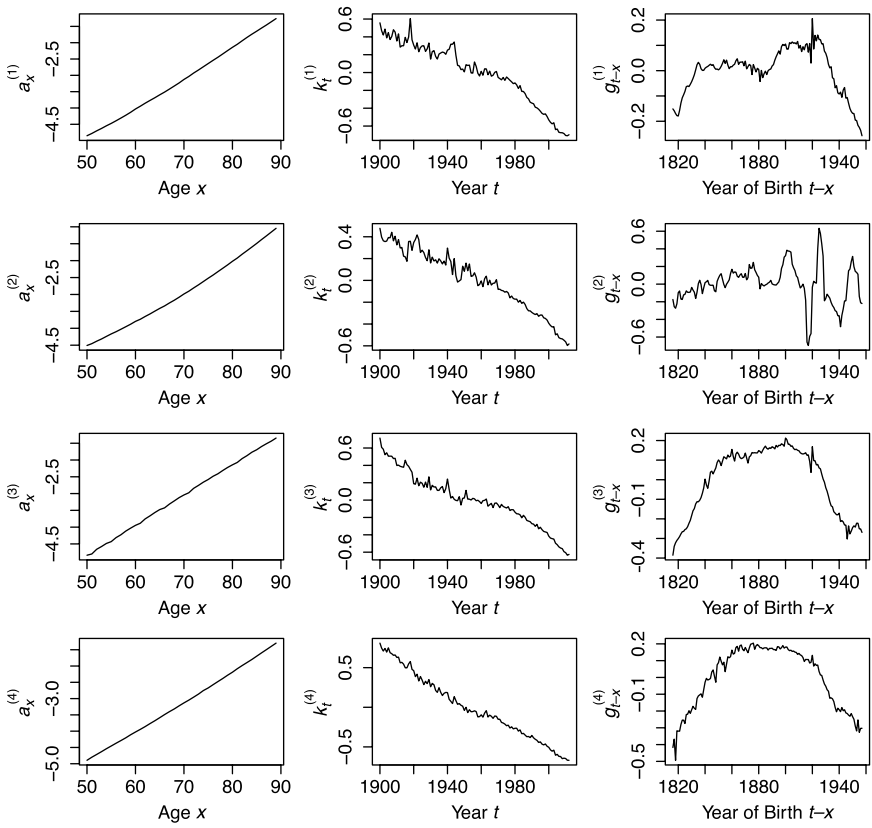


FIGURE 1: Parameter estimates for the age-period-cohort models.

3. THE MARGINAL MODELS OF PERIOD EFFECTS

The forecast of future mortality rates for an existing cohort using the APC model requires the values of $a_x^{(i)}$, $k_t^{(i)}$ and $g_{t-x}^{(i)}$. While $a_x^{(i)}$ and $g_{t-x}^{(i)}$ for $x \in [50, 89]$ and $t - x \in [1811, 1962]$ are among the parameter estimates obtained from fitting historical data to the APC model, $k_t^{(i)}$ for $t > 2012$ is not available and hence requires further modelling and forecasting. Time series models (Lee and Carter, 1992; Cairns et al., 2006, 2011; Chen et al., 2015; Wang et al., 2015; Chen et al., 2017) are frequently used for modelling the period effects. In this paper, we follow the route of Copula-GARCH method, which is a two-step approach first proposed by Jondeau and Rockinger (2006) to model the dependence between stock market returns. In the first step, we use ARIMA-GARCH as the marginal model to capture the autocorrelation and conditional heteroscedasticity in each series of period effects and obtain standardized residuals.

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

TABLE 1

BICs OF CANDIDATE ARIMA($p, 1, q$) MODELS FOR $k_t^{(1)}$.

p	q	BIC	p	q	BIC	p	q	BIC
0	0	−303.28	1	0	−309.21	2	0	−308.92
0	1	−313.83	1	1	−311.32	2	1	−306.84
0	2	−310.73	1	2	−306.93	2	2	−303.27
...	...	...	...	...	...	...	...	...

To determine the marginal model, we first select the lag orders for ARIMA-GARCH and then estimate the parameters using MLE approach. Recall that Figure 1 shows a linear downward trend in $k_t^{(i)}$, for $i = 1, \ldots, 4$. Using augmented Dickey–Fuller test, we find that taking first difference of $k_t^{(i)}$ removes its linear trend and obtains stationarity. Therefore, we consider an ARIMA model with a difference order of 1 as the marginal conditional mean model for $k_t^{(i)}$. An ARIMA($p, 1, q$)-GARCH(P, Q) model for $k_t^{(i)}$can be expressed as

$$\begin{aligned}
&\Delta k_t^{(i)} + \sum_{l=1}^{p} \phi_l \Delta k_{t-l}^{(i)} = \mu + \sum_{j=1}^{q} \theta_j \epsilon_{t-j+1}^{(i)}, \\
&\epsilon_t^{(i)} = \sigma_t^{(i)} z_t^{(i)}, \\
&\left(\sigma_t^{(i)}\right)^2 = \omega + \sum_{j=1}^{P} \beta_j \left(\sigma_{t-j}^{(i)}\right)^2 + \sum_{j=1}^{Q} \alpha_j \left(\epsilon_{t-j}^{(i)}\right)^2,
\end{aligned}$$

where $\Delta k_t^{(i)} = k_t^{(i)} - k_{t-1}^{(i)}$ and $z_t^{(i)} \sim i.i.d.(0, 1)$. For stationary and positivity, the GARCH model has the following parameter constraints: $\omega > 0$, $\alpha_j \geq 0$, $\beta_j \geq 0$, and $\sum_{j=1}^{Q} \alpha_j + \sum_{j=1}^{P} \beta_j < 1$.

In the following, we use $k_t^{(1)}$ as an example to show how we determine the orders of the ARIMA-GARCH model using both graphical and quantitative analysis. To select the AR and MA orders of the ARIMA model, we first examine the autocorrelation and partial autocorrelation for $\Delta k_t^{(1)}$. As shown in Figure 2, the ACF of $\Delta k_t^{(1)}$ displays a sharp cutoff at lag 1, while the PACF decays more slowly. The patterns of ACF and PACF plots suggest an MA order of 1. We then fit ARIMA($p, 1, q$) models with $p, q \leq 5$ to the estimated $k_t^{(1)}$ and choose the model with the lowest Bayesian information criterion (BIC). Table 1 presents an excerpt of the comparison of various ARIMA($p, 1, q$) models. The BIC comparison also recommends that ARIMA(0,1,1) fits $k_t^{(1)}$ best. We perform the same analysis on $k_t^{(i)}$ for $i = 1, 2, 3, 4$ and find that ARIMA(0,1,1) appears sufficient based on visual examination and also yields the lowest BIC for all $k_t^{(i)}$ series. Therefore, we use ARIMA(0,1,1) as the marginal conditional mean model.

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

TABLE 2

p-VALUES OF ARCH EFFECT TESTS FOR $\epsilon_t^{(1)}$.

	Lag	1	2	3	4	5
p-value	Ljung–Box	0.0129	0.0228	0.0147	0.0323	0.0576
	ARCH-LM	0.0144	0.0414	0.0456	0.0829	0.1433

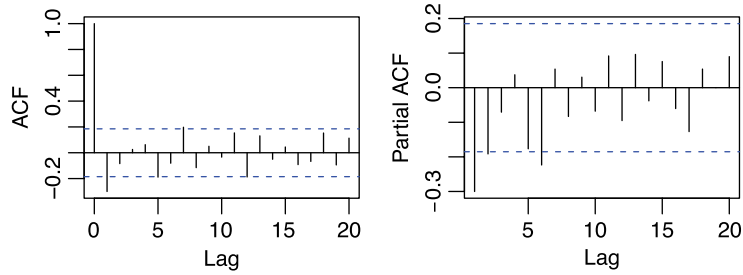


FIGURE 2: ACF and PACF plots for $\Delta k_t^{(1)}$.

After fitting the selected ARIMA model to $k_t^{(1)}$, we obtain the corresponding residuals $\epsilon_t^{(1)}$. To examine ARCH effect in the residuals, we plot the time series, ACF and PACF of the squared residuals $\left(\epsilon_t^{(1)}\right)^2$ in Figure 3. The series of the squared residuals displays periods of increased variation most notably in the first half of the twentieth century. The ACF and PACF at lag 1 are both significant, thereby indicating the presence of volatility clustering in the residual series. We further test the ARCH effect by performing the Ljung–Box test on the squared residual series and the ARCH-LM test (Engle, 1982) on the residual series. The p-values of the two tests are presented in Table 2. Both tests suggests that the residual series $\epsilon_t^{(1)}$ exhibits conditional heteroscedasticity for lags 1, 2 and 3 at 5% significance level.

To select the conditional variance model, we consider a collection of ARIMA(0,1,1)-GARCH(P, Q) models. The candidate GARCH orders include (0,1), (1,0) and (1,1). The candidate residual distributions include normal, skewed normal, Student, skewed Student, generalized error, skewed generalized error, generalized hyperbolic, normal inverse Gaussian and Johnson's SU. We again choose the model with the lowest BIC for the $k_t^{(1)}$ series.

We perform the same analysis to test ARCH effect and select the conditional variance model for each $k_t^{(i)}$ series. Table 3 presents the selected ARIMA-GARCH models for $k_t^{(i)}$ and their corresponding parameter estimates. It is easy to verify that all the estimated GARCH models satisfy the GARCH parameter constraints. Shape parameter is the parameter for residual distribution. Normal distribution does not have any shape parameter, while Student's t has one shape parameter as the degree of freedom.

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

TABLE 3

PARAMETER ESTIMATES AND THEIR STANDARD ERRORS (IN PARENTHESES) OF THE SELECTED ARIMA-GARCH MODEL FOR $k_t^{(i)}$.

