ANALYZING MORTALITY BOND INDEXES VIA HIERARCHICAL FORECAST RECONCILIATION

BY

HAN LI AND QIHE TANG

ABSTRACT

In recent decades, there has been significant growth in the capital market for mortality- and longevity-linked bonds. Therefore, modeling and forecasting the mortality indexes underlying these bonds have crucial implications for risk management in life insurance companies. In this paper, we propose a hierarchical reconciliation approach to constructing probabilistic forecasts for mortality bond indexes. We apply this approach to analyzing the Swiss Re Kortis bond, which is the first "longevity trend bond" introduced in the market. We express the longevity divergence index associated with the bond's principal reduction factor (PRF) in a hierarchical setting. We first adopt time-series models to obtain forecasts on each hierarchical level, and then apply a minimum trace reconciliation approach to ensure coherence of forecasts across all levels. Based on the reconciled probabilistic forecasts of the longevity divergence index, we estimate the probability distribution of the PRF of the Kortis bond, and compare our results with those stated in Standard and Poor's report on pre-sale information. We also illustrate the strong performance of the approach by comparing the reconciled forecasts with unreconciled forecasts as well as those from the bottom-up approach and the optimal combination approach. Finally, we provide first insights on the interest spread of the Kortis bond throughout its risk period 2010–2016.

KEYWORDS

Forecast reconciliation, probabilistic forecast, time-series models, mortality modeling, the Kortis bond.

1. INTRODUCTION

Securitization, often used in the context of general insurance and loans or mortgage products, has also become an important risk mitigation tool for

Astin Bulletin 49(3), 823-846. doi:10.1017/asb.2019.19

the life insurance industry in recent decades (Cox *et al.*, 2000; Cowley and Cummins, 2005; MacMinn *et al.*, 2006; Cummins and Trainar, 2009). For example, mortality catastrophe bonds have been used by a number of insurance and reinsurance companies to transfer extreme mortality risks to the capital market. Most of the current mortality bonds have a principal reduction factor (PRF) calculated from a single mortality index. As the mortality index is often constructed in a complex manner, which involves age-specific mortality rates across multiple populations, accurate modeling and forecasting of the PRF become a main challenge to the pricing of such bonds. Since the PRF is generally a known function of age-country-specific mortality rates, combining the information from forecasts of these disaggregate series with forecasts of the PRF itself could lead to more accurate pricing of the mortality bond. In this paper, we propose a forecast reconciliation approach to achieve this.

In the literature, there have been a number of works on the design and pricing of mortality bonds (see, e.g., Cairns *et al.*, 2005, 2006; Lin and Cox, 2008; Chen and Cox, 2009; Bauer *et al.*, 2010; Deng *et al.*, 2012; Biagini *et al.*, 2013; Lin *et al.*, 2013; Bauer and Kramer, 2016; Braun, 2016; Chulia *et al.*, 2016; Chen *et al.*, 2017; Stupfler and Yang, 2018). These led to a rapid development in mortality modeling. In particular, contributions have been made in the areas of continuous-time stochastic models, multivariate time-series models, and copula models. However, to the best of our knowledge, most existing works look at either the bottom-level data series (i.e., age-country-specific mortality rates), or the top-level data series (i.e., the mortality bond index itself) alone when assessing the underlying risk.

In some other applications, it has been shown that forecast accuracy can be improved by taking into account available data at all levels (Athanasopoulos *et al.*, 2009; Hyndman *et al.*, 2014). Motivated by this, we utilize and combine information from both aggregate and disaggregate mortality rate series throughout the construction of the bond index. Note that if forecasts are produced for individual series independently, it is very unlikely that they will add up in the same hierarchical structure as the original data since aggregation constraints are not incorporated into the forecasting process. To address this issue, we adopt an optimal forecast reconciliation approach proposed by Wickramasuriya *et al.* (2018). Forecast reconciliation is a methodology by which forecasts on different levels are adjusted to ensure that certain aggregation constraints are fulfilled (Hyndman and Athanasopoulos, 2014). As the estimation of the PRF requires full probabilistic mortality forecasts, we extend the approach developed by Wickramasuriya *et al.* (2018) from point forecasts to interval forecasts based on the sampling algorithm proposed by Jeon *et al.* (2019).

In this paper, we conduct an empirical case study on the Swiss Re Kortis bond, and apply the proposed forecast reconciliation approach to modeling the "Longevity Divergence Index." According to Standard and Poor's (2010), the PRF of the Kortis bond depends on the divergence in mortality improvements between the UK population (England & Wales only) and the US population.

https://doi.org/10.1017/asb.2019.19 Published online by Cambridge University Press

It has attracted attention from both practitioners and academics. Hunt and Blake (2015) analyzed the design of the bond and proposed a co-integration time-series approach to modeling the mortality dynamics. Chen *et al.* (2017) proposed a two-factor copula model with a generalized autoregressive score structure to capture the mortality dependence of the two populations. In both papers, the authors conducted research only based on bottom-level age-specific mortality rates from the two countries. Our method can be easily distinguished from theirs by utilizing information on all levels to produce coherent forecasts for the longevity divergence index.

We contribute to the existing literature in threefold. First, we introduce a new approach to the modeling and forecasting of the mortality bond index, which utilizes all available information and guarantees coherence. As we do not attempt to jointly model age-specific mortality rates across different countries, our method is less affected by the "curse of dimensionality" compared to other approaches. Second, we are among the first to incorporate state-of-the-art forecast reconciliation techniques into the analysis of mortality-/longevity-linked securities, after Shang and Hyndman (2017) and Shang and Haberman (2017) who reconciled point forecasts of the regional mortality rates in Japan. In contrast to these studies, we are the first to perform probabilistic forecast reconciliation in mortality modeling to the best of our knowledge. The empirical results show that our method provides reliable probabilistic forecasts for the Kortis bond divergence index. Finally, our study offers first insights into changes in the distribution of the Kortis bond PRF over the risk period 2010–2016.

The rest of the paper is organized as follows. In Section 2, we provide a review of the Kortis longevity trend bond and describe the data and models used. Section 3 introduces the minimum trace forecast reconciliation approach and applies this approach to the hierarchical time series constructed in Section 2. In Section 4, we comment on the interest spread of the bond and discuss the changes in the distribution of PRF over time. Section 5 concludes.

2. Swiss Re Kortis bond

2.1. Background

Launched by Swiss Re Kortis Capital Ltd. in December 2010, the Kortis bond is promoted as the first "Longevity Trend Bond" in the market (for detailed information, see Standard and Poor's, 2010). The risk period is from 1st January 2009 to 31st December 2016 and the bond received a rating of "BB+" from Standard and Poor's.

The PRF of the Kortis bond is linked to a longevity divergence index measured based on the difference between the 8-year mortality improvement rate of the UK male population aged 75–85, and that of the US male population aged 55–65. During the risk period, the bond will be "triggered" if the longevity

https://doi.org/10.1017/asb.2019.19 Published online by Cambridge University Press

TABLE 1

CUMULATIVE AND 6-YEAR ANNUALIZED LOSS PROBABILITIES ESTIMATED BY THE RMS.

	Cumulative (%)	6-year annualized (%)
Attachment probability	5.31	0.88
Exhaustion probability	1.81	0.30
Expected loss	3.27	0.55

divergence index reaches 3.4%. Bondholders will lose their entire initial investments if the index reaches 3.9%. In exchange for the risk that the investors' principal might be reduced, the bond pays quarterly coupons at a rate of 5% above the 3-month London interbank offered rate (LIBOR).[1]

On behalf of Standard and Poor's, the Risk Management Solutions (RMS) performed the risk modeling and initial pricing of the Kortis bond. RMS applied an epidemiological modeling approach that incorporates information on causes of death mortality improvements, medical advancements, and likely future mortality drivers (Standard and Poor's, 2010). The modeling approach is quite different from traditional statistical mortality models (see more details in Blake *et al.*, 2013). Based on this approach, RMS calculated the distribution of the PRF. Table 1 shows the cumulative and 6-year annualized loss probabilities published in Standard and Poor's pre-sale report, page 12.[2]

2.2. The structure of the Kortis bond

Consider the UK male population aged 75–85 and the US male population aged 55–65. Let $m_t^j(x)$ denote male mortality rate at age x and time t for population j. According to Standard and Poor's (2010), the "Longevity Divergence Index" is constructed in the following steps:

- First, we calculate the average mortality improvement rate at time t over the last 8 years, for age x and population j:

$$\triangle m_t^j(x) = 1 - \left(\frac{m_t^j(x)}{m_{t-8}^j(x)} \right)^{1/8}. \tag{2.1}$$

- Second, we average the mortality improvement rates obtained in Equation (2.1) across ages x_1 to x_2 for each time t and population j:

$$\triangle m_t^j(x_1, x_2) = \frac{1}{x_2 - x_1 + 1} \sum_{x=x_1}^{x_2} \triangle m_t^j(x). \tag{2.2}$$