Series		$k_t^{(1)}$	$k_t^{(2)}$	$k_t^{(3)}$	$k_t^{(4)}$
ARIMA order		(0,1,1)	(0,1,1)	(0,1,1)	(0,1,1)
ARIMA parameters	μ	−0.015469	−0.011648	−0.013792	−0.012326
		(0.001482)	(0.0015)	(0.001602)	(0.001025)
	θ_1	−0.366555	−0.427229	−0.414796	−0.575486
		(0.085465)	(0.083377)	(0.073163)	(0.077392)
GARCH order		(1,1)	(1,1)	(1,1)	(1,1)
GARCH parameters	ω	0.000018	0.000018	0.000007	0.000007
		(0.000022)	(0.000025)	(0.000025)	(0.000038)
	α_1	0.200905	0.160109	0.090078	0.095694
		(0.093225)	(0.078496)	(0.048508)	(0.084014)
	β_1	0.798089	0.833240	0.894928	0.886842
		(0.071084)	(0.055275)	(0.043612)	(0.109028)
Residual distribution		Normal	Student	Student	Normal
Shape parameter		N/A	4.871383	2.428820	N/A
			(1.816525)	(2.428820)	
Loglikelihood		185.7374	202.1227	217.2574	239.5558

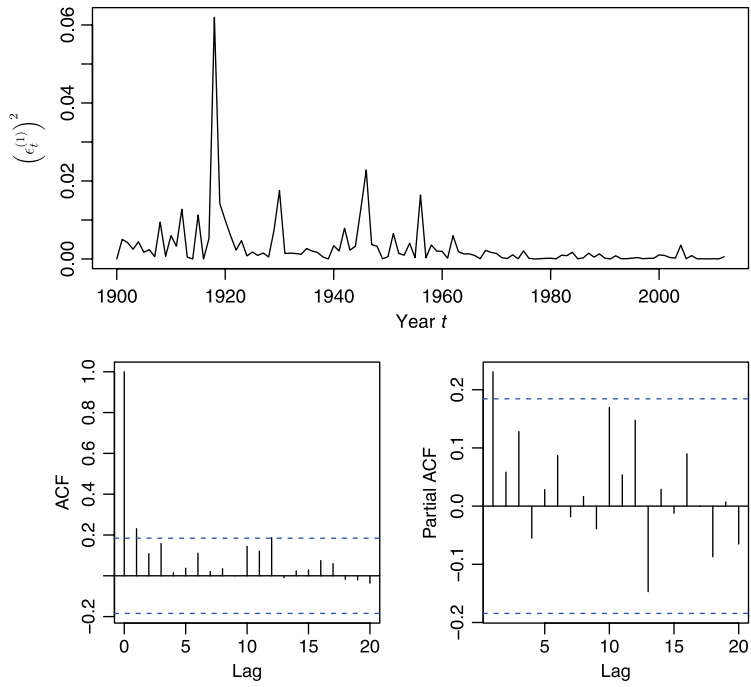


FIGURE 3: Time series, ACF and PACF plots for $\left(\epsilon_t^{(1)}\right)^2$.

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

We check model adequacy by performing Ljung–Box autocorrelation test on both standardized residuals and squared standardized residuals for each population. The Ljung–Box test p-values for 10 lags are all greater than 0.1, indicating that we cannot reject the null hypothesis of no autocorrelation in standardized residuals and squared standardized residuals at 10% significance level.

4. Modelling dependence structure with R-vine copulas

4.1. R-vine construction

The Sklar theorem (Sklar, 1959) states that for n variables $Y_1, Y_2, \ldots, Y_n$, there exists an n-dimensional copula such that the joint cumulative distribution function (CDF), $F(y_1, y_2, \ldots, y_n)$, can be expressed as

$$F(y_1, y_2, \ldots, y_n) = C\big(F_1(y_1), F_2(y_2), \ldots, F_n(y_n)\big),$$

where $F_i(.)$ denotes the marginal CDF of Y_i and $C(.)$ is the copula CDF.

In this paper, we consider some commonly used elliptical and Archimedean copulas, including Gaussian, t, Frank, Clayton, Gumbel, and rotated versions of Clayton and Gumbel copulas. Both Gaussian and Frank copulas have no tail dependence. t-copula has symmetric tail dependence. Clayton copula allows lower tail dependence but no upper tail dependence. Gumbel copula allows upper tail dependence but no lower tail dependence. A detailed description of these copulas can be found in Appendix A.

Standard multivariate copulas impose the same dependence structure on any pair of variables and hence are inflexible for modelling complex high-dimensional dependence. Vine copula, first proposed by Joe (1996) and further developed by Bedford and Cooke (2001, 2002), avoids the problem by building the joint distribution from bivariate copulas and allowing arbitrary copula for the dependence between any pair of variables.

A regular vine $\mathcal{V}$ on n variables is a sequence of $n-1$ connected trees $T_1, T_2, \ldots, T_{n-1}$, and satisfies the following properties (Bedford and Cooke, 2001, 2002):

- T_1 is a tree with node set $N_1 = \{1, \ldots, n\}$ and edge set E_1.
- For $i = 2, \ldots, n-1$, T_i is a connected tree with node set $N_i = E_{i-1}$, of which the cardinality is $n-(i-1)$, and edge set E_i.
- The proximity condition: If two nodes in T_i, for $i = 2, \ldots, n-1$, are joined by an edge, their corresponding edges in T_{i-1} must share a common node.

R-vine includes two simpler tree structures: canonical vine (C-vine) of which each tree has a node connected to all other nodes and drawable vine (D-vine) of which each tree is a path.

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

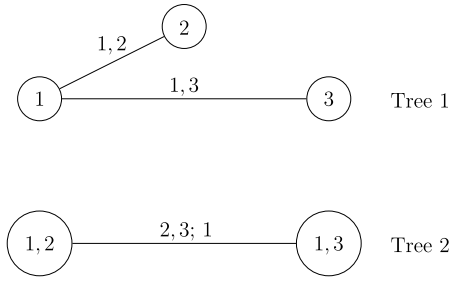


FIGURE 4: A three-dimensional R-vine structure.

To understand R-vine structure, let us first study a three-dimensional case. An illustration of a three-dimensional R-vine is shown in Figure 4. The node and edge sets of the first tree T_1 are $\{1, 2, 3\}$ and $\{\{1, 2\}, \{1, 3\}\}$, respectively. The node and edge sets of the second tree T_2 are $\{\{1, 2\}, \{1, 3\}\}$ and $\{2, 3\}$, respectively. In the second tree, $\{1\}$ is the conditioning set associated with the edge $\{2, 3\}$. Note that a three-dimensional R-vine is both a C-vine and a D-vine.

Next let us consider a random vector $\boldsymbol{Y} = (Y_1, Y_2, Y_3)$ which exhibits the R-vine dependence illustrated in Figure 4. It is well known that the joint density function for three variables Y_1, Y_2 and Y_3 can be written as

$$f(y_1, y_2, y_3) = f_1(y_1) \cdot f(y_2|y_1) \cdot f(y_3|y_2, y_1),$$

where $f_i(.)$ denotes the marginal density function of Y_i. Each of the conditional density function $f(.|.)$ can be further decomposed using conditional copulas as follows:

$$\begin{aligned} f(y_2|y_1) &= c_{12}(F_1(y_1), F_2(y_2)) \cdot f_2(y_2), \\ f(y_3|y_1, y_2) &= c_{23;1}(F_{2|1}(y_2|y_1), F_{3|1}(y_3|y_1)) \cdot f(y_3|y_1) \\ &= c_{23;1}\left(F_{2|1}(y_2|y_1), F_{3|1}(y_3|y_1)\right) \cdot c_{13}(F_1(y_1), F_3(y_3)) \cdot f_3(y_3), \end{aligned}$$

where c_{ij} and $c_{ij;k}$ are bivariate copula densities, and

$$\begin{aligned} c_{ij}(F_i(y_i), F_j(y_j)) &= \frac{\partial^2 C(F_i(y_i), F_j(y_j))}{\partial F_i(y_i) \partial F_j(y_j)}, \\ c_{ij;k}(F_{i|k}(y_i|y_k), F_{j|k}(y_j|y_k)) &= \frac{\partial^2 C(F_{i|k}(y_i|y_k), F_{j|k}(y_j|y_k))}{\partial F_{i|k}(y_i|y_k) \partial F_{j|k}(y_j|y_k)}. \end{aligned}$$

The joint density function of $\boldsymbol{Y}$ can be rewritten as follows:

$$\begin{aligned} f(y_1, y_2, y_3) = c_{23;1}(F_{2|1}(y_2|y_1), F_{3|1}(y_3|y_1)) \cdot c_{13}(F_1(y_1), F_3(y_3)) \cdot c_{12}(F_1(y_1), F_2(y_2)) \\ \cdot f_1(y_1) \cdot f_2(y_2) \cdot f_3(y_3), \end{aligned}$$

where $c_{23;1}(F_{2|1}(y_2|y_1), F_{3|1}(y_3|y_1)) \cdot c_{13}(F_1(y_1), F_3(y_3)) \cdot c_{12}(F_1(y_1), F_2(y_2))$ is the copula density.

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

This R-vine structure first models the dependence between Y_1 and Y_2 and that between Y_1 and Y_3. This R-vine then, conditioning on Y_1, models the conditional dependence between Y_2 and Y_3. If a different set of edges is used in T_1 then a new expression of joint density function will be derived and a new R-vine will be constructed. Therefore, there are three possible R-vine structures for this three-dimensional case.

This construction approach can be extended to an n-dimensional case. Let us denote the bivariate copula density associated with an edge $e = \{j(e), k(e)\} \in E_i$ by

$$c_{j(e),k(e);D(e)}(F_{j(e)|D(e)}(y_{j(e)}|\boldsymbol{y}_{D(e)}), F_{k(e)|D(e)}(y_{k(e)}|\boldsymbol{y}_{D(e)})),$$

where $D(e)$ is the conditioning set associated with the edge e and $\boldsymbol{y}_{D(e)} = \{y_i | i \in D(e)\}$. Bedford and Cooke (2001, 2002) showed that the joint density function of $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ can be written as

$$f(y_1, \ldots, y_n) = \prod_{k=1}^{n} f_k(y_k) \times \prod_{i=1}^{n-1} \prod_{e \in E_i} c_{j(e),k(e);D(e)}(F_{j(e)|D(e)}(y_{j(e)}|\boldsymbol{y}_{D_{(e)}}), F_{k(e)|D(e)}(y_{k(e)}|\boldsymbol{y}_{D_{(e)}})). \tag{4.1}$$

When the marginal density of Y_i is uniform on [0, 1], equation (4.1) represents the R-vine copula density $c(\boldsymbol{y}|\mathcal{V}, \boldsymbol{B}, \boldsymbol{\theta})$ and

$$c(\boldsymbol{y}|\mathcal{V}, \boldsymbol{B}, \boldsymbol{\theta}) = \prod_{i=1}^{n-1} \prod_{e \in E_i} c_{j(e),k(e);D(e)}(F_{j(e)|D(e)}(y_{j(e)}|\boldsymbol{y}_{D_{(e)}}), F_{k(e)|D(e)}(y_{k(e)}|\boldsymbol{y}_{D_{(e)}})), \tag{4.2}$$

where $\mathcal{V}$, $\boldsymbol{B}$ and $\boldsymbol{\theta}$ represent the R-vine structure, set of bivariate copulas associated with the R-vine and parameter vector, respectively.

Since different sets of nodes and edges yield different R-vines and we allow an arbitrary choice of copula for each pair of variables, the number of joint dependences that can be constructed using R-vine increase rapidly with the dimension n.

4.2. R-vine selection and estimation

Dißmann et al. (2013) proposed a sequential method to select and estimate R-vine structure. This method captures high dependence at the lower layers of the R-vine, thereby leaving lower dependence to be modelled at the higher layers. Bivariate copulas are specified for pairs of variables capturing the strongest pairwise dependencies present in the dataset. The selection procedure is summarized as follows:

1. Calculate empirical pairwise Kendall's tau $\hat{\tau}_{j(e),k(e)}$ for all possible pairs of variables $Y_{j(e)}$ and $Y_{k(e)}$ with $j(e), k(e) = 1, 2 \ldots, n$.

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

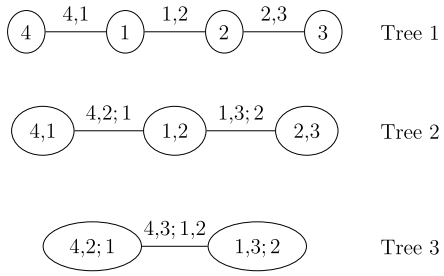


FIGURE 5: Selected R-vine tree structure for $z_t^{(i)}$, $i = 1, 2, 3, 4$.

2. Select the spanning tree that maximizes the sum of absolute empirical Kendall's tau, that is,

$$\max \sum_{e=\{j(e),k(e)\}\in E_1} |\hat{\tau}_{j(e),k(e)}|.$$

3. For each edge in the selected spanning tree T_1, select bivariate copulas for the associated pairs of variables based on BIC and estimate the corresponding parameter(s) using MLE method.
4. For the ith tree, calculate the empirical Kendall's tau $\hat{\tau}_{j(e),k(e)|D(e)}$ for all conditional variable pairs $Y_{j(e)}|\boldsymbol{Y}_{D(e)}$ and $Y_{k(e)}|\boldsymbol{Y}_{D(e)}$ which are connected by an edge satisfying proximity condition.
5. Select the spanning tree T_i that maximizes the sum of absolute empirical Kendall's taus, that is,

$$\max \sum_{e=\{j(e),k(e)|D(e)\}\in E_i} |\hat{\tau}_{j(e),k(e)|D(e)}|.$$

6. For each edge in the selected spanning tree T_i, select bivariate copulas for the associated pairs of variables based on BIC and estimate the corresponding parameter(s) using MLE method.
7. Perform Steps 4 to 6 for $i = 2, 3, \ldots n-1$ sequentially.

4.3. Estimated R-vine copula for mortality dependence

Modelling the dependence between standardized residuals $z_t^{(i)}$, $i = 1, 2, 3, 4$ requires a four-dimensional R-vine copula. Note that a four-dimensional R-vine is either a C-vine or a D-vine. The R-vine structure selected using the sequential selection and estimation approach is displayed in Figure 5. We also calculate pairwise estimated Kendall's tau between the standardized residuals in Table 4. It is easy to verify that the edge set $\{\{4, 1\}, \{1, 2\}, \{2, 3\}\}$ maximizes the sum of absolute empirical Kendall's tau for the first tree.

Table 5 presents the estimated parameters of the selected R-vine copula. The bivariate copulas in Tree 1 have relatively high Kendall's tau and also

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

TABLE 4

PAIRWISE ESTIMATED KENDALL'S TAU BETWEEN $z_t^{(i)}$, $i = 1, 2, 3, 4$.

Variable	$z_t^{(1)}$	$z_t^{(2)}$	$z_t^{(3)}$	$z_t^{(4)}$
$z_t^{(1)}$	1.00			
$z_t^{(2)}$	0.36	1.00		
$z_t^{(3)}$	0.22	0.27	1.00	
$z_t^{(4)}$	0.37	0.36	0.24	1.00

TABLE 5

SELECTED R-VINE STRUCTURE AND PARAMETER ESTIMATES.

						Tail dependence	
Tree	Edge	Copula	First parameter	Second parameter	tau	Upper	Lower
1	2,3	Gumbel	1.41 (0.11)	–	0.29	0.37	–
	1,2	Survival Gumbel	1.60 (0.12)	–	0.38	–	0.46
	4,1	Student's t	0.55 (0.07)	3.42 (1.54)	0.37	0.32	0.32
2	1,3; 2	Independent	–	–	0.00	–	–
	4,2; 1	Frank	2.19 (0.61)	–	0.23	–	–
3	4,3; 1,2	Independent	–	–	0.00	–	–

TABLE 6

GOODNESS OF FIT FOR VARIOUS MULTIVARIATE COPULAS.

Copula	Loglikelihood	Number of parameters	AIC	BIC
Gaussian	55.12	6	−98.24	−81.93
Student's t	60.88	7	−107.76	−88.73
Gumbel	44.66	1	−87.32	−79.88
Clayton	46.76	1	−91.52	−84.08
Frank	46.68	1	−91.36	−83.92
R-vine	60.68	5	−111.37	−97.77

capture possible upper and lower tail dependences. Tree 2 includes an independent copula and a Frank copula which has relatively low Kendall's tau and no tail dependence. Tree 3 only involves an independent copula. As a result, the most dependence is captured in Tree 1, echoing the motivation of the sequential selection method by Dißmann et al. (2013).

To study the advantage of R-vine copula over standard multivariate copulas, we also fit multivariate Gaussian, t, Gumbel, Clayton and Frank copulas to the standardized residual series. Table 6 summarizes the loglikelihood, Akaike information criterion (AIC), and BIC of these copulas. It is clear that R-vine provides much better goodness of fit than other copulas.

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

5. Markov regime-switching copula

5.1. Identifying regimes

Wang et al. (2015), Chuliá et al. (2016) and Chen et al. (2017) used time-varying copulas to capture changes of mortality dependence over time. They modelled the evolution of copula parameters by functions of lagged variables and/or some score functions. While these time-varying copulas are shown to yield better goodness of fit than stationary ones, no model interpretation is found in these papers. Given the estimated models, there is still no indication of how mortality dependence has changed over time and of what drove these changes. To bridge this gap, we consider a simple Markov regime-switching model for the evolution of mortality dependence.

We examine the occurrence of regime switching in the historical data through the rolling window analysis proposed by Stöber and Czado (2014) and Fink et al. (2017). Each window consists of 40 consecutive data points. For simplicity, we assume that each regime of the Markov switching R-vine exhibits the same tree structure with that of the single-regime R-vine shown in Figure 5. We further assume that a bivariate copula associated with the regime-switching R-vine is either Gaussian, Student's t, Gumbel or Survival Gumbel. These copula families include four potential tail dependence behaviours: none, symmetric, upper and lower correlations.

For each edge in Tree 1, we fit the four families of bivariate copulas to the corresponding bivariate data in a rolling window. We repeat the fitting exercise for all rolling windows. We then plot the BICs of the fitted bivariate copulas and examine whether the best-fitting copula family has changed over time. The top left panel of Figure 6 presents the BICs of the four bivariate copulas for the edge {4,1}. Gumbel copula fits the best for the windows with starting observation in 1900–1942. Survival Gumbel performs the best for the windows with starting observation in 1955–1972. Gaussian copula yields the lowest BICs for the windows with starting observation in 1943–1955. However, the BICs of Gaussian copula in these windows are only marginally better than those of Gumbel or Survival Gumbel. Therefore, switching from Gumbel in early periods to Survival Gumbel in later periods appears sufficient for the edge {4,1}.

The observed switching coincides with the fact that the first half of the twentieth century underwent several mortality disruptions while the second half of the century was much more peaceful. The flu epidemic in 1918, the World War I between 1914–1916 and the World War II between 1940–1945 caused severe mortality deteriorations in many countries. After the World War II, the European countries experienced great economic development, thereby inducing strong mortality improvement. We will further analyse the causes of regime changes after fitting a regime-switching model.

Similar graphical analysis of the other two panels in Figure 6 indicates that switching from Gaussian to Survival Gumbel is most suitable for the edge

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

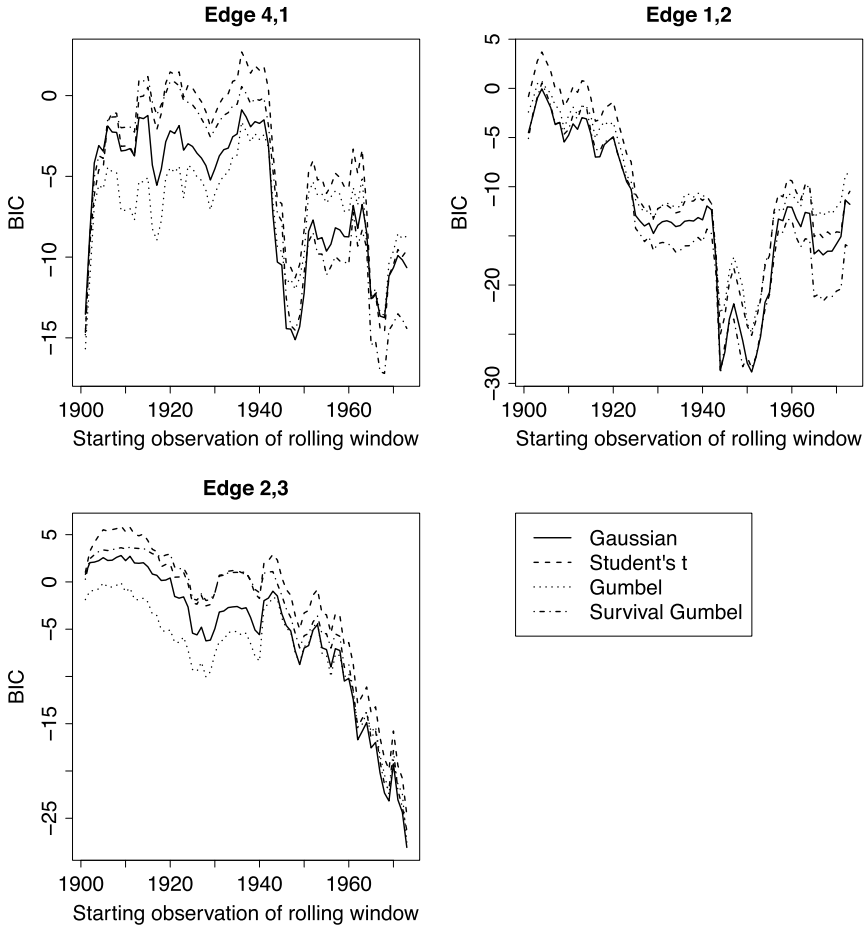


FIGURE 6: Rolling window analysis for the edges in Tree 1.

{1,2}, and switching from Gumbel to Gaussian for the edge {2,3}. For Trees 2 and 3, we only consider switching of copula parameters but not copula families since the dependence captured by these two trees are low and switching copula families only marginally improve the model. In addition, we replace the Frank copula for the edge {4,2;1} by Gaussian in order to be consistent with the copula families used in Tree 1.

Ideally, we should search all possible combinations of bivariate copulas and tree structures for the regime-switching model and find the one that best fits our data. However, such an approach is very time-consuming due to the large number of combinations and the iterative numerical optimization procedure for maximizing loglikelihood.

The selected tree structure and bivariate copulas using the rolling window analysis are summarized in Table 7 with regime 1 representing peaceful periods and regime 2 representing periods in which mortality disruptions occur.

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

TABLE 7

TREE STRUCTURE AND BIVARIATE COPULAS SELECTED FOR THE REGIME-SWITCHING R-VINE COPULA.