- Third, we construct the Longevity Divergence Index Value (LDIV) at time t as

$$\text{LDIV}_t = \triangle m_t^{\text{UK}}(75, 85) - \triangle m_t^{\text{US}}(55, 65). \tag{2.3}$$

https://doi.org/10.1017/asb.2019.19 Published online by Cambridge University Press

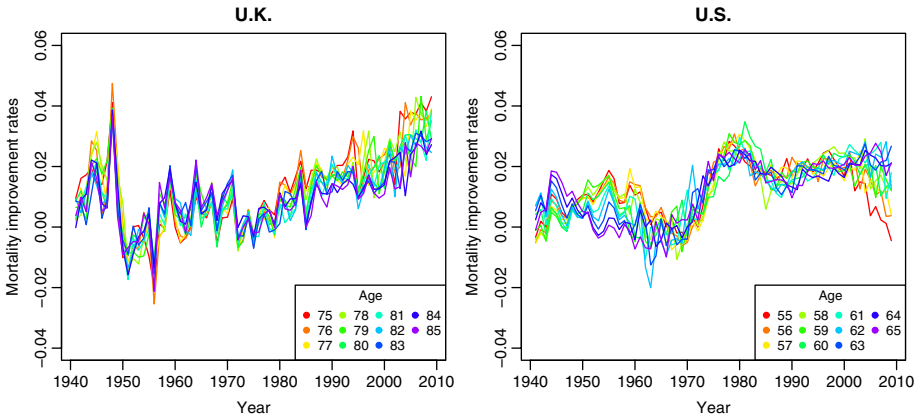

FIGURE 1: Historical annualized age-specific mortality improvement rates.

The PRF is then calculated as

$$\text{PRF} = \frac{[\text{LDIV}_t - 3.4\%]_+ - [\text{LDIV}_t - 3.9\%]_+}{3.9\% - 3.4\%}, \tag{2.4}$$

where 3.4% is referred to as the "point of attachment," and 3.9% is referred to as the "point of exhaustion."

2.3. Data

We collect historical mortality data from 1933 to 2009 for the UK males aged 75–85 and the US males aged 55–65. The deaths and exposures data are obtained from the Human Mortality Database (HMD).[3]

In Figure 1, we plot the historical annualized age-specific mortality improvement rates of the two populations. For both countries, the shape of mortality improvement is quite homogeneous within the selected age range. However, among the UK age groups, the mortality improvement rates for the "younger olds" (mid-to-late 70s) had a more substantial increase in the last two to three decades. In contrast, among the US age groups, even though the "younger olds"(mid-to late 50s) had higher mortality improvement rates in the 1950s and early 1960s, the improvement rates of these age groups seem to have experienced a significant decline since early 2000s.

In Figure 2, we then plot the historical annualized average mortality improvement rates $\triangle m_t^{\text{UK}}(75, 85)$ and $\triangle m_t^{\text{US}}(55, 65)$ for the UK and the US, respectively. We can see that for the UK, apart from the sudden drop around the early 1950s, the mortality improvement rate has kept rising for the last couple of decades. On the other hand, the upward trend in mortality improvement rate for the US seems to be less apparent.

Figure 3 shows the historical LDIV constructed based on the mortality improvement rates plotted in Figure 2. It shows no obvious overall upward

https://doi.org/10.1017/asb.2019.19 Published online by Cambridge University Press

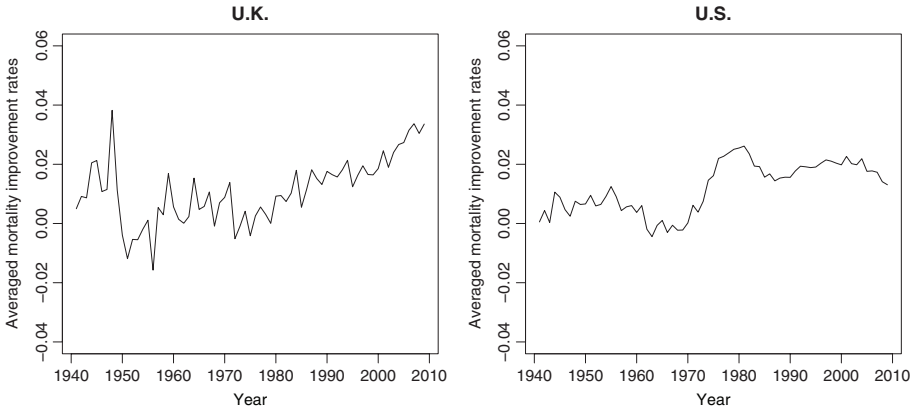

FIGURE 2: Historical annualized average mortality improvement rates.

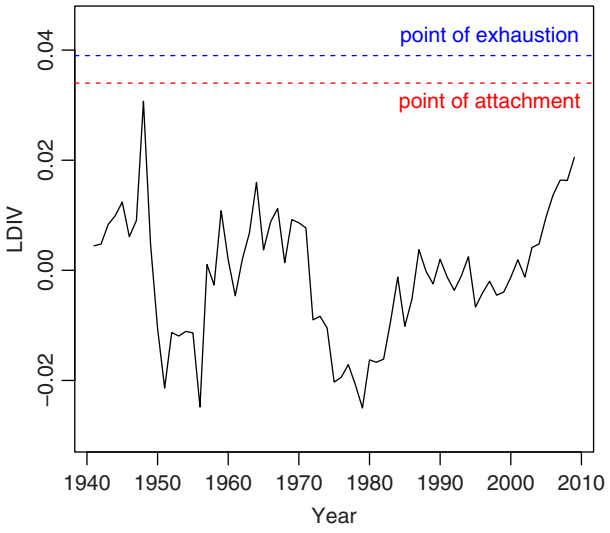

FIGURE 3: Historical LDIV up to year 2009.

or downward trend in the historical index value, besides the seemingly gradual increase in the index value after the mid-1970s. More importantly, it is shown in Figure 3 that historically the LDIV has always stayed below the point of attachment. In fact, the maximum value of the index is only 3.07%. Therefore, the Kortis bond is designed to protect the issuer from extreme mortality experience.

2.4. The ARIMA–GARCH framework

As mentioned earlier, two main references on quantitative analysis of the Kortis bond are Hunt and Blake (2015) and Chen *et al.* (2017). In Hunt and

https://doi.org/10.1017/asb.2019.19 Published online by Cambridge University Press

Blake (2015), mortality rates were fitted by age-period-cohort models. The dynamics between the two countries' mortality experience was then captured by co-integration time-series models. Chen *et al.* (2017) adopted a fairly different approach. They first fitted the mortality improvement rate for each age using independent ARIMA–GARCH models, and then applied a factor copula model to capture the mortality dependence.

Different from the abovementioned references, we directly model the 8-year annualized age-country-specific mortality improvement rate calculated as $\triangle m_t^j(x)$ in Section 2.2, which is precisely as is specified in the construction of the LDIV (Standard and Poor's, 2010). There are similarities between our approach and that of Chen *et al.* (2017) since both involve time-series models.[4] However, we note that Chen *et al.* (2017) chose to model the 1-year log mortality improvement rates instead. A GARCH component is considered here as the heteroskedasticity in mortality dynamics is well recognized in the literature (see discussions in Lee and Miller, 2001). We conduct the Ljung–Box test on the residuals from ARIMA models first and only include the GARCH component when the null hypothesis is rejected at 5% level of significance. The Akaike Information Criteria (AIC) is used to select the optimal ARIMA–GARCH model for each age group included in this work (Akaike, 1974).

Besides analyzing the age-specific mortality improvement rates used to calculate the LDIV, it is also important to consider the mortality improvement at aggregated levels. As bottom-level data (in our case, the age-country-specific improvement rates) are more volatile and noisy, they are generally more difficult to model, especially with limited sample sizes (Shlifer and Wolff, 1979; Schwarzkopf *et al.*, 1988; Athanasopoulos *et al.*, 2009). On the other hand, top-level data, despite the fact that they may exhibit a loss of information due to the aggregation process, are often less noisy and thus provide a clearer picture of any underlying trends. Therefore, modeling average mortality improvement rates of the two countries and their longevity divergence index will provide us with additional information on the pricing of the Kortis bond. We apply the same modeling approach to $\triangle m_t^{\mathrm{UK}}(75, 85)$, $\triangle m_t^{\mathrm{US}}(55, 65)$, and LDIV_t. The model selection results are shown in Table 2, where p is the order of the AR model, d is the order of differencing, q is the order of the MA model, m is the order of the ARCH model, and n is the order of the GARCH model.[5]

3. A FORECAST RECONCILIATION APPROACH

As described in Section 2.2, the construction of the LDIV is achieved via a hierarchical structure. We plot this three-level hierarchical tree of the longevity divergence index in Figure 4. At the bottom of the hierarchy, we have age-specific mortality improvement rates for the UK and US male populations. At the middle level, we have average mortality improvement rates calculated from the bottom level. At the top of the hierarchy, we have the LDIV, which is the difference between the UK and US average mortality improvement rates.

https://doi.org/10.1017/asb.2019.19 Published online by Cambridge University Press

TABLE 2

SELECTED ARIMA–GARCH MODELS BASED ON THE AIC.