Tree	Edge	Regime 2	Regime 1
1	4,1	Gumbel	Survival Gumbel
	1,2	Gaussian	Survival Gumbel
	2,3	Gumbel	Gaussian
2	1,3; 2	Independent	Independent
	4,2; 1	Gaussian	Gaussian
3	4,3; 1,2	Independent	Independent

5.2. Loglikelihood of regime-switching R-vine copula

Suppose that the regime of mortality dependence at time t depends on a latent Markov process s_t. Let $s_t = 1$ if mortality dependence is in regime 1 and $s_t = 2$ if mortality dependence is in regime 2. Assuming that there are in total K regimes, the transition probability matrix among the regimes can be written as

$$P = \begin{pmatrix} P_{11} & P_{12} & \cdots & P_{1K} \\ P_{21} & P_{22} & \cdots & P_{2K} \\ \cdots & \cdots & \cdots & \cdots \\ P_{K1} & P_{K2} & \cdots & P_{KK} \end{pmatrix},$$

where P_{ij} represents the probability of transition from regime i to regime j. The stationary distribution of the Markov process is represented by a $K \times 1$ vector $\boldsymbol{\pi}$ which solves the following equation:

$$\boldsymbol{\pi}' P = \boldsymbol{\pi}'.$$

Let $\boldsymbol{u}_t = (u_{1t}, \ldots, u_{nt})$ be the uniformly distributed pseudo observations of the standardized residuals $(z_t^{(1)}, \ldots, z_t^{(n)})$ and $\tilde{\boldsymbol{u}}_t = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_t)$. The parameter vector of the regime-switching copula is denoted by $\boldsymbol{\theta}' = (\boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_K', \boldsymbol{\theta}_{mc}')$, where $\boldsymbol{\theta_i}$ is the parameter vector of the bivariate copulas associated with regime i for $i = 1, \ldots, K$, and $\boldsymbol{\theta}_{mc}$ is a vector that contains all the elements of P. The joint density function $f(\tilde{\boldsymbol{u}}_T|\boldsymbol{\theta})$ can be written into product of conditional densities as follows:

$$\begin{aligned} f(\tilde{\boldsymbol{u}}_T|\boldsymbol{\theta}) &= f(\boldsymbol{u}_1|\boldsymbol{\theta}) \prod_{t=2}^{T} f(\boldsymbol{u}_t|\tilde{\boldsymbol{u}}_{t-1}, \boldsymbol{\theta}) \\ &= \left[\sum_{k=1}^{K} f\left(\boldsymbol{u}_1|s_1 = k, \boldsymbol{\theta}_k\right) \Pr\left(s_1 = k|\boldsymbol{\theta}_{mc}\right) \right] \\ &\quad \cdot \prod_{t=2}^{T} \left[\sum_{k=1}^{K} f\left(\boldsymbol{u}_t|s_t = k, \boldsymbol{\theta}_k\right) \Pr\left(s_t = k|\tilde{\boldsymbol{u}}_{t-1}, \boldsymbol{\theta}\right) \right], \end{aligned} \tag{5.1}$$

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

where $f(\boldsymbol{u}_t|s_t = k, \boldsymbol{\theta}_k)$ is the joint density of the regime-k R-vine copula evaluated at $\boldsymbol{u}_t$ using equation (4.2) and $\Pr(s_1 = k|\boldsymbol{\theta}_{mc})$ is assumed to be equal to $\boldsymbol{\pi}_k$, the kth element of $\boldsymbol{\pi}$.

To obtain the values of $\Pr(s_t = k|\tilde{\boldsymbol{u}}_{t-1}, \boldsymbol{\theta}_{mc})$, we apply the iterative method developed in Hamilton (1989). Let $\boldsymbol{\Omega}_{t|t}$ and $\boldsymbol{\Omega}_{t|t-1}$ be $K \times 1$ probability vectors with the kth element defined as follows:

$$(\boldsymbol{\Omega}_{t|t-1})_k = \Pr(s_t = k|\tilde{\boldsymbol{u}}_{t-1}, \boldsymbol{\theta}),$$
$$(\boldsymbol{\Omega}_{t|t})_k = \Pr(s_t = k|\tilde{\boldsymbol{u}}_t, \boldsymbol{\theta}).$$

As shown in Appendix B, we have the following expressions:

$$(\boldsymbol{\Omega}_{t|t-1})_k = \sum_{j=1}^{K} P_{jk}(\boldsymbol{\Omega}_{t-1|t-1})_j, \tag{5.2}$$

$$(\boldsymbol{\Omega}_{t|t})_k = \frac{f(\boldsymbol{u}_t|s_t = k, \boldsymbol{\theta}_k)(\boldsymbol{\Omega}_{t|t-1})_k}{\sum_{j=1}^{K} f(\boldsymbol{u}_t|s_t = j, \boldsymbol{\theta}_j)(\boldsymbol{\Omega}_{t|t-1})_j}. \tag{5.3}$$

We then calculate $\boldsymbol{\Omega}_{t|t-1}$ and $\boldsymbol{\Omega}_{t|t}$ iteratively for $t = 1, 2, \ldots, T$ with the starting value $\boldsymbol{\Omega}_{1|0} = \boldsymbol{\pi}$.

5.3. Model estimation

Due to the latent state variables and complex dependence structure, parameter estimation by direct loglikelihood maximization is difficult. Stöber and Czado (2014) proposed a stepwise expectation–maximization (EM) algorithm to estimate regime-switching R-vine copula. A detailed description of the algorithm is presented in Appendix C.

Table 8 presents the parameter estimates and goodnesses of fit of the selected regime-switching R-vine copula. We note that the Kendall's taus of the bivariate copulas in regime 2 are smaller than those in regime 1, indicating that regime 2 is a low-dependence regime. Furthermore, the Gumbel parameters of edges {2,3} and {4,1} in regime 2 are not significantly different from 1 and the Gaussian parameter of edge {1,2} is not significantly different from 0 at 5% significance level. In contrast, the copula parameters for regime 1 are mostly higher than those of the single-regime R-vine copula shown in Table 5, suggesting a stronger dependence. Therefore, using regime-switching copula leads us to one high-dependence regime and one low-dependence regime.

To examine which years correspond to regime 1 and which years correspond to regime 2, we calculate the smoothed probability of mortality dependence in regime k in year t, $(\boldsymbol{\Omega}_{t|T})_k = \Pr(s_t = k|\tilde{\boldsymbol{u}}_T, \boldsymbol{\theta})$. Kim (1994) shows that the smoothed probability can be calculated using the following equation with backward iterations:

$$(\boldsymbol{\Omega}_{t|T})_k = \frac{\sum_{j=1}^{K} (\boldsymbol{\Omega}_{t+1|T})_j \cdot (\boldsymbol{\Omega}_{t|t})_k \cdot P_{kj}}{(\boldsymbol{\Omega}_{t+1|t})_j}$$

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

TABLE 8

PARAMETER ESTIMATES AND GOODNESSES OF FIT OF THE SELECTED REGIME-SWITCHING R-VINE COPULA.

Regime 1						
					Tail dependence	
Tree	Edge	Family	Parameter	tau	Upper	Lower
1	2,3	Gaussian	0.72 (0.06)	0.51	–	
	1,2	Survival Gumbel	2.49 (0.42)	0.60	–	0.68
	4,1	Survival Gumbel	2.00 (0.24)	0.50	–	0.59
2	1,3; 2	Independent	–	0.00	–	–
	4,2; 1	Gaussian	0.33 (0.13)	0.22	–	–
3	4,3; 1,2	Independent	–	0.00	–	–
Regime 2						
					Tail dependence	
Tree	Edge	Family	Parameter	tau	Upper	Lower
1	2,3	Gumbel	1.13 (0.11)	0.12	0.16	–
	1,2	Gaussian	0.23 (0.15)	0.15	–	–
	4,1	Gumbel	1.26 (0.16)	0.21	0.27	–
2	1,3; 2	Independent	–	0.00	–	–
	4,2; 1	Gaussian	0.29 (0.13)	0.19	–	–
3	4,3; 1,2	Independent	–	0.00	–	–
$p_{11} = 0.91(0.06)$, $p_{22} = 0.88(0.06)$						
Loglike=71.68, AIC=−123.36, BIC=−96.18						

Figure 7 plots the smoothed probability of mortality dependence in regime 1 using the regime-switching R-vine copula. It is interesting to see that there are four periods of low dependence: 1903–1908, 1914–1927, 1937–1950 and 1989–1996. There are several well-known events that occurred in these periods. World War I took place in 1914–1918. Spanish flu broke out in 1918. World War II lasted from 1939 to 1945. Health care system reforms occurred in several European countries around 1990. France introduced budgets in the public hospital sector in 1985. UK introduced substantial changes in its National Health System beginning in 1990. Italy started a reform to improve chronic probability of deficits and efficiency in service provision in 1993. During these turbulent periods, mortality dependence has a low probability in the high-dependence regime. A possible explanation is that health care reform or recovery from catastrophic events often lasts several years and takes different paths in different countries.

We also observe a probability of 0.02 in the high-dependence regime for 1918, the year that Spanish flu broke out. Since the flu epidemic broke out and caused mortality jumps at approximately the same time in these countries, we initially expected a high probability in the high-dependence regime. This discrepancy between our expectation and the model result is caused by

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

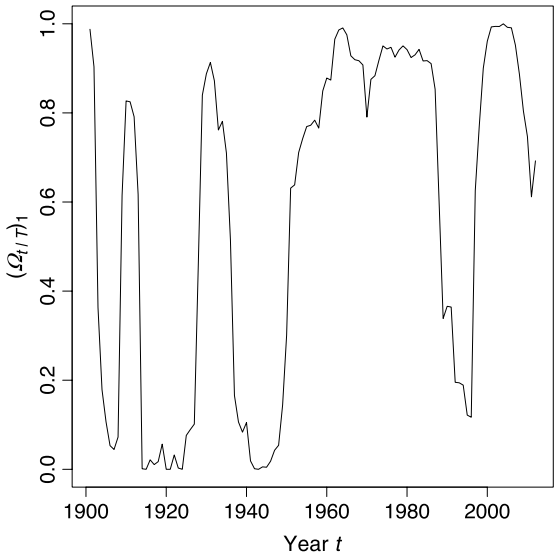


FIGURE 7: Smoothed probability of the process being in regime 1 using the regime-switching R-vine copula.

the preference given to staying in the same regime when maximizing the likelihood for the regime-switching R-vine copula. Equations (5.1) and (5.2) indicate that the basic component of loglikelihood is $f(\boldsymbol{u}_t|s_t = k, \boldsymbol{\theta}_k)P_{jk}(\boldsymbol{\Omega}_{t-1|t-1})_j$. The mortality dependence of 1917 more likely falls into the low-dependence regime 2 since the smoothed probability of being in regime 2 in 1917 is 0.01. While transitioning to regime 1 in 1918 better matches the history and provides a higher value for the conditional density function $f(\boldsymbol{u}_{1918}|s_t = 1, \boldsymbol{\theta}_1)$, the product $f(\boldsymbol{u}_{1918}|s_t = 1, \boldsymbol{\theta}_1)P_{21}(\boldsymbol{\Omega}_{1917|1917})_2$ can be significantly lower than the product $f(\boldsymbol{u}_{1918}|s_t = 2, \boldsymbol{\theta}_2)P_{22}(\boldsymbol{\Omega}_{1917|1917})_2$ which corresponds to staying in regime 2 due to low transition probability P_{21} (estimated to be 0.12). Therefore, staying in regime 2 in 1918 is preferred in order to maximize the loglikelihood.