Series	ARIMA			GARCH		Series	ARIMA			GARCH	
	p	d	q	m	n		p	d	q	m	n
LDIV_t	1	0	0	1	2						
$\triangle m_t^{\text{UK}}(75,85)$	0	1	1	1	1	$\triangle m_t^{\text{US}}(55,65)$	0	1	0	1	1
$\triangle m_t^{\text{UK}}(75)$	0	1	2	1	2	$\triangle m_t^{\text{US}}(55)$	1	1	0	0	0
$\triangle m_t^{\text{UK}}(76)$	2	1	1	2	2	$\triangle m_t^{\text{US}}(56)$	0	1	0	0	0
$\triangle m_t^{\text{UK}}(77)$	0	1	1	1	2	$\triangle m_t^{\text{US}}(57)$	0	1	0	1	1
$\triangle m_t^{\text{UK}}(78)$	0	1	1	1	2	$\triangle m_t^{\text{US}}(58)$	1	1	0	1	1
$\triangle m_t^{\text{UK}}(79)$	0	1	1	1	1	$\triangle m_t^{\text{US}}(59)$	0	1	1	1	1
$\triangle m_t^{\text{UK}}(80)$	0	1	1	0	0	$\triangle m_t^{\text{US}}(60)$	1	1	0	1	1
$\triangle m_t^{\text{UK}}(81)$	0	1	1	1	1	$\triangle m_t^{\text{US}}(61)$	0	1	3	1	1
$\triangle m_t^{\text{UK}}(82)$	0	1	1	1	1	$\triangle m_t^{\text{US}}(62)$	0	1	1	1	3
$\triangle m_t^{\text{UK}}(83)$	0	1	1	1	1	$\triangle m_t^{\text{US}}(63)$	0	1	1	2	2
$\triangle m_t^{\text{UK}}(84)$	0	1	2	3	1	$\triangle m_t^{\text{US}}(64)$	0	1	0	0	0
$\triangle m_t^{\text{UK}}(85)$	0	1	1	1	1	$\triangle m_t^{\text{US}}(65)$	0	1	1	0	0

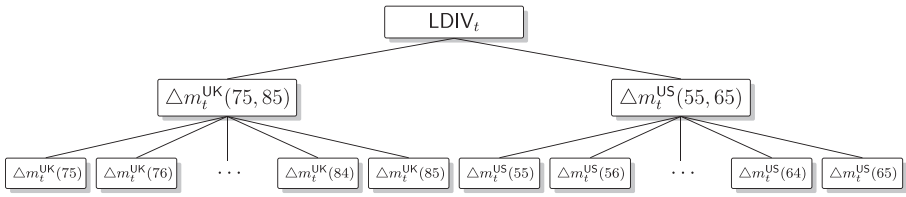

FIGURE 4: Hierarchical tree of the longevity divergence index.

When modeling and forecasting the LDIV, it is important to consider information on all levels. It is expected that the combination of all such information will provide a more accurate projection of the future LDIV. Based on the selected optimal ARIMA–GARCH models in Section 2.3, it is easy for us to produce forecasts for each time series shown in Figure 4. We refer to these as "base" forecasts. Ideally, we want the forecasts to add up in a way that is consistent with the underlying hierarchical structure. However, in reality it is very unlikely that base forecasts will add up in the same manner as the original data (Hyndman and Athanasopoulos, 2014). Therefore, we need to adjust these base forecasts to ensure that they become coherent and follow the aggregation constraints. To do this, we need to reconcile the forecasts at each level taking into account information at other levels.

For the hierarchical structure of the LDIV, we have the following aggregation constraints for all values of t:

$$\frac{1}{11}\sum_{a=75}^{85}\triangle m_t^{\text{UK}}(a)=\triangle m_t^{\text{UK}}(75,85), \tag{3.1}$$

https://doi.org/10.1017/asb.2019.19 Published online by Cambridge University Press

$$\frac{1}{11}\sum_{b=55}^{65}\triangle m_t^{\mathrm{US}}(b) = \triangle m_t^{\mathrm{US}}(55,65), \tag{3.2}$$

$$\triangle m_t^{\mathrm{UK}}(75,85) - \triangle m_t^{\mathrm{US}}(55,65) = \mathrm{LDIV}_t. \tag{3.3}$$

3.1. The minimum trace reconciliation method

There has been a rich literature on forecast reconciliation (see, e.g., Stone *et al.*, 1942; Shlifer and Wolff, 1979; Schwarzkopf *et al.*, 1988; Weale, 1988; Dangerfield and Morris, 1992; Kahn, 1998; Zellner and Tobias, 2000; Athanasopoulos *et al.*, 2009; Hyndman *et al.*, 2011; Van Erven and Cugliari, 2015; Wickramasuriya *et al.*, 2018). Traditionally, the most common techniques to forecast hierarchical time series are the "bottom-up" and "top-down" methods.[6] The "bottom-up" method simply aggregates all bottom-level base forecasts to produce forecasts at higher levels in the hierarchy. In doing so, no information at the bottom level is lost. However, the major disadvantage of this method is that since the bottom-level series are generally more noisy, they are more difficult to model and forecast (Hyndman and Athanasopoulos, 2014). Moreover, the method does not take into account the correlation structure of the errors among disaggregated series.

The appropriateness of forecast reconciliation was formally established in Van Erven and Cugliari (2015) for the first time. Along this direction, an optimal combination method has been proposed to reconcile hierarchical time series (Hyndman *et al.*, 2011). The method aims to combine the base forecasts at all levels to achieve better forecasting results. A regression model is adopted to combine and reconcile base forecasts. The method has shown superior forecasting performance to traditional reconciliation methods and has been applied to various disciplines (see, e.g., Athanasopoulos *et al.*, 2009; Capistrán *et al.*, 2010; Borges *et al.*, 2013; Syntetos *et al.*, 2016; Shang and Hyndman, 2017).

Wickramasuriya *et al.* (2018) extended the work of Hyndman *et al.* (2011) and introduced a minimum trace (MinT) reconciliation approach. The approach aims to minimize the sum of variances of the reconciled forecast errors and thus to find the minimum variance unbiased estimates of the forecasts. Moreover, Wickramasuriya *et al.* (2018) provided a theoretical proof to justify the use of variances and covariances of base forecast errors in their approach, which is missing in the previous literature. As such, although dependence is often ignored in producing base forecasts, it is taken into account in the reconciliation process. In addition, Wickramasuriya *et al.* (2018) have shown that the reconciled forecasts will be at least as good as the base forecasts, which guarantees the effectiveness of the approach.

In our analysis, we adopt the MinT approach to reconciling forecasts of the LDIV and mortality improvement rates. Before looking into details of the MinT approach, we first introduce some notation and terminologies to be used in this section.

https://doi.org/10.1017/asb.2019.19 Published online by Cambridge University Press

For $t \in [1, T]$,

- let $\boldsymbol{y}_t = (\text{LDIV}_t,\ \triangle m_t^{\text{UK}}(75, 85),\ -\triangle m_t^{\text{US}}(55, 65),\ \frac{1}{11}\triangle m_t^{\text{UK}}(75),\ \ldots,\ \frac{1}{11}\triangle m_t^{\text{UK}}(85),\ -\frac{1}{11}\triangle m_t^{\text{US}}(55),\ \ldots,\ -\frac{1}{11}\triangle m_t^{\text{US}}(65))'$ be a vector that contains observations of all series in the hierarchy;
- let $\boldsymbol{b}_t = (\frac{1}{11}\triangle m_t^{\text{UK}}(75),\ \ldots,\ \frac{1}{11}\triangle m_t^{\text{UK}}(85),\ -\frac{1}{11}\triangle m_t^{\text{US}}(55), \ldots, -\frac{1}{11}\triangle m_t^{\text{US}}(65))'$ be a vector that contains observations at the bottom level only.

We can then link these two vectors by the equation

$$\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{b}_t, \tag{3.4}$$

where $\boldsymbol{S}$ is a "summing matrix" of dimension 25×22, which aggregates age-country-specific mortality rates to construct the LDIV. It is given by

$$\boldsymbol{S} = \begin{pmatrix} 1\,1 & 1 & \ldots\ldots & 1 & 1 & 1 \\ 1\,1\ldots & 1 & 0 & 0\ldots & 0 & \\ 0\,0\ldots & 0 & 1 & 1\ldots & 1 & \\ & & \boldsymbol{I}_{22} & & & \end{pmatrix},$$

where $\boldsymbol{I}_{22}$ denotes a 22×22 identity matrix. The aggregation constraints (3.1)–(3.3) are reflected by the first three rows of the matrix $\boldsymbol{S}$.

Let $\hat{\boldsymbol{y}}_{T+h}$ be a vector of h-step-ahead base forecasts of all series in the hierarchy, and $\hat{\boldsymbol{b}}_{T+h}$ be a vector of h-step-ahead base forecasts of bottom-level series only. We produce these base forecasts based on the ARIMA–GARCH models selected in Table 2. According to Wickramasuriya *et al.* (2018), all linear reconciliation methods can be expressed as

$$\tilde{\boldsymbol{y}}_{T+h} = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_{T+h}, \tag{3.5}$$

for some selected matrix $\boldsymbol{P}$ of dimension 22×25, where $\tilde{\boldsymbol{y}}_{T+h}$ represents a vector of reconciled forecasts for all levels and satisfies the aggregation constraints. The choice of $\boldsymbol{P}$ is not unique. For example, for the bottom-up reconciliation approach, $\boldsymbol{P}$ is chosen as

$$\boldsymbol{P} = (\boldsymbol{0}_{22\times 3}, \boldsymbol{I}_{22}), \tag{3.6}$$

where $\boldsymbol{0}_{22\times 3}$ is a zero matrix of dimension 22×3. Therefore, the method simply extracts the bottom-level base forecasts and sums them up to form forecasts for higher levels.

Hyndman *et al.* (2011) proposed an optimal combination approach in which they expressed the base forecasts as

$$\hat{\boldsymbol{y}}_{T+h} = \boldsymbol{S}\boldsymbol{\beta}_{T+h} + \boldsymbol{\epsilon}_{T+h}, \tag{3.7}$$

https://doi.org/10.1017/asb.2019.19 Published online by Cambridge University Press

where $\boldsymbol{\beta}_{T+h} = \mathbb{E}[\boldsymbol{b}_{T+h}|\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T]$ is the unknown mean of bottom-level base forecasts, and $\boldsymbol{\epsilon}_{T+h}$ is the error term which has a mean vector zero and a covariance matrix $\boldsymbol{\Sigma}_h$. As shown by Hyndman *et al.* (2011), if $\boldsymbol{\Sigma}_h$ is known, we can find the generalized least squares (GLS) estimator of $\boldsymbol{\beta}_{T+h}$ as

$$\hat{\boldsymbol{\beta}}_{T+h} = (\boldsymbol{S}'\boldsymbol{\Sigma}_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{\Sigma}_h^{\dagger}\hat{\boldsymbol{y}}_{T+h}, \tag{3.8}$$

where $\boldsymbol{\Sigma}_h^{\dagger}$ is the Moore–Penrose generalized inverse of $\boldsymbol{\Sigma}_h$ as $\boldsymbol{\Sigma}_h$ is likely to be singular.

We then obtain the reconciled forecasts by

$$\tilde{\boldsymbol{y}}_{T+h} = \boldsymbol{S}\hat{\boldsymbol{\beta}}_{T+h} = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{\Sigma}_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{\Sigma}_h^{\dagger}\hat{\boldsymbol{y}}_{T+h}. \tag{3.9}$$

This implies that $\boldsymbol{P} = (\boldsymbol{S}'\boldsymbol{\Sigma}_h^{\dagger}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{\Sigma}_h^{\dagger}$. Since $\boldsymbol{SPS} = \boldsymbol{S}$, the reconciled forecasts are shown to be unbiased given that base forecasts are also unbiased (Hyndman *et al.*, 2011).

However, as discussed and proven by Wickramasuriya *et al.* (2018), $\boldsymbol{\Sigma}_h$ is generally not known nor identifiable. Wickramasuriya *et al.* (2018) provided an alternative estimation of $\boldsymbol{P}$ by minimizing the trace of $\mathrm{VAR}[\boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}_{t+h}|\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_t]$, which is the covariance matrix of the in-sample reconciled forecast errors. This method is referred to as the MinT reconciliation approach.

Let $\boldsymbol{W}_h$ be a positive definite covariance matrix of the h-step-ahead in-sample base forecast errors, that is,

$$\boldsymbol{W}_h = \mathbb{E}[\hat{\boldsymbol{e}}_{t+h}\hat{\boldsymbol{e}}'_{t+h}|\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_t], \tag{3.10}$$

where $\hat{\boldsymbol{e}}_{t+h} = \boldsymbol{y}_{t+h} - \hat{\boldsymbol{y}}_{t+h}$.[7] One can verify that[8]

$$\mathrm{VAR}[\boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}_{t+h}|\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_t] = \boldsymbol{SPW}_h\boldsymbol{P}'\boldsymbol{S}. \tag{3.11}$$

The optimal reconciliation matrix is then given by

$$\boldsymbol{P} = (\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1}. \tag{3.12}$$

Thus, the only difference between the GLS solution given by Hyndman *et al.* (2011) and the MinT solution given by Wickramasuriya *et al.* (2018) is the covariance matrix used in the estimation of $\boldsymbol{P}$. However, in the GLS solution, $\boldsymbol{\Sigma}_h$ is generally not identifiable. This issue of the unidentifiable covariance matrix was ultimately resolved by Wickramasuriya *et al.* (2018). The contribution of Wickramasuriya *et al.* (2018) was to recognize that this unidentifiable matrix was not required at all to minimize the trace of the reconciliation error covariance matrix. The MinT reconciliation approach therefore gives a more practically feasible solution, and we choose to adopt the MinT approach to reconcile forecasts of the LDIV for the Kortis bond.[9]

https://doi.org/10.1017/asb.2019.19 Published online by Cambridge University Press

3.2. Reconciling probabilistic forecasts of the LDIV

As described in the previous section, the MinT approach provides an effective solution for reconciling point forecasts. However, in order to price the Kortis bond, it is required to estimate the entire distribution of the PRF. Therefore, we need to reconcile not only the point forecasts but also the probabilistic forecasts of the LDIV.

In the existing literature, there are very few attempts to reconcile probabilistic forecasts and construct reconciled prediction intervals. To the best of our knowledge, only papers tackling this issue are Ben Taieb *et al.* (2017) and Jeon *et al.* (2019). Both works aim to adjust probabilistic forecasts to ensure that the aggregation constraints are met. In particular, Jeon *et al.* (2019) are the first to utilize information on full probabilistic distributions of all levels to produce the reconciled probabilistic forecasts. Moreover, Jeon *et al.* (2019) provided a generalization of the MinT approach.

Following Jeon *et al.* (2019), we first define the following terms:

- Let $\boldsymbol{f}(\boldsymbol{y}_{T+h}|\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T)$ be the probabilistic distribution of h-step-ahead forecasts;
- Let $\hat{\boldsymbol{y}}^i_{T+h}$ denote the ith sample of base forecasts generated from $\boldsymbol{f}(\boldsymbol{y}_{T+h}|\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T)$;
- A sample of size N generated from $\boldsymbol{f}(\boldsymbol{y}_{T+h}|\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T)$ is denoted by $\hat{\boldsymbol{Y}}$, where $\hat{\boldsymbol{Y}} = (\hat{\boldsymbol{y}}^1_{T+h}, \hat{\boldsymbol{y}}^2_{T+h}, \ldots, \hat{\boldsymbol{y}}^N_{T+h})$.

Clearly, there is no guarantee (in fact, it is highly unlikely) that each column of $\hat{\boldsymbol{Y}}$ will satisfy the aggregation constraints required by the hierarchical structure. Therefore, a reconciliation process is needed as follows:

$$\tilde{\boldsymbol{Y}} = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{Y}}, \tag{3.13}$$

where $\boldsymbol{S}$ and $\boldsymbol{P}$ are the same as defined in Section 3.1,[10] and $\tilde{\boldsymbol{Y}}$, a matrix of dimension $25 \times N$, is the reconciled forecasts for N sample paths.