Although the regime-switching R-vine copula provides insight into the evolution of mortality dependence, it does not yield better goodness of fit than the single-regime R-vine copula. Due to the additional six parameters (two transition probabilities and four copula parameters), the BIC of the regime-switching copula is slightly higher than that of the single-regime R-vine. Since we have observed that all three parameters that are associated with the first tree in regime 2 are insignificant, we consider a reduced regime-switching R-vine of which regime 2 assumes independence. Table 9 reports the parameter estimates and goodnesses of fit of the reduced regime-switching copula. This copula yields the lowest BIC among all the models we have considered so far.

Figure 8 plots the smoothed probabilities of mortality dependence in regime 1 using the reduced regime-switching copula. Figure 8 is same as Figure 7 except that it is more jagged than Figure 7 during the turbulent periods. In addition, we observe that year 1918 has a probability of 0.82

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

TABLE 9

PARAMETER ESTIMATES AND GOODNESSES OF FIT OF THE REDUCED REGIME-SWITCHING R-VINE COPULA.

Regime 1						
					Tail dependence	
Tree	Edge	Family	Parameter	tau	Upper	Lower
1	2,3	Gaussian	0.65 (0.07)	0.45	–	–
	1,2	Survival Gumbel	2.17 (0.27)	0.54	–	0.62
	4,1	Survival Gumbel	1.98 (0.22)	0.50	–	0.58
2	1,3; 2	Independent	–	0.00	–	–
	4,2; 1	Gaussian	0.34 (0.11)	0.22	–	–
3	4,3; 1,2	Independent	–	0.00	–	–
Regime 2 Independent						
$p_{11} = 0.91(0.06)$, $p_{22} = 0.75(0.13)$ Loglike=67.90, AIC=−123.8, BIC=−107.49						

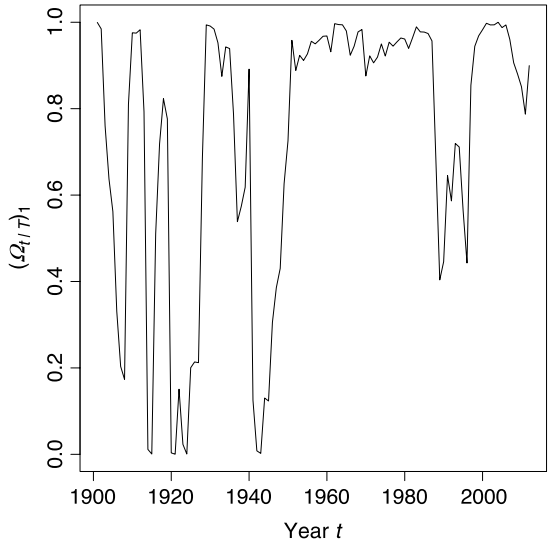


FIGURE 8: Smoothed probability of mortality dependence in regime 1 using the reduced regime-switching R-vine copula.

in high-dependence regime in Figure 8 as compared to 0.02 in Figure 7. To understand this change, we again refer to the loglikelihood component $f(\boldsymbol{u}_t|s_t = k, \boldsymbol{\theta}_k)P_{jk}(\boldsymbol{\Omega}_{t-1|t-1})_j$. The mortality dependence is more likely in regime 1 in 1917 since $(\boldsymbol{\Omega}_{1917|T})_1 = 0.72$. Staying in regime 1 in 1918 yields a higher conditional density function $f(\boldsymbol{u}_{1918}|s_t = 1, \boldsymbol{\theta}_1)$ because the high-dependence regime 1 better matches the simultaneous mortality jump that occurred in several

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

countries in 1918. The transition probability P_{11} is also high, thereby leading to a large value for $f(\boldsymbol{u}_t|s_t = k, \boldsymbol{\theta}_k)P_{jk}(\boldsymbol{\Omega}_{t-1|t-1})_j$. Therefore, the loglikelihood maximization prefers staying in regime 1 in 1918.

6. Mortality forecast and longevity hedge

6.1. A case study

Based on previous analysis, mortality dependence may fall in either the high-dependence regime or the independent regime. In the sample period of 1901–2012, 88 out of 112 years have a greater than 60% probability of being in the high-dependence regime. From the perspective of longevity hedge, higher mortality dependence between the population underlying the hedger's liability and that underlying the hedging instrument often means more effective hedging. When the hedging horizon is short, the hedger may predict that a major event with significant impact on mortality is unlikely to occur and thus assume that future mortality dependence stays in the high regime. As a result, the longevity hedge may be deemed more effective under this assumption in comparison to using the regime-switching copula.

In this section, we set up a case study to illustrate how different opinions of future mortality dependence evolution may affect the assessment of longevity hedge effectiveness based on the estimated regime-switching copula. Let us assume that an annuity provider issues annuities to 60-year-old individuals from the four populations at the beginning of 2013, with 25% of policies in each population. Starting from age 70, the annuity pays \$1 to a surviving annuitant at the beginning of each year until the annuitant reaches age 90. The annuity provider is interested in hedging the longevity risk in its annuity liability over a 10-year horizon.

In particular, the annuity provider plans to use S-forwards to hedge the volatility in its annuity liability in 10 years. An S-forward involves the exchange of a realised survival rate of a given population cohort in return for a fixed survival rate that has been agreed upon at the beginning of the contract. As a longevity risk hedger, the annuity provider should receive the realised survival rate and pay the fixed survival rate. To maximize hedge effectiveness, we assume that the population cohort underlying the S-forward is also 60 years old in 2013.

Denote ${}_tp^{(i)}_{x,t_0}$ as the t-year survival rate of an individual from population (i) who is x years old at the beginning of year t_0. Assuming constant force of mortality between consecutive integer ages, ${}_tp^{(i)}_{x,t_0}$ can be approximated by

$$\begin{aligned} {}_tp^{(i)}_{x,t_0} &= \prod_{k=0}^{t-1} p^{(i)}_{x+k,t_0+k} \\ &\approx \prod_{k=0}^{t-1} e^{-m^{(i)}_{x+k,t_0+k}}. \end{aligned}$$

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

Assume that annual effective interest rate is 3%. Assessed in year 10, the value of a T-year S-forward written on the survival rate of population (i) with a notional amount of 1 is

$$v_s(T) = E\left[1.03^{10-T}\left({}_Tp^{(i)}_{60,2013} - {}_T\hat{p}^{(i)}_{60,2013}\right)|\mathcal{F}_{2023}\right],$$

where ${}_T\hat{p}^{(i)}_{60,2013}$ is the fixed survival rate agreed upon at the beginning of the contract and $\mathcal{F}_{2023}$ entails all the mortality information up to the beginning of year 2023. Depending on the value of T, $v_s(T)$ can be rewritten as

$$v_s(T) = \begin{cases} 1.03^{10-T}\left({}_Tp^{(i)}_{60,2013} - {}_T\hat{p}^{(i)}_{60,2013}\right) & T \le 10, \\ 1.03^{10-T}\left({}_{10}p^{(i)}_{60,2013} \times E\left[{}_{T-10}p^{(i)}_{70,2023}|\mathcal{F}_{2023}\right] - {}_T\hat{p}^{(i)}_{60,2013}\right) & T > 10. \end{cases}$$

Assuming no death count risk, the value of the hedger's annuity liability assessed at the beginning of 2023 is

$$\begin{aligned} v_l &= E\left[0.25\sum_{i=1}^{4}\sum_{j=0}^{19} 1.03^{-j} \times {}_{10}p^{(i)}_{60,2013} \times {}_jp^{(i)}_{70,2023}|\mathcal{F}_{2023}\right] \\ &= 0.25\sum_{i=1}^{4}\sum_{j=0}^{19} 1.03^{-j} \times {}_{10}p^{(i)}_{60,2013} \times E\left[{}_jp^{(i)}_{70,2023}|\mathcal{F}_{2023}\right]. \end{aligned}$$

Suppose that the hedger uses a basket of n S-forwards to hedge the longevity risk. The maturity and notional amount of the kth S-forward are denoted by T_k and x_k, respectively. The hedge effectiveness is measured by the variance reduction of the hedger's liability in 10 years and can be expressed as follows:

$$1 - \frac{\mathrm{var}[v_l + \sum_{k=1}^{n} x_k v_s(T_k)]}{\mathrm{var}[v_l]}.$$

The hedging objective is to find the set of notional amounts, $\{x_1, x_2, \ldots, x_n\}$, that maximizes the variance reduction.

Let $X = [x_1, x_2, \ldots, x_n]'$, $V_s = [V_s(T_1), \ldots, V_s(T_n)]'$ and $V_{ls} = [\mathrm{cov}(v_l, v_s(T_1)), \mathrm{cov}(v_l, v_s(T_2)), \ldots, \mathrm{cov}(v_l, v_s(T_n))]'$, where the prime symbol denotes the transpose of a vector or matrix. Denote the variance–covariance matrix of V_s by $\Sigma(V_s)$. As shown in Appendix D, the notional amount vector that maximizes the hedge effectiveness is

$$X = -(\Sigma(V_s))^{-1} V_{ls}.$$

6.2. Simulating mortality path and mortality forecast

To calculate the optimal notional amount and maximum hedge effectiveness, we first use simulation to obtain sample variance/covariance as approximations

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

for var(v_l), $\Sigma(V_s)$ and V_{ls}. We begin with simulating one mortality paths over the 10-year hedge horizon. Let $t = 2012$.

1. Simulate a value for s_{t+1} using a Bernoulli random number generator with the success rate equal to $P_{11} \times (\Omega_{2012|2012})_1 + P_{21} \times (\Omega_{2012|2012})_2$.
2. If $s_{t+1} = 1$, then mortality dependence is in regime 1 at time $t+1$. Simulate a four-dimensional vector of uniform random numbers using the regime 1 R-vine copula. If $s_{t+1} = 0$, then mortality dependence is in regime 2 at time $t+1$. Simulate a four-dimensional vector of independent uniform random numbers.
3. Transform the uniform random number vector to the simulated standardized residual vector of the GARCH model using the inverse CDF of the residual distributions, which are normal for the first and fourth entries of the vector and Student's t for the second and third entries.
4. Calculate the volatility of the GARCH model at time $t+1$ and determine the simulated $k_{t+1}^{(i)}$.
5. Calculate the simulated $m_{x,t+1}^{(i)}$.
6. Repeat step 1–5 sequentially for $t = 2013, 2014, \ldots, 2022$, with the success rate of the Bernoulli random number generator changed to $P_{s_t,1}$ which is the transition probability from regime s_t to regime 1.

If the annuity provider believes that there will be no turbulence in the next 10 years and mortality dependence will stay in the high-dependence regime, we simply replace the success rate of the Bernoulli random number generator by 1 and keep the remaining steps unchanged. If the annuity provider expects significant health care reforms in the next 10 years and mortality dependence is likely to stay in the independent regime, we replace the success rate of the Bernoulli random number generator by 0 and again keep the remaining steps unchanged.

We repeat the above procedures to generate 5000 mortality paths over the 10-year horizon. At the beginning of 2023, we evaluate the annuity value v_l and S-forward value $v_s(T_k)$ based on each simulated mortality path. If $T_k \leq 10$, $v_s(T_k)$ can be easily calculated. However, if $T_k \geq 10$, then the evaluation of $v_s(T_k)$ involves an expectation, the calculation of which requires another set of simulated mortality paths for 2023 to $2012 + T_k$. Therefore, nested simulation arises. Due to limited computational power, we simulate a set of 1000 mortality paths for 2023–2042 for each mortality path simulated for 2013–2022. In total, 5000×1000 mortality paths need to be simulated for obtaining 5000 simulations of $v_s(T_k)$ and calculating the sample variance of $v_s(T_k)$ for $T_k \geq 10$. The liability value v_l is evaluated in a similar approach.

6.3. Hedging results

We suppose that the hedger uses S-forwards written on the survival rates of population (2) for the hedge. The reasons are twofold: first, simulation results

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

TABLE 10

MAXIMUM ACHIEVABLE HEDGE EFFECTIVENESS BASED ON DIFFERENT BELIEFS OF FUTURE MORTALITY DEPENDENCE EVOLUTION.

S-forward maturity	Variance reduction		
$[T_1, T_2, \ldots, T_n]$	Case 1	Case 2	Case 3
5	0.3680	0.3613	0.1941
10	0.7056	0.6373	0.3520
15	0.8298	0.7334	0.4091
20	0.8321	0.7326	0.4109
25	0.8226	0.7250	0.4085
30	0.8008	0.7157	0.4058
[10, 15]	0.8340	0.7354	0.4112
[15, 20]	0.8338	0.7354	0.4115
[10, 15, 20]	0.8343	0.7357	0.4119

suggest that policies issued to population (2) contribute most to the volatility of the annuity liability; secondly, population (2) has the highest overall mortality dependence with other populations.

We examine the hedge effectiveness based on the following three different beliefs of future mortality dependence evolution:

- Case 1: The mortality dependence will stay in regime 1 for the next 10 years.
- Case 2: The mortality dependence will transition between regimes 1 and 2 according to the regime-switching model.
- Case 3: The mortality dependence will stay in regime 2 for the next 10 years.

The three cases exhibit increasingly lower mortality dependence amongst the four populations. Table 10 summarizes the maximum achievable hedge effectiveness using S-forwards of various maturities under the three cases.

It is obvious that the maximum achievable variance reduction is significantly different among the three cases. Since Case 3 assumes complete independence, the S-forward written on the survival rate of population (2) is only effective in hedging the liability volatility contributed by policies issued to population (2). In contrast, Case 1 assumes high mortality dependence over the next 10 years. As a result, the S-forward is also effective in hedging the liability volatility contributed by policies issued to other populations, hence providing higher variance reduction. Case 2 allows transitions between high dependence and independence, thereby yielding variance reduction between those resulting from Cases 1 and 3.

In addition, we find that a single S-forward with a maturity of 15 years and longer yields much higher variance reduction than that with a shorter maturity in all three cases. However, when the S-forward maturity exceeds 15 years, variance reduction only improves marginally with maturity. We also consider a combination of S-forwards with different maturities, but the

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

increase of variance reduction from the additional S-forwards is insignificant. Given similar level of variance reduction, a simpler hedge strategy with shorter maturity should be preferred due to its lower transaction cost and higher liquidity. Therefore, a 15-year S-forward appears to be the best choice among the strategies considered in Table 10.

7. A HIGHER-DIMENSIONAL EXAMPLE

We use a four-dimensional example in previous sections for its easy illustration of how R-vine copula and regime-switching R-vine copula work. It is also easier to explain how the rolling window analysis is performed and to match the occurrence of regime switching with historical events for a small number of time series. The advantage of vine copula is obvious despite that the dimension of the above example is not high. We summarize the advantages of the R-vine copula as follows:

- Flexibility: Allowing an arbitrary choice of copula instead of forcing the same copula for each pair of variables enables us to construct a R-vine copula that best matches the true dependence. This can be verified by comparing the loglikelihoods of the R-vine and standard multivariate copulas in Table 6. The loglikelihood of R-vine is the second highest, only lower than that of the Student's t copula by 0.2, while the R-vine copula has two less parameters than the Student's t copula. We should also note that the selection of bivariate copulas for R-vine uses BIC as criterion. If we change the criterion to maximizing loglikelihood of each bivariate copula, the loglikelihood of the R-vine copula can attain 64.86.
- Less parameters: Since most dependence is captured at the lower levels of a R-vine as a result of the sequential selection procedure, we expect some independent bivariate copulas at the higher levels. Consequently, the number of parameters required for R-vine copulas should be smaller than that for multivariate Gaussian/t copulas.
- Computational ease: The selection and estimation of R-vine use a sequential method, thereby avoiding optimizing all the parameters in one step. This feature is particularly useful in high-dimensional problems for which one-stage optimization can be computationally challenging.

The methodologies developed in previous sections are readily applicable for a higher-dimensional example. However, the rolling window analysis is more complex since we need to identify the number of regimes and bivariate copulas associated with each edge in each regime manually. Let us consider an eight-dimensional example which includes both male and female populations of the four countries under consideration. We use superscript (i) for $i = 5, 6, 7, 8$, to represent the female populations of Italy, France, England and Wales, and Switzerland, respectively.

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

We follow the same procedures described in Sections 2 and 3 to decompose mortality data into age, period and cohort effects to select and estimate the marginal ARIMA-GARCH model for the period effect $k_t^{(i)}$, and to obtain the standardized residuals from fitting the ARIMA-GARCH model to $k_t^{(i)}$. We then estimate an R-vine copula for the eight residual series.

For simplicity, we assume independence for Trees 3 and above; hence we truncate the R-vine copula at Tree 2. It is natural to simplify an R-vine copula by assuming independence at high level trees, since most dependence is captured at the lower level trees due to the sequential estimation method. A more stringent method for determining the truncation level is to apply the statistical model selection techniques proposed in Brechmann et al. (2012). In comparison, much more deliberation is required when deciding which parameters should be reduced for standard multivariate Gaussian/Student's t copulas. We also confine the selection of bivariate copulas to independent, Gaussian, Student's t, Gumbel and rotated Gumbels. Table 11 summarizes the tree structure and parameter estimates of the selected R-vine. We observe that there is always an edge in tree 1 connecting the male and female populations of the same country and the Kendall's taus for these edges are high. For example, nodes 3 and 7 represent the male and female populations of England and Wales, respectively. These two nodes are connected by the edge {3,7}. The selected bivariate copula for this edge is Gaussian with a Kendall's tau of 0.81. The high dependence between male and female populations of the same country is reasonable since the two populations share the same health care system, culture, diet and other factors that impact mortality. Since the first tree captures the strongest dependence, it is expected that these edges appear in the first tree.

We further consider a regime-switching R-vine for the eight-dimensional example. We perform rolling window analysis for each pair of variables in Tree 1 to identify copulas associated with each regime. The rolling window analysis suggests the following switches of regimes:

- the best-fitting copula for edge $\{2, 3\}$ switches from Gumbel in turbulent periods to Guassian in peaceful periods;
- the best-fitting copula for edge $\{6, 2\}$ switches from Gumbel in turbulent periods to Gaussian in peaceful periods;
- the best-fitting copula for edge $\{8, 6\}$ switches from Survival Gumbel in turbulent periods to Gumbel in peaceful periods.

In addition, we find that the copula parameters associated with the second tree in the low-dependence regime are insignificant. Therefore, we assume independent copulas for Tree 2 and above in the low-dependence regime. The estimated regime-switching R-vine copula is shown in Table 12. Regime 1 that represents the peaceful periods again exhibits higher dependence than regime 2 that represents the turbulent periods. The regime-switching R-vine copula also yields a lower BIC than the single-regime R-vine shown in Table 11, thereby indicating better goodness of fit.

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

TABLE 11

ESTIMATED R-VINE COPULA FOR THE EIGHT-DIMENSIONAL EXAMPLE.

					Tail dependence	
Tree	Edge	Copula	Parameter	tau	Upper	Lower
1	3,7	Gaussian	0.96 (0.01)	0.81	–	–
	2,3	Gumbel	1.41 (0.11)	0.29	0.37	–
	5,1	Gaussian	0.93 (0.01)	0.75	–	–
	6,2	Gumbel	2.99 (0.24)	0.67	0.74	–
	6,5	Gaussian	0.62 (0.05)	0.43	–	–
	8,4	Gaussian	0.81 (0.03)	0.60	–	–
	8,6	Gumbel	1.67 (0.13)	0.40	0.49	–
2	2,7;3	Independent	–	0.00	–	–
	6,3;2	Independent	–	0.00	–	–
	6,1;5	Independent	–	0.00	–	–
	8,2;6	Independent	–	0.00	–	–
	8,5;6	Survival Gumbel	1.28 (0.09)	0.22	–	0.28
	6,4;8	Gaussian	0.25 (0.09)	0.16	–	–
3 and above		Independent				
Loglikelihood: 444.16, AIC: −870.32, BIC: −845.85						

8. CONCLUSIONS

In this paper, we aim to develop copulas to capture the mortality dependence of multiple populations and to understand the historical evolution of mortality dependence. We first apply a simple APC model to decompose the historical mortality rates of each population into age, period and cohort effects. We further model the period effects of each population by ARIMA-GARCH and obtain standardized residuals. Finally, we construct copulas to model the dependence between the standardized residuals.

In particular, we introduce R-vine copula to model mortality dependence. Different from standard multivariate copulas, R-vine copula is built from bivariate copulas. An R-vine copula is very flexible in modelling high-dimensional dependence because it allows an arbitrary choice of bivariate copula for each pair of variables. The R-vine copula fits our four-dimensional data significantly better than commonly used multivariate copulas such as Gaussian copula, *t* copula and Gumbel copula. We further consider a regime-switching R-vine copula. We begin with a rolling window analysis to identify copulas associated with each regime. We then estimate the regime-switching model with an EM algorithm. Our estimated model shows that mortality dependence switches between two regimes, one with high dependence and the other with low dependence.

It is interesting to note that the peaceful periods without major events show high probability in the high-dependence regime. The years in which extreme mortality-deteriorating events such as epidemics break out also exhibit

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

TABLE 12

ESTIMATED REGIME-SWITCHING R-VINE COPULA FOR THE EIGHT-DIMENSIONAL EXAMPLE.