We can construct the predictive probabilistic distribution and thus prediction intervals from the reconciled samples $\tilde{\boldsymbol{Y}}$. The problem is then down to how to construct unreconciled forecasts $\hat{\boldsymbol{Y}}$. To this end, Jeon *et al.* (2019) introduced three schemes, namely the "stacked sample", "permuted sample," and "ranked sample." The empirical results in Jeon *et al.* (2019) show that the "ranked sample" method has the best performance among the three schemes, in particular when there is a positive correlation between the underlying series. In this work, we adopt the "ranked sample" method and construct our sample in the following steps:

- Define $\hat{\boldsymbol{Z}}_m$ to be an $N \times 1$ column matrix, which contains base forecasts of size N for the mth series in $\boldsymbol{y}_{T+h}$. For example, $\hat{\boldsymbol{Z}}_1$ represents N samples generated from the distribution $\boldsymbol{f}(\text{LDIV}_{T+h}|\text{LDIV}_1, \text{LDIV}_2, \ldots, \text{LDIV}_T)$.

https://doi.org/10.1017/asb.2019.19 Published online by Cambridge University Press

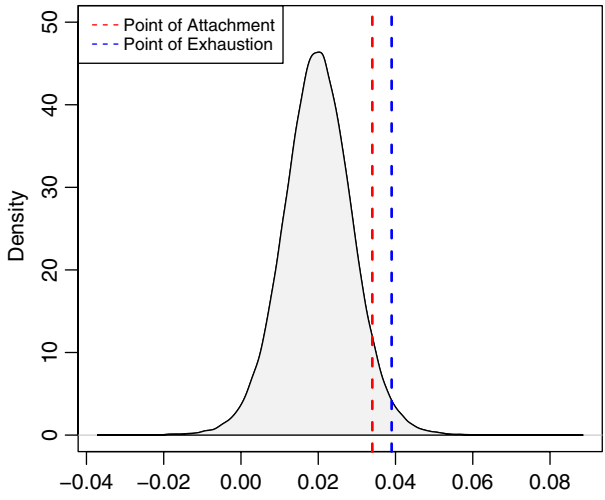

FIGURE 5: Density of the LDIV in 2016.

- Arrange the elements in $\hat{\boldsymbol{Z}}_m$ in ascending order to form $\hat{\boldsymbol{Z}}_m^R$.
- Let $\hat{\boldsymbol{Y}}^R = \left(\hat{\boldsymbol{Z}}_1^R, \hat{\boldsymbol{Z}}_2^R, \ldots, \hat{\boldsymbol{Z}}_{25}^R\right)'$ be the ranked sample of size N for all series.

In this way, we obtain reconciled forecasts for N sample paths as

$$\tilde{\boldsymbol{Y}} = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{Y}}^R. \tag{3.14}$$

For each age-specific mortality improvement rate, average mortality improvement rate, and LDIV, we generate 10,000 ranked samples to build base probabilistic forecasts. Based on these samples, we reconcile the probabilistic forecasts of the LDIV and then estimate the distribution of the PRF. In Figure 5, we plot the forecast density of the LDIV in 2016. To clearly illustrate the implication of the results, we present selected estimated quantiles of the PRF in Table 3.[11] For comparison, we also include the results from the RMS, the base forecasts, the bottom-up approach, and the optimal combination GLS approach.

It can be seen that our estimates from the MinT reconciliation approach are highly consistent with the RMS estimates. The expected loss based on the MinT approach is only 0.14% higher than the figure published in the RMS report. The fact that the expected loss and the conditional expected loss resulting from the MinT approach are so close to the RMS estimates will also lead to similar results for pricing. Even though our approach to projecting the LDIV is quite different from the approach used by RMS, these results in Table 3 still to a certain extent ensure the reliability of our reconciliation approach. However, until more information on the modeling details from the RMS becomes available, we are not able to make further comments on the accuracy of LDIV forecasts.

https://doi.org/10.1017/asb.2019.19 Published online by Cambridge University Press

TABLE 3

ESTIMATED QUANTILES OF THE PRF OF THE KORTIS BOND.

		Estimated probability				
LDIV≥	PRF≥	Bottom up	Base	GLS	MinT	RMS
3.40%	0%	29.32%	0.33%	2.11%	5.61%	5.31%
3.50%	20%	27.51%	0.25%	1.65%	4.51%	4.32%
3.60%	40%	25.76%	0.19%	1.29%	3.63%	3.48%
3.70%	60%	24.09%	0.14%	1.01%	2.92%	2.82%
3.80%	80%	22.51%	0.10%	0.78%	2.34%	2.28%
3.90%	100%	20.98%	0.07%	0.60%	1.89%	1.81%
Expected loss		25.00%	0.18%	1.21%	3.41%	3.27%
Conditional expected loss		85%	53%	57%	61%	62%

On the other hand, we can clearly see that the bottom-up method potentially overestimates the quantiles of the PRF. As a result, the estimated expected loss and the conditional expected loss based on the bottom-up method are much higher than the MinT and RMS estimates. Indeed, we observe that the expected loss from the bottom-up method is almost 8 times as large as the MinT and RMS estimates. On the contrary, we find that the estimates from the base forecasts are substantially lower than the MinT and RMS estimates. This may be because aggregation has led to a loss of information from lower levels, and as a result the base forecasts on the top level do not fully reflect the future volatility of the index. Consequently, the expected loss and the conditional expected loss calculated from the base forecasts appear to be too low given the market spread of the bond. The GLS quantile estimates are between the bottom-up estimates and the base forecast estimates. However, they are still well below the RMS figures, such that the resulting expected loss is less than half of published expected loss (3.27%) by the RMS. These results well bring out the effectiveness of the MinT reconciliation method.

We conclude that the reconciliation approach gives much more reliable probabilistic forecasts of the LDIV than the other methods included in our analysis, and at the same time ensures that all aggregation constraints in the hierarchy are met. In Figures 6, 7, 8, and 9, we plot the 90%, 95%, and 99% prediction intervals of the LDIV over the period 2010–2016, based on the bottom-up forecasts, base forecasts, GLS forecasts, and MinT forecasts, respectively. These plots are consistent with our findings from Table 3.

4. DISCUSSIONS ON THE INTEREST SPREAD OF THE KORTIS BOND

As mentioned in Section 2.1, the interest spread of the Kortis bond at issue was 5%. Although it is unclear whether a secondary market existed for the Kortis bond between 2011 and 2016, Lane and Beckwith (2011, 2012, 2013,

https://doi.org/10.1017/asb.2019.19 Published online by Cambridge University Press

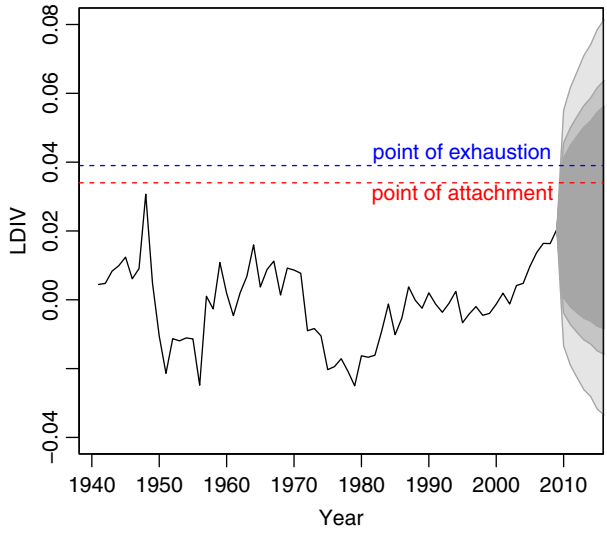

FIGURE 6: The 90%, 95%, and 99% prediction intervals of LDIV from the bottom-up method.

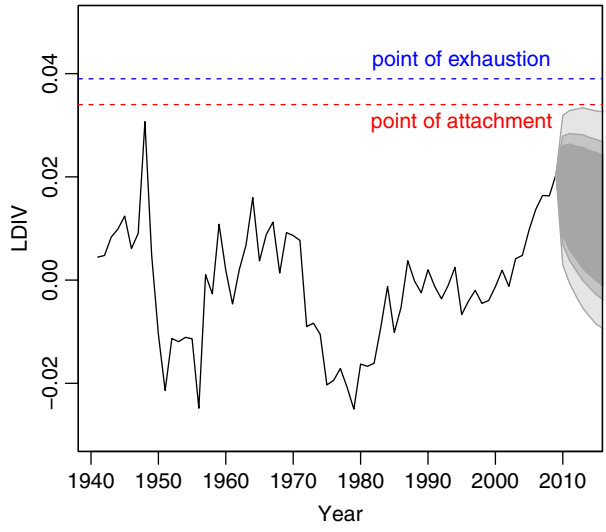

FIGURE 7: The 90%, 95%, and 99% prediction intervals of LDIV from the base forecasts.

2014, 2015, 2016, 2017) published "Average Market Indications" for the interest spread of the bond at the end of each quarter since March 2011. We plot these published figures in Figure 10. It can be seen that the market-indicated spread of the bond has a decreasing trend over time, which is in line with many other catastrophe bonds.[12] However, it is worth noting that the payment

https://doi.org/10.1017/asb.2019.19 Published online by Cambridge University Press

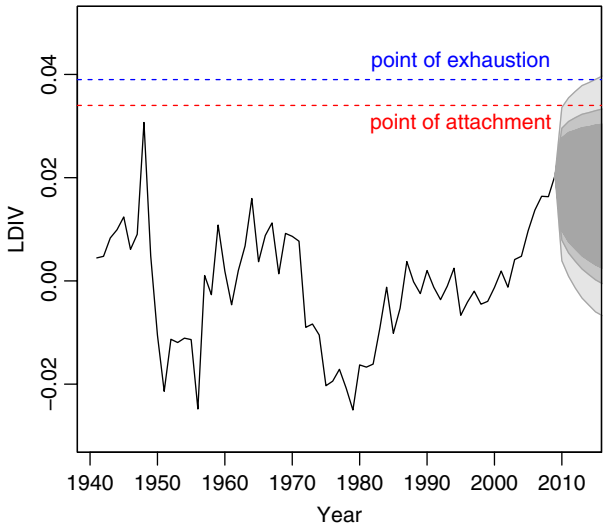

FIGURE 8: The 90%, 95%, and 99% prediction intervals of LDIV from the GLS reconciliation.

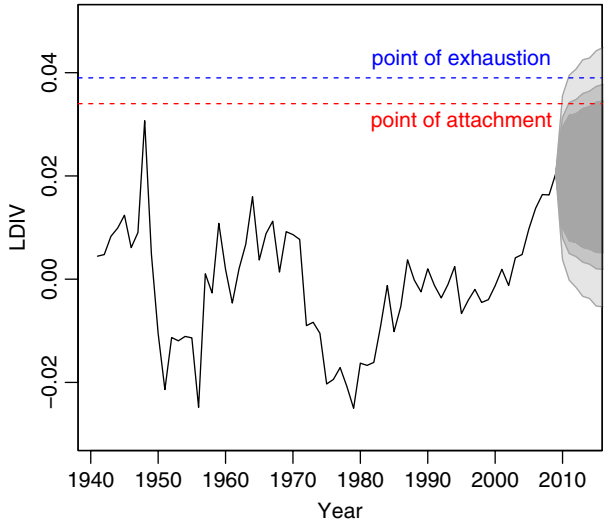

FIGURE 9: The 90%, 95%, and 99% prediction intervals of LDIV from the MinT reconciliation.

structure of the Kortis bond is different from traditional catastrophe bonds as the bondholders' principal payment can only be reduced at the end of the risk period.

Assuming that a secondary market for mortality bonds will emerge and develop in the future, it is important for us to keep track of the movements of

https://doi.org/10.1017/asb.2019.19 Published online by Cambridge University Press

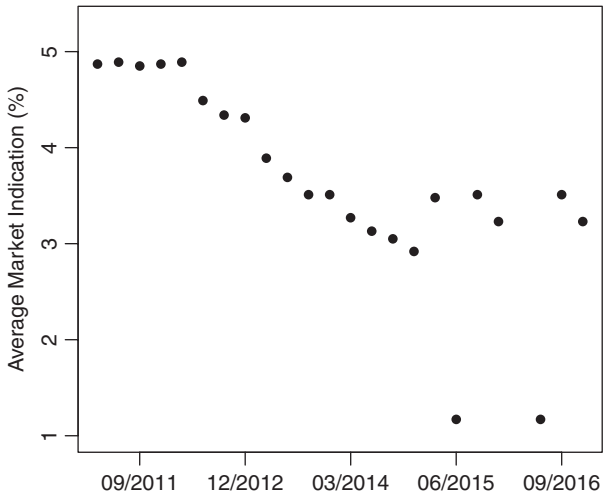

FIGURE 10: Indicated spread of the Koris bond.

mortality indexes and update the forecasts of the PRF regularly. The Kortis bond reached its maturity on the 31st of December 2016, and information on mortality improvement rates up to 2016 has now become available. Therefore, in this work, we provide first evidence on changes in the distribution of the PRF since its issue date. We repeat the reconciliation process described in Section 3 with additional mortality data from 2010 to 2015. The results for estimated quantiles of the PRF are shown in Table 4.[13]

We can see that the expected loss based on our approach shows a decreasing trend, which is consistent with the trend we discovered in the average market indications earlier on in this section. On the other hand, there is no clear trend found in the conditional expected loss during the same period. In summary, as time goes by, the LDIV becomes less likely to hit the point of attachment/exhaustion.

To obtain more insights on the changes in the distribution of the PRF over time, we plot the historical LDIV with additional data from 2010 to 2016. The crosses in Figure 11 represent "new information" which came in since the issue of the Kortis bond. We can see that the LDIV reached its maximum in 2011. After that, there was a slight decrease in the value of the index. The observed value of the LDIV in 2016 is 2.09%, which is still far below 3.4%, the point of attachment of the bond. Even though the added observations of the LDIV only illustrate changes in the most aggregated level during 2010–2016, they still validate the estimation results in Table 4 to a certain degree. For example, there was a jump in the LDIV in 2011, and according to Table 4 the bondholders' principal is more likely to be reduced (thus the expected loss tends to be higher) compared to the previous year.

https://doi.org/10.1017/asb.2019.19 Published online by Cambridge University Press

TABLE 4

ESTIMATED QUANTILES OF THE PRF WITH ADDITIONAL MORTALITY INFORMATION.

		Estimated probability					
LDIV≥	PRF≥	2010	2011	2012	2013	2014	2015
3.40%	0%	4.34%	4.98%	1.86%	1.29%	2.12%	2.70%
3.50%	20%	3.58%	3.76%	1.41%	1.15%	1.82%	1.97%
3.60%	40%	2.77%	2.98%	1.02%	1.02%	1.56%	1.46%
3.70%	60%	2.14%	2.36%	0.74%	0.92%	1.32%	1.00%
3.80%	80%	1.63%	1.71%	0.55%	0.81%	1.12%	0.71%
3.90%	100%	1.25%	1.24%	0.40%	0.72%	0.93%	0.56%
Expected loss		2.59%	2.78%	1.00%	0.98%	1.47%	1.35%
Conditional expected loss		60%	56%	54%	76%	69%	50%

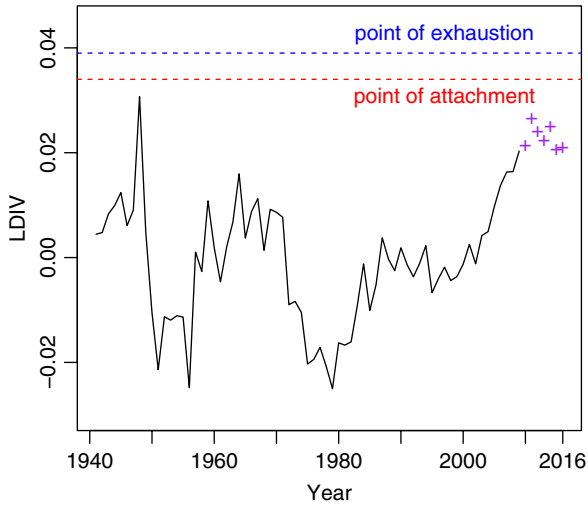

FIGURE 11: Historical LDIV up to year 2016.

In addition, to illustrate the out-of-sample point forecast performance of the MinT approach during 2010–2016, for each individual series and level in the hierarchy, we calculate the root mean squared forecast error (RMSFE) based on the bottom-up method, the base forecasts, the GLS approach, and the MinT approach. We find that the MinT reconciliation approach provides the most accurate point forecasts compared to other methods for a majority of individual series, and improves the overall forecast accuracy at all levels in the hierarchy. These results are reported in Tables A1 and A2 in the Appendix, with bold figures highlighting the method that provides the best forecast performance.

https://doi.org/10.1017/asb.2019.19 Published online by Cambridge University Press

5. Conclusions

The desire to transfer and hedge mortality and longevity risk has led to the emergence and growth of the longevity capital market in recent decades. Therefore, accurate projections of mortality bond indexes are of fundamental importance for the pricing of mortality bonds. In this paper, we propose a hierarchical forecast reconciliation approach to constructing the probabilistic forecasts of mortality bond indexes and demonstrate the strong performance of our method.

A MinT reconciliation method (Wickramasuriya *et al.*, 2018) is applied together with the probabilistic forecast sampling algorithm proposed by Jeon *et al.* (2019) to estimate the distribution of the PRF for the Kortis bond. The estimated quantiles of the PRF based on our approach are very close to the figures published in Standard and Poor's (2010). We also compare our results with the bottom-up and base forecasts. It is shown that in the case of Kortis bond, forecast reconciliation is an effective tool to utilize all available information and provide more accurate forecasts. In addition, we apply the proposed method with the most recent mortality data and provide first insights on the changes in distribution of PRF throughout the risk period of the bond.

Acknowledgments

We would like to thank Dr Jooyoung Jeon, Dr Anastasios Panagiotelis and Dr Fotios Petropoulos for sharing with us their paper Jeon *et al.* (2019) prior to its publication. We would also like to thank two anonymous referees for their insightful comments which have greatly improved our work.