Regime 1						
					Tail dependence	
Tree	Edge	Copula	Parameter	tau	Upper	Lower
1	3,7	Gaussia	0.97 (0.01)	0.84	–	–
	2,3	Gaussian	0.67 (0.06)	0.46	–	–
	5,1	Gaussia	0.95 (0.01)	0.79	–	–
	6,2	Gaussia	0.94 (0.01)	0.77	–	–
	6,5	Gaussia	0.76 (0.04)	0.55	–	–
	8,4	Gaussia	0.86 (0.02)	0.67	–	–
	8,6	Survival Gumbel	1.98 (0.20)	0.49	–	0.58
2	2,7;3	Independent	–	0.00	–	
	6,3;2	Independent	–	0.00	–	–
	6,1;5	Independent	–	0.00	–	–
	8,2;6	Independent	–	0.00	–	–
	8,5;6	Survival Gumbel	1.32 (0.13)	0.24	–	0.31
	6,4;8	Gaussian	0.28 (0.11)	0.18	–	–
3 and above		Independent				
Regime 2						
				Kendall's	Tail dependence	
Tree	Edge	Copula	Parameter	tau	Upper	Lower
1	3,7	Gaussian	0.93 (0.02)	0.77	–	–
	2,3	Independent	–	0.00	–	–
	5,1	Gaussian	0.88 (0.03)	0.69	–	–
	6,2	Gumbel	2.03 (0.27)	0.51	0.59	–
	6,5	Gaussian	0.38 (0.13)	0.25	–	–
	8,4	Gaussian	0.69 (0.07)	0.48	–	–
	8,6	Gumbel	1.49 (0.18)	0.33	0.41	–
2 and above		Independent				
$p_{11} = 0.93(0.04)$, $p_{22} = 0.86(0.06)$						
Loglikelihood: 467.28, AIC: −900.56, BIC: −854.46						

a high chance of being in the high-dependence regime because the break-out often occurs in the same year for multiple countries. However, the recovery from such events often takes different paths in different countries, thereby leading to a period of low dependence following these events. The period of health reforms in European countries also shows low probability in the high-dependence regime because the reforms of different countries began in different years and introduced different measures to improve health care.

It is important to note that a drawback of our approach is that common trend among multi-population mortalities is not considered. Our approach assumes a unit root for each period effect series and subsequently uses differenced period effect series for further analysis. As a result, valuable information

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

contained in the original series, an example of which is common trend or cointegrating relations, may be lost. Multivariate time series are cointegrated if all of the series are integrated of order d and a linear combination of this collection is integrated of lower order. We ran Johansen's test (Johansen, 1991) for the four male period effect series and detected two cointegrating relations at 5% significance level. The occurrence of simultaneous mortality jumps and the high dependence in the peaceful periods could be a manifestation of the cointegrating relations present the data. Ignoring these cointegrating relations, our approach may not fully capture the mortality dependence. For more detailed discussion of cointegration, we refer readers to Engle and Granger (1987) and Johansen (1991). In the future research, we plan to quantitatively compare the performance of copula approach and VECM, which explicitly incorporates cointegration relations, in modelling mortality dependence.

Our model can be easily adapted to assess the effectiveness of standardized longevity hedge. Based on the estimated model, we allow the hedger to express different opinions on the evolution of future mortality dependence and project mortality rates with such opinions incorporated. We use a case study to illustrate the difference in hedge effectiveness caused by these opinions. Our model can also be applied to the risk management of multinational insurance/annuity providers. The mortality dependence between policyholders from different population groups may significantly impact the tail distribution of the insurance/annuity liability or the extent of natural hedge between insurance and annuity liabilities, which in turn determines the capital requirement of the provider.

In this paper, the transition between two regimes is abrupt. However, recovery from extreme mortality deteriorating events or health reforms may last a number of years. Transitioning from the low-dependence regime to the high-dependence regime may take small steps and a smooth transition may be more appropriate. Therefore, a possible future research direction is to consider a two-regime copula with smooth transitions. Another possible direction is to examine whether the R-vine tree structure also evolves over time and to achieve higher model flexibility by considering different tree structures in different regimes. In addition, we are also interested in developing a pricing method for longevity hedge based on the regime-switching copula.

NOTE

1. Downloaded from Human Mortality Database website http://www.mortality.org/ on February 12, 2017.

REFERENCES

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

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**, 245–268.

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

BEDFORD, T. and COOKE, R.M. (2002) Vines—a new graphical model for dependent random variables. *The Annals of Statistics*, **30**, 1031–1068.
BELTRÁN-SÁNCHEZ, H., CRIMMINS, E.M. and FINCH, C.E. (2012) Early cohort mortality predicts the cohort rate of aging: an historical analysis. *Journal of developmental origins of health and disease*, **3**, 380–386.
BRECHMANN, E.C. and CZADO, C. (2013) Risk management with high-dimensional vine copulas: an analysis of the euro stoxx 50. *Statistics & Risk Modeling*, **30**, 307–342.
BRECHMANN, E.C., CZADO, C. and AAS, K. (2012) Truncated regular vines in high dimensions with application to financial data. *Canadian Journal of Statistics*, **40**, 68–85.
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**, 687–718.
CAIRNS, A.J.G., BLAKE, D., DOWD, K., COUGHLAN, G.D. and KHALAF-ALLAH, M. (2011) Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin*, **41**, 29–59.
CHAI, C.M.H., KUEN SIU, T. and ZHOU, X. (2013) A double-exponential garch model for stochastic mortality. *European Actuarial Journal*, **3**, 385–406.
CHEN, H., MACMINN, R. and SUN, T. (2015) Multi-population mortality models: a factor copula approach. *Insurance: Mathematics and Economics*, **63**, 135–146.
CHEN, H., MACMINN, R.D. and SUN, T. (2017) Mortality dependence and longevity bond pricing: A dynamic factor copula mortality model with the gas structure. *Journal of Risk and Insurance*, **84**, 393–415.
CHOLLETE, L., HEINEN, A. and VALDESOGO, A. (2009) Modeling international financial returns with a multivariate regime-switching copula. *Journal of Financial Econometrics*, **7**, 437–480.
CHULIÁ, H., GUILLÉN, M. and URIBE, J.M. (2016) Modeling longevity risk with generalized dynamic factor models and vine-copulae. *ASTIN Bulletin*, **46**, 165–190.
DIßMANN, J., BRECHMANN, E.C., CZADO, C. and KUROWICKA, D. (2013) Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*, **59**, 52–69.
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**, 334–356.
ENGLE, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, **50**, 987–1007.
ENGLE, R.F. and GRANGER, C.W.J. (1987) Co-integration and error correction: representation, estimation, and testing. *Econometrica*, **55**, 251–276.
FINK, H., KLIMOVA, Y., CZADO, C. and STÖBER, J. (2017) Regime switching vine copula models for global equity and volatility indices. *Econometrics*, **5**, 3. doi:10.3390/econometrics5010003.
GAO, Q. and HU, C. (2009) Dynamic mortality factor model with conditional heteroskedasticity. *Insurance: Mathematics and Economics*, **45**, 410-423.
HAMILTON, J.D. (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, **57**, 357–384.
JANSSEN, F. (2005) Cohort patterns in mortality trends among the elderly in seven european countries, 1950–99. *International Journal of Epidemiology*, **34**, 1149–1159.
JARNER, S.F. and KRYGER, E.M. (2011) Modelling adult mortality in small populations: the saint model. *ASTIN Bulletin*, **41**, 377–418.
JOE, H. (1996) *Families of m-variate distributions with given margins and* $m(m-1)/2$ *bivariate dependence parameters*. Lecture Notes–Monograph Series, pp. 120–141. Hayward, CA: Institute of Mathematical Statistics.
JOHANSEN, S. (1991) Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, **59**, 1551–1580.
JONDEAU, E. and ROCKINGER, M. (2006) The copula-garch model of conditional dependencies: an international stock market application. *Journal of International Money and Finance*, **25**, 827–853.
KIM, C.-J. (1994) Dynamic linear models with markov-switching. *Journal of Econometrics*, **60**, 1–22.
KIM, C.-J. and NELSON, C.R. (1999) *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*. Cambridge, Massachusetts: MIT Press.
LEE, R.D. and CARTER, L.R. (1992) Modeling and forecasting u. s. mortality. *Journal of the American Statistical Association*, **87**, 659–671.

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

LI, J.S.-H. and HARDY, M.R. (2011) Measuring basis risk in longevity hedges. *North American Actuarial Journal*, **15**, 177–200.
LI, J.S.-H. ZHOU, R. and HARDY, M. (2015) A step-by-step guide to building two-population stochastic mortality models. *Insurance: Mathematics and Economics*, **63**, 121–134.
LI, N. and LEE, R. (2005) Coherent mortality forecasts for a group of populations: an extension of the lee-carter method. *Demography*, **42**, 575–594.
LIN, T., WANG, C.-W. and TSAI, C.C.-L. (2015) Age-specific copula-ar-garch mortality models. *Insurance: Mathematics and Economics*, **61**, 110–124.
NIKOLOULOPOULOS, A.K., JOE, H. and LI, H. (2012) Vine copulas with asymmetric tail dependence and applications to financial return data. *Computational Statistics & Data Analysis*, **56**, 3659–3673.
OSMOND, C. (1985) Using age, period and cohort models to estimate future mortality rates. *International Journal of Epidemiology*, **14**, 124–129.
SKLAR, A. (1959) Fonctions de répartition àn dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris*, **8**, 229–231.
STÖBER, J. and CZADO, C. (2014) Regime switches in the dependence structure of multidimensional financial data. *Computational Statistics and Data Analysis*, **76**, 672–686.
WANG, C.-W., YANG, S.S. and HUANG, H.-C. (2015) Modeling multi-country mortality dependence and its application in pricing survivor index swaps—a dynamic copula approach. *Insurance: Mathematics and Economics*, **63**, 30–39.
WANG, Z. and LI, J.S.-H. (2016) A dcc-garch multi-population mortality model and its applications to pricing catastrophic mortality bonds. *Finance Research Letters*, **16**, 103–111.
WILMOTH, J.R. (2005) On the relationship between period and cohort mortality. *Demographic Research*, **S4**, 231–280.
YANG, S.S. and WANG, C.-W. (2013) Pricing and securitization of multi-country longevity risk with mortality dependence. *Insurance: Mathematics and Economics*, **52**, 157–169.
ZHOU, R., LI, J.S.-H. and TAN, K.S. (2013) Pricing standardized mortality securitizations: a two-population model with transitory jump effects. *Journal of Risk and Insurance*, **80**, 733–774.
ZHOU, R., WANG, Y., KAUFHOLD, K., LI, J.S.-H. and TAN, K.S. (2014) Modeling period effects in multi-population mortality models: applications to Solvency II. *North American Actuarial Journal*, **18**, 150–167.

RUI ZHOU
Department of Economics
University of Melbourne
Parkville VIC 3010,
Australia
E-mail: rui.zhou@unimelb.edu.au

APPENDIX A. COMMONLY USED COPULAS

Let $U_i = F_i(y_i)$, for $i = 1, 2, \ldots, n$. A copula CDF can be defined as

$$C(u_1, u_2, \ldots, u_n) = \Pr[U_1 \leq u_1, U_2 \leq u_2, \ldots, U_n \leq u_n].$$

In this paper, we consider the following commonly used copulas:

(i) The Gaussian copula with correlation matrix R, $R \in [-1, 1]^{n \times n}$ can be written as

$$C^{\text{Gauss}}(u_1, u_2, \ldots, u_n) = \Phi_R(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \ldots, \Phi^{-1}(u_n)),$$

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

TABLE A1

KENDALL'S TAU AND TAIL DEPENDENCE FORMULA FOR COMMONLY USED COPULAS.