Notes

1. LIBOR is a key global benchmark interest rate at which banks offer to lend funds to each other.
2. As Kortis bond is the first longevity trend bond introduced in the market, there are no other benchmark estimates other than the RMS estimates for its loss distribution (see discussions in Hunt and Blake, 2015; Chen *et al.*, 2017). The quantile estimates and the expected loss of the Kortis bond published by RMS have been frequently used by other financial service companies (see, e.g., Lane and Beckwith, 2011).
3. The HMD mortality database can be found at http://www.mortality.org/
4. For other examples that apply time-series models to mortality modeling, see Engle (2001), Gao and Hu (2009), Giacometti *et al.* (2012), Sarpong (2013), Wang *et al.* (2013), Lin *et al.* (2015), and Chen *et al.* (2015).
5. `R` packages `forecast` (Hyndman and Khandakar, 2008) and `rugarch` (Ghalanos, 2019) are used to select optimal models and generate forecasts.
6. Since in the case of the Kortis bond we primarily focus on the forecasts of LDIV, we do not further discuss the top-down method. For more details of each method, see Hyndman and Athanasopoulos (2014).
7. In our analysis, although the individual series themselves are mostly nonstationary, the in-sample forecast errors we used to compute the $\boldsymbol{W}_h$ matrix are in fact stationary. Consequently, the MinT approach is valid in our study.

https://doi.org/10.1017/asb.2019.19 Published online by Cambridge University Press

8. For a detailed proof, see Wickramasuriya *et al.* (2018), Appendix A.1.

9. There are several choices of the estimator of $\boldsymbol{W}_h$ proposed in Wickramasuriya *et al.* (2018), Section 2.4. In this work, we have used the "shrinkage" estimator as described there.

10. Note that for the bottom-up reconciliation, $\boldsymbol{P} = (\boldsymbol{0}_{22\times 3}, \boldsymbol{I}_{22})$, which is the same as in Section 3.1.

11. The MinT reconciliation method is implemented by `R` package `hts` (Hyndman *et al.*, 2018).

12. We do not attempt to comment further on the values of the market-indicated spreads of the Kortis bond since it is beyond the scope of the paper; for further details, please refer to Lane and Beckwith (2011, 2012, 2013, 2014, 2015, 2016, 2017).

13. Note that in Table 4, "2010" indicates that the reconciled forecasts are based on information up to 2010. The same rule applies to the other years.

References

Akaike, H. (1974) A new look at the statistical model identification. *IEEE transactions on automatic control*, **19**(6), 716–723.

Athanasopoulos, G., Ahmed, R.A. and Hyndman, R.J. (2009) Hierarchical forecasts for australian domestic tourism. *International Journal of Forecasting*, **25**(1), 146–166.

Bauer, D. and Kramer, F. (2016) The risk of a mortality catastrophe. *Journal of Business & Economic Statistics*, **34**(3), 391–405.

Bauer, D., Börger, M. and Russ, J. (2010) On the pricing of longevity-linked securities. *Insurance: Mathematics and Economics*, **46**(1), 139–149.

Ben Taieb, S., Taylor, J.W. and Hyndman, R.J. (2017) Coherent probabilistic forecasts for hierarchical time series. In *International Conference on Machine Learning*, pp. 3348–3357.

Biagini, F., Rheinländer, T. and Widenmann, J. (2013) Hedging mortality claims with longevity bonds. *ASTIN Bulletin: The Journal of the IAA*, **43**(2), 123–157.

Blake, D., Cairns, A., Coughlan, G., Dowd, K. and MacMinn, R. (2013) The new life market. *Journal of Risk and Insurance*, **80**(3), 501–558.

Borges, C.E., Penya, Y.K. and Fernandez, I. (2013) Evaluating combined load forecasting in large power systems and smart grids. *IEEE Transactions on Industrial Informatics*, **9**(3), 1570–1577.

Braun, A. (2016) Pricing in the primary market for cat bonds: new empirical evidence. *Journal of Risk and Insurance*, **83**(4), 811–847.

Cairns, A.J., Blake, D., Dawson, P. and Dowd, K. (2005) Pricing the risk on longevity bonds. *Life and Pensions*, **1**(2), 41–44.

Cairns, A.J., Blake, D. and Dowd, K. (2006) Pricing death: frameworks for the valuation and securitization of mortality risk. *ASTIN Bulletin: The Journal of the IAA*, **36**(1), 79–120.

Capistrán, C., Constandse, C. and Ramos-Francia, M. (2010) Multi-horizon inflation forecasts using disaggregated data. *Economic Modelling*, **27**(3), 666–677.

Chen, H. and Cox, S.H. (2009) Modeling mortality with jumps: applications to mortality securitization. *Journal of Risk and Insurance*, **76**(3), 727–751.

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**(S1), 393–415.

Chulia, H., Guillen, M. and Uribe, J.M. (2016) Modeling longevity risk with generalized dynamic factor models and vine-copulae. *ASTIN Bulletin: The Journal of the IAA*, **46**(1), 165–190.

Cowley, A. and Cummins, J.D. (2005) Securitization of life insurance assets and liabilities. *Journal of Risk and Insurance*, **72**(2), 193–226.

https://doi.org/10.1017/asb.2019.19 Published online by Cambridge University Press

COX, S.H., FAIRCHILD, J.R. and PEDERSEN, H.W. (2000) Economic aspects of securitization of risk. *ASTIN Bulletin: The Journal of the IAA*, **30**(1), 157–193.
CUMMINS, J.D. and TRAINAR, P. (2009) Securitization, insurance, and reinsurance. *Journal of Risk and Insurance*, **76**(3), 463–492.
DANGERFIELD, B.J. and MORRIS, J.S. (1992) Top-down or bottom-up: Aggregate versus disaggregate extrapolations. *International Journal of Forecasting*, **8**(2), 233–241.
DENG, Y., BROCKETT, P.L. and MACMINN, R.D. (2012) Longevity/mortality risk modeling and securities pricing. *Journal of Risk and Insurance*, **79**(3), 697–721.
ENGLE, R. (2001) Garch 101: The use of arch/garch models in applied econometrics. *Journal of economic perspectives*, **15**(4), 157–168.
GAO, Q. and HU, C. (2009) Dynamic mortality factor model with conditional heteroskedasticity. *Insurance: Mathematics and Economics*, **45**(3), 410–423.
GHALANOS, A. (2019) Introduction to the rugarch package. (Version 1.4-1). *http://cran.r-project.org/web/packages/rugarch*.
GIACOMETTI, R., BERTOCCHI, M., RACHEV, S.T. and FABOZZI, F.J. (2012). A comparison of the Lee–Carter model and AR–ARCH model for forecasting mortality rates. *Insurance: Mathematics and Economics*, **50**(1), 85–93.
HUNT, A. and BLAKE, D. (2015) Modelling longevity bonds: Analysing the swiss re kortis bond. *Insurance: Mathematics and Economics*, **63**, 12–29.
HYNDMAN, R.J. and ATHANASOPOULOS, G. (2014) Forecasting: Principles and practice. https://otexts.com/fpp2/.
HYNDMAN, R.J. and KHANDAKAR, Y. (2008) Automatic time series forecasting: The forecast package for r. *Journal of Statistical Software*, **27**(3), 1–22.
HYNDMAN, R.J., AHMED, R.A., ATHANASOPOULOS, G. and SHANG, H.L. (2011) Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis*, **55**(9), 2579–2589.
HYNDMAN, R.J., ATHANASOPOULOS, G., *et al.* (2014) Optimally reconciling forecasts in a hierarchy. *Foresight: The International Journal of Applied Forecasting*, **35**, 42–48.
HYNDMAN, R.J., LEE, A., WANG, E., WICKRAMASURIYA, S. and WANG, M.E. (2018) Package 'hts'.
JEON, J., PANAGIOTELIS, A. and PETROPOULOS, F. (2019) Probabilistic forecast reconciliation with applications to wind power and electric load, *European Journal of Operational Research*, https://doi.org/10.1016/j.ejor.2019.05.020.
KAHN, K.B. (1998) Revisiting top-down versus bottom-up forecasting. *The Journal of Business Forecasting*, **17**(2), 14.
LANE, M. and BECKWITH, R. (2011) Prague spring or Louisiana morning? Annual review for the four quarters, Q2 2010 to Q1 2011. *Tech. Rep. Lane Finacial LLC.*
LANE, M. and BECKWITH, R. (2012) More return; more risk: Annual review for the four quarters, Q2 2011 to Q1 2012. *Tech. Rep. Lane Finacial LLC.*
LANE, M. and BECKWITH, R. (2013) Soft markets ahead!? Annual review for the four quarters, Q2 2012 to Q1 2013. *Tech. Rep. Lane Finacial LLC.*
LANE, M. and BECKWITH, R. (2014) Straw hats in winter: Annual review for the four quarters, Q2 2013 to Q1 2014. *Tech. Rep. Lane Finacial LLC.*
LANE, M. and BECKWITH, R. (2015) Crawling along or coming off bottom? Annual review for the four quarters, Q2 2014 to Q1 2015. *Tech. Rep. Lane Finacial LLC.*
LANE, M. and BECKWITH, R. (2016) Trace data twenty one months on - ILS trade or quote data? Annual review for the four quarters, Q2 2015 to Q1 2016. *Tech. Rep. Lane Finacial LLC.*
LANE, M. and BECKWITH, R. (2017) Annual review and commentary for the four quarters, Q2 2016 to Q1 2017. *Tech. Rep. Lane Finacial LLC.*
LEE, R. and MILLER, T. (2001) Evaluating the performance of the lee-carter method for forecasting mortality. *Demography*, **38**(4), 537–549.
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.
LIN, Y. and COX, S.H. (2008) Securitization of catastrophe mortality risks. *Insurance: Mathematics and Economics*, **42**(2), 628–637.

LIN, Y., LIU, S. and YU, J. (2013) Pricing mortality securities with correlated mortality indexes. *Journal of Risk and Insurance*, **80**(4), 921–948.
MACMINN, R., BROCKETT, P. and BLAKE, D. (2006) Longevity risk and capital markets. *Journal of Risk and Insurance*, **73**(4), 551–557.
SARPONG, S.A. (2013) Modeling and forecasting maternal mortality; an application of ARIMA models. *International Journal of Applied Science and Technology*, **3**(1), 19–28.
SCHWARZKOPF, A.B., TERSINE, R.J. and MORRIS, J.S. (1988) Top-down versus bottom-up forecasting strategies. *The International Journal Of Production Research*, **26**(11), 1833–1843.
SHANG, H.L. and HABERMAN, S. (2017) Grouped multivariate and functional time series forecasting: An application to annuity pricing. *Insurance: Mathematics and Economics*, **75**, 166–179.
SHANG, H.L. and HYNDMAN, R.J. (2017) Grouped functional time series forecasting: An application to age-specific mortality rates. *Journal of Computational and Graphical Statistics*, **26**(2), 330–343.
SHLIFER, E. and WOLFF, R. (1979) Aggregation and proration in forecasting. *Management Science*, **25**(6), 594–603.
STANDARD AND POOR'S (2010) Presale information: Kortis capital ltd. *Tech. Rep. Standard and Poors*.
STONE, R., CHAMPERNOWNE, D.G. and MEADE, J.E. (1942) The precision of national income estimates. *The Review of Economic Studies*, **9**(2), 111–125.
STUPFLER, G. and YANG, F. (2018) Analyzing and predicting cat bond premiums: A financial loss premium principle and extreme value modeling. *ASTIN Bulletin: The Journal of the IAA*, **48**(1), 375–411.
SYNTETOS, A.A., BABAI, Z., BOYLAN, J.E., KOLASSA, S. and NIKOLOPOULOS, K. (2016) Supply chain forecasting: Theory, practice, their gap and the future. *European Journal of Operational Research*, **252**(1), 1–26.
VAN ERVEN, T. and CUGLIARI, J. (2015) Game-theoretically optimal reconciliation of contemporaneous hierarchical time series forecasts. In *Modeling and Stochastic Learning for Forecasting in High Dimensions*, pp. 297–317. Cham, Switzerland: Springer.
WANG, C.-W., HUANG, H.-C. and LIU, I.-C. (2013) Mortality modeling with non-gaussian innovations and applications to the valuation of longevity swaps. *Journal of Risk and Insurance*, **80**(3), 775–798.
WEALE, M. (1988) The reconciliation of values, volumes and prices in the national accounts. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, **151**(1), 211–221.
WICKRAMASURIYA, S.L., ATHANASOPOULOS, G. and HYNDMAN, R.J. (2018) Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *Journal of the American Statistical Association*, **114**(526), 804–819.
ZELLNER, A. and TOBIAS, J. (2000) A note on aggregation, disaggregation and forecasting performance. *Journal of Forecasting*, **19**(5), 457–465.

HAN LI (Corresponding author)
Department of Actuarial Studies and Business Analytics
Macquarie University
Sydney, NSW 2109, Australia
E-Mail: han.li@mq.edu.au.

QIHE TANG
School of Risk and Actuarial Studies
UNSW Sydney
Sydney, NSW 2052, Australia
E-Mail: qihe.tang@unsw.edu.au.

https://doi.org/10.1017/asb.2019.19 Published online by Cambridge University Press

Department of Statistics and Actuarial Science
University of Iowa
Iowa City, IA 52242, USA
E-Mail: qihe-tang@uiowa.edu.

APPENDIX A

For each individual series, the RMSFE is defined as follows:

$$\text{RMSFE} = \sqrt{\frac{1}{h}\sum_{k=1}^{h}\left(m_{T+k} - \hat{m}_{T+k}\right)^2}, \tag{A.1}$$

where h represents the length of the forecast horizon, m_{T+k} represents the kth actual observation in the holdout sample, and $\hat{m}_{T+k}$ represents the corresponding point forecast.

For each hierarchical level, we compute the average RMSFE by calculating the mean of RMSFE across all individual series within each level. Therefore, the average RMSFE is defined as

$$\text{Average RMSFE} = \frac{1}{n_j}\sum_{i=1}^{n_j}\text{RMSFE}_i, \tag{A.2}$$

where n_j represents the number of individual series at level j in the hierarchy, and RMSFE_i represents the RMSFE of individual series i at level j.

TABLE A1

RMSFE (×100) OF OUT-OF-SAMPLE FORECASTS DURING 2010–2016.

Series	Bottom up	Base	GLS	MinT	Series	Bottom up	Base	GLS	MinT
LDIV_t	0.38	0.81	0.53	**0.34**					
$\triangle m_t^{\text{UK}}(75,85)$	1.08	0.92	**0.86**	**0.86**	$\triangle m_t^{\text{US}}(55,65)$	**0.95**	0.96	1.20	0.97
$\triangle m_t^{\text{UK}}(75)$	NA	1.76	**1.57**	1.73	$\triangle m_t^{\text{US}}(55)$	NA	0.89	**0.53**	0.69
$\triangle m_t^{\text{UK}}(76)$	NA	1.29	**1.11**	1.13	$\triangle m_t^{\text{US}}(56)$	NA	0.60	0.66	**0.45**
$\triangle m_t^{\text{UK}}(77)$	NA	1.28	1.11	**0.97**	$\triangle m_t^{\text{US}}(57)$	NA	**0.56**	0.78	0.58
$\triangle m_t^{\text{UK}}(78)$	NA	1.13	0.95	**0.92**	$\triangle m_t^{\text{US}}(58)$	NA	1.80	1.81	**1.63**
$\triangle m_t^{\text{UK}}(79)$	NA	0.88	**0.73**	0.77	$\triangle m_t^{\text{US}}(59)$	NA	1.59	1.57	**1.26**
$\triangle m_t^{\text{UK}}(80)$	NA	0.78	0.68	**0.67**	$\triangle m_t^{\text{US}}(60)$	NA	**1.10**	1.33	1.17
$\triangle m_t^{\text{UK}}(81)$	NA	1.40	1.21	**1.18**	$\triangle m_t^{\text{US}}(61)$	NA	2.54	**1.77**	**1.77**
$\triangle m_t^{\text{UK}}(82)$	NA	0.82	0.80	**0.79**	$\triangle m_t^{\text{US}}(62)$	NA	2.31	1.64	**1.05**
$\triangle m_t^{\text{UK}}(83)$	NA	1.13	0.99	**0.96**	$\triangle m_t^{\text{US}}(63)$	NA	2.01	1.91	**1.90**
$\triangle m_t^{\text{UK}}(84)$	NA	0.74	**0.62**	0.66	$\triangle m_t^{\text{US}}(64)$	NA	**0.77**	1.01	0.81
$\triangle m_t^{\text{UK}}(85)$	NA	1.17	**1.00**	**1.00**	$\triangle m_t^{\text{US}}(65)$	NA	**1.06**	1.26	1.08

https://doi.org/10.1017/asb.2019.19 Published online by Cambridge University Press

TABLE A2

AVERAGE RMSFE (×100) OF OUT-OF-SAMPLE FORECASTS DURING 2010–2016.

Level	Bottom up	Base	GLS	MinT
Top	0.38	0.81	0.53	**0.34**
Middle	1.02	0.94	1.03	**0.92**
Bottom	NA	1.26	1.14	**1.05**

https://doi.org/10.1017/asb.2019.19 Published online by Cambridge University Press