		Tail dependence	
Copula	Kendall's tau	Upper	Lower
Gaussian	$\frac{2}{\pi}\arcsin(\rho)$	0	0
t	$\frac{2}{\pi}\arcsin(\rho)$	$2t_{v+1}\left(-\sqrt{v+1}\sqrt{\frac{1-\rho}{1+\rho}}\right)$	$2t_{v+1}\left(-\sqrt{v+1}\sqrt{\frac{1-\rho}{1+\rho}}\right)$
Clayton	$\frac{\theta}{\theta+2}$	$2^{-\frac{1}{\theta}}$	0
Gumbel	$1-\frac{1}{\theta}$	0	$2-2^{\frac{1}{\theta}}$
Frank	$1-\frac{4}{\theta}+4\frac{\int_0^\theta \frac{x/\theta}{\exp(x)-1}dx}{\theta}$	0	0

where Φ_R is the joint CDF of an n-dimensional standard multivariate normal distribution with correlation matrix R and Φ^{-1} is the inverse CDF of a standard normal distribution.

(ii) The Student's t copula with correlation matrix R and degrees of freedom v can be written as

$$C^T(u_1,u_2,\ldots,u_n)=T_{R,v}(T_v^{-1}(u_1),T_v^{-1}(u_2),\ldots,T_v^{-1}(u_n)),$$

where $T_{R,v}$ is the joint CDF of an n-dimensional standard multivariate Student's t distribution with v degrees of freedom and T_v^{-1} is the inverse CDF of a standard Student's t distribution with v degrees of freedom.

(iii) The Gumbel copula with parameter θ, $\theta\in[1,\infty)$ can be written as

$$C^G(u_1,u_2,\ldots,u_n)=\exp\left[-\left(\sum_{i=1}^{n}(-\ln u_i)^\theta\right)^{1/\theta}\right].$$

(iv) The Frank copula with parameter θ, $\theta\in\mathbb{R}\setminus\{0\}$ can be written as

$$C^F(u_1,u_2,\ldots,u_n)=-\frac{1}{\theta}\ln\left[1+\frac{\prod_{i=1}^{n}(\exp(-\theta u_i)-1)}{\exp(-\theta)-1}\right].$$

(v) The Clayton copula with parameter θ, $\theta>0$ can be written as

$$C^C(u_1,u_2,\ldots,u_n)=\left[\max\left\{\sum_{i=1}^{n}u_i^{-\theta}-N+1;0\right\}\right]^{-1/\theta}.$$

When selecting bivariate copulas for pair variables, we also consider rotated version of Clayton and Gumbel copulas. When rotating them by 90°, 180° and 270°, the copula CDFs are given as follows:

$$\begin{aligned}C_{90}(u_1,u_2)&=u_2-C(1-u_1,u_2),\\ C_{180}(u_1,u_2)&=u_1+u_2-1+C(1-u_1,1-u_2),\\ C_{270}(u_1,u_2)&=u_1-C(u_1,1-u_2).\end{aligned}$$

Table A1 summarizes the formulas used for calculating Kendall's tau and upper and lower dependences of the bivariate copulas considered in this paper.

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

APPENDIX B. PROOFS OF EQUATIONS (5.2) AND (5.3)

Proof.

$$
\begin{aligned}
(\boldsymbol{\Omega}_{t|t-1})_k &= \Pr(s_t = k|\tilde{\boldsymbol{u}}_{t-1}, \boldsymbol{\theta}) \\
&= \sum_{j=1}^{K} \Pr(s_t = k, s_{t-1} = j|\tilde{\boldsymbol{u}}_{t-1}, \boldsymbol{\theta}) \\
&= \sum_{j=1}^{K} \Pr(s_t = k|s_{t-1} = j, \tilde{\boldsymbol{u}}_{t-1}, \boldsymbol{\theta}) \Pr(s_{t-1} = j|\tilde{\boldsymbol{u}}_{t-1}, \boldsymbol{\theta}) \\
&= \sum_{j=1}^{K} \Pr(s_t = k|s_{t-1} = j) \Pr(s_{t-1} = j|\tilde{\boldsymbol{u}}_{t-1}, \boldsymbol{\theta}) \\
&= \sum_{j=1}^{K} P_{jk} (\boldsymbol{\Omega}_{t-1|t-1})_j \\
(\boldsymbol{\Omega}_{t|t})_k &= \Pr(s_t = k|\tilde{\boldsymbol{u}}_t, \boldsymbol{\theta}) \\
&= \frac{f(s_t = k, \boldsymbol{u}_t|\tilde{\boldsymbol{u}}_{t-1}, \boldsymbol{\theta})}{f(\boldsymbol{u}_t|\tilde{\boldsymbol{u}}_{t-1}, \boldsymbol{\theta})} \\
&= \frac{f(s_t = k, \boldsymbol{u}_t|\tilde{\boldsymbol{u}}_{t-1}, \boldsymbol{\theta})}{\sum_{j=1}^{K} f(s_t = j, \boldsymbol{u}_t|\tilde{\boldsymbol{u}}_{t-1}, \boldsymbol{\theta})} \\
&= \frac{f(\boldsymbol{u}_t|s_t = k, \tilde{\boldsymbol{u}}_{t-1}, \boldsymbol{\theta}) \Pr(s_t = k|\tilde{\boldsymbol{u}}_{t-1}, \boldsymbol{\theta})}{\sum_{j=1}^{K} f(\boldsymbol{u}_t|s_t = j, \tilde{\boldsymbol{u}}_{t-1}, \boldsymbol{\theta}) \Pr(s_t = j|\tilde{\boldsymbol{u}}_{t-1}, \boldsymbol{\theta})} \\
&= \frac{f(\boldsymbol{u}_t|s_t = k, \boldsymbol{\theta}_k)(\boldsymbol{\Omega}_{t|t-1})_k}{\sum_{j=1}^{K} f(\boldsymbol{u}_t|s_t = j, \boldsymbol{\theta}_j)(\boldsymbol{\Omega}_{t|t-1})_j}.
\end{aligned}
$$

□

APPENDIX C. STEPWISE EM ALGORITHM (STÖBER AND CZADO, 2014)

Let $\boldsymbol{\theta}^l$ be the estimated parameter set from the lth iteration. This algorithm maximizes the expected pseudo loglikelihood function which is defined as follows:

$$
\begin{aligned}
Q(\boldsymbol{\theta}^{l+1}; \tilde{\boldsymbol{u}}_T, \boldsymbol{\theta}^l) &= \int_{\tilde{\boldsymbol{s}}_T} \log\left(f(\tilde{\boldsymbol{u}}_T, \tilde{\boldsymbol{s}}_T|\boldsymbol{\theta}^{l+1})\right) \Pr(\tilde{\boldsymbol{s}}_T|\tilde{\boldsymbol{u}}_T, \boldsymbol{\theta}^l) \\
&= \sum_{t=1}^{T} \int_{\tilde{\boldsymbol{s}}_T} \log\left(f(\boldsymbol{u}_t|s_t, \boldsymbol{\theta}_1^{l+1}, \ldots, \boldsymbol{\theta}_K^{l+1})\right) \Pr(\tilde{\boldsymbol{s}}_T|\tilde{\boldsymbol{u}}_T, \boldsymbol{\theta}^l) \\
&\quad + \int_{\tilde{\boldsymbol{s}}_T} \left[\sum_{t=1}^{T} \log\left(\Pr\left(s_t|s_{t-1}, \boldsymbol{\theta}_{mc}^{l+1}\right)\right)\right] \Pr(\tilde{\boldsymbol{s}}_T|\tilde{\boldsymbol{u}}_T, \boldsymbol{\theta}^l) \qquad \text{(C.1)}
\end{aligned}
$$

where $\tilde{\boldsymbol{s}}_t = (s_1, \ldots, s_t)$ and $\int_{\tilde{\boldsymbol{s}}_T} g(\tilde{\boldsymbol{s}}_T) = \sum_{s_1=1}^{K} \ldots \sum_{s_T=1}^{K} g(\tilde{\boldsymbol{s}}_T)$.

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

The algorithm iterates the expectation and maximization steps as follows:

(a) Expectation: calculate the smoothed probability vector $\boldsymbol{\Omega}_{t|T}$ given the parameter set $\boldsymbol{\theta}^l$
(b) Maximization: maximize $Q(\boldsymbol{\theta}^{l+1};\tilde{\boldsymbol{u}}_T, \boldsymbol{\theta}^l)$ with respect to $\boldsymbol{\theta}^{l+1}$.

The first term of the RHS of equation (C.1) depends only on the copula parameters $\boldsymbol{\theta}_1^{l+1}, \ldots, \boldsymbol{\theta}_K^{l+1}$, while the second term only depends on the transition probability vector $\boldsymbol{\theta}_{\text{mc}}^{l+1}$. Kim and Nelson (1999) showed that the transition probabilities maximizing the pseudo loglikelihood are given by

$$P_{ij}^{l+1} = \frac{\sum_{t=1}^{T} \Pr(s_t = j, s_{t-1} = i|\tilde{\boldsymbol{u}}_T, \boldsymbol{\theta}^l)}{\sum_{t=1}^{T} \Pr(s_{t-1} = i|\tilde{\boldsymbol{u}}_T, \boldsymbol{\theta}^l)}.$$

However, the copula parameters maximizing the first term of the expected pseudo loglikelihood cannot be derived analytically and hence need to be found using numerical methods. For a high-dimensional problem, the joint estimation of the copula parameters is computationally challenging. As a result, Stöber and Czado (2014) recommended the stepwise estimation procedure proposed by Aas et al. (2009) with the weight of the year t observation set to $\Pr(s_t|\tilde{\boldsymbol{u}}_T, \boldsymbol{\theta}^l)$.

APPENDIX D. MAXIMIZING HEDGE EFFECTIVENESS

The maximization of hedge effectiveness can be written as

$$\begin{aligned}
&\max_X \; 1 - \frac{\text{var}[v_l + X'V_s]}{\text{var}[v_l]} \\
&= \min_X \text{var}(v_l + X'V_s) \\
&= \min_X \text{var}(v_l) + \text{var}(X'V_s) + 2\text{cov}(v_l, X'V_s) \\
&= \min_X \text{var}(v_l) + X'\Sigma(V_s)X + 2X'V_{ls}.
\end{aligned}$$

The vector X that maximizes the hedge effectiveness should satisfy the following:

$$\begin{aligned}
\frac{\partial}{\partial X}\text{var}(v_l + X'V_s) &= 0 \\
2X'\Sigma(V_s) + 2V_{ls}' &= 0 \\
X &= -(\Sigma(V_s))^{-1}V_{ls}.
\end{aligned}$$

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press

https://doi.org/10.1017/asb.2019.10 Published online by Cambridge University Press