Analyzing Mortality Bond Indexes via Hierarchical Forecast Reconciliation

Han Li\textsuperscript{[a]} and Qihe Tang\textsuperscript{[b],[c]}

\textsuperscript{[a]}Department of Actuarial Studies and Business Analytics, Macquarie University
\textsuperscript{[b]}School of Risk and Actuarial Studies, UNSW Sydney
\textsuperscript{[c]}Department of Statistics and Actuarial Science, University of Iowa

March 1, 2019

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 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 principal reduction factor of the Kortis bond, and compare our results with those stated in Standard and Poors’ 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.

\textit{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 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 becomes 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. (2018).

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 U.K. population (England & Wales only) and the U.S. population. 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 U.K. male population aged 75 to 85, and that of the U.S. male population aged 55 to 65. During the risk period, the bond will be “triggered” if the longevity 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 three-month London interbank offered rate (LIBOR).\(^1\)

\(^1\)LIBOR is a key global benchmark interest rate at which banks offer to lend funds to each other.
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

Table 1: Cumulative and 6-year annualized loss probabilities estimated by the RMS

<table>
<thead>
<tr>
<th>Cumulative (%)</th>
<th>6-year annualized (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attachment probability</td>
<td>5.31</td>
</tr>
<tr>
<td>Exhaustion probability</td>
<td>1.81</td>
</tr>
<tr>
<td>Expected loss</td>
<td>3.27</td>
</tr>
</tbody>
</table>

2.2 The structure of the Kortis bond

Consider the U.K. male population aged 75 to 85 and the U.S. male population aged 55 to 65. Let \( m_j^t(x) \) 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 \( j \):
  \[
  \Delta m_j^t(x) = 1 - \left( \frac{m_j^t(x)}{m_j^{t-8}(x)} \right)^{1/8}.
  \]  

- Second, we average the mortality improvement rates obtained in Equation (1) across ages \( x_1 \) to \( x_2 \) for each time \( t \) and population \( j \):
  \[
  \Delta m_j^t(x_1, x_2) = \frac{1}{x_2 - x_1 + 1} \sum_{x=x_1}^{x_2} \Delta m_j^t(x).
  \]  

- Third, we construct the Longevity Divergence Index Value (LDIV) at time \( t \) as
  \[
  \text{LDIV}_t = \Delta m_t^{UK}(75, 85) - \Delta m_t^{US}(55, 65).
  \]

²As 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).
The PRF is then calculated as

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

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 U.K. males aged 75 to 85 and the U.S. males aged 55 to 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 U.K. 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 U.S. 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.

![Figure 1: Historical annualized age-specific mortality improvement rates](image-url)

³The HMD mortality database can be found at [http://www.mortality.org/](http://www.mortality.org/)
Figure 2: Historical annualized average mortality improvement rates

Figure 3: Historical LDIV up to year 2009
In Figure 2, we then plot the historical annualized average mortality improvement rates $\Delta m_{t}^{UK}(75, 85)$ and $\Delta m_{t}^{US}(55, 65)$ for the U.K. and the U.S. respectively. We can see that for the U.K., 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 U.S. 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 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 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 above-mentioned references, we directly model the 8-year annualized age-country-specific mortality improvement rate calculated as $\Delta m_{t}^{j}(x)$ 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. 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

---

4For other examples that apply time series models to mortality modeling, see Engle (2001); Gao and Hu (2009); Giacometti et al. (2012); Surpong (2013); Wang et al. (2013); Lin et al. (2015); Chen et al. (2015).
of the Kortis bond. We apply the same modeling approach to $\Delta m_{t}^{\text{UK}}(75,85)$, $\Delta m_{t}^{\text{US}}(55,65)$, and $\text{LDIV}_t$. 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.\footnote{R packages \texttt{forecast} (Hyndman and Khandakar, 2008) and \texttt{rugarch} (Ghalanos, 2019) are used to select optimal models and generate forecasts.}

Table 2: Selected ARIMA-GARCH models based on the AIC

<table>
<thead>
<tr>
<th>Series</th>
<th>ARIMA</th>
<th>GARCH</th>
<th>Series</th>
<th>ARIMA</th>
<th>GARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{LDIV}_t$</td>
<td>1 0 0 1 2</td>
<td></td>
<td>$\Delta m_{t}^{\text{US}}(55,65)$</td>
<td>0 1 0 1 1</td>
<td></td>
</tr>
<tr>
<td>$\Delta m_{t}^{\text{UK}}(75,85)$</td>
<td>0 1 1 1 1</td>
<td></td>
<td>$\Delta m_{t}^{\text{US}}(55)$</td>
<td>1 1 0 0 0</td>
<td></td>
</tr>
<tr>
<td>$\Delta m_{t}^{\text{UK}}(75)$</td>
<td>0 1 1 1 1</td>
<td></td>
<td>$\Delta m_{t}^{\text{US}}(56)$</td>
<td>0 1 0 0 0</td>
<td></td>
</tr>
<tr>
<td>$\Delta m_{t}^{\text{UK}}(76)$</td>
<td>2 1 1 2 2</td>
<td></td>
<td>$\Delta m_{t}^{\text{US}}(57)$</td>
<td>0 1 0 1 1</td>
<td></td>
</tr>
<tr>
<td>$\Delta m_{t}^{\text{UK}}(77)$</td>
<td>0 1 1 1 2</td>
<td></td>
<td>$\Delta m_{t}^{\text{US}}(58)$</td>
<td>1 1 0 1 1</td>
<td></td>
</tr>
<tr>
<td>$\Delta m_{t}^{\text{UK}}(78)$</td>
<td>0 1 1 1 2</td>
<td></td>
<td>$\Delta m_{t}^{\text{US}}(59)$</td>
<td>0 1 1 1 1</td>
<td></td>
</tr>
<tr>
<td>$\Delta m_{t}^{\text{UK}}(79)$</td>
<td>0 1 1 1 1</td>
<td></td>
<td>$\Delta m_{t}^{\text{US}}(60)$</td>
<td>1 1 0 1 1</td>
<td></td>
</tr>
<tr>
<td>$\Delta m_{t}^{\text{UK}}(80)$</td>
<td>0 1 1 1 0</td>
<td></td>
<td>$\Delta m_{t}^{\text{US}}(61)$</td>
<td>0 1 3 1 1</td>
<td></td>
</tr>
<tr>
<td>$\Delta m_{t}^{\text{UK}}(81)$</td>
<td>0 1 1 1 1</td>
<td></td>
<td>$\Delta m_{t}^{\text{US}}(62)$</td>
<td>0 1 1 1 3</td>
<td></td>
</tr>
<tr>
<td>$\Delta m_{t}^{\text{UK}}(82)$</td>
<td>0 1 1 1 1</td>
<td></td>
<td>$\Delta m_{t}^{\text{US}}(63)$</td>
<td>0 1 1 2 2</td>
<td></td>
</tr>
<tr>
<td>$\Delta m_{t}^{\text{UK}}(83)$</td>
<td>0 1 1 1 1</td>
<td></td>
<td>$\Delta m_{t}^{\text{US}}(64)$</td>
<td>0 1 0 0 0</td>
<td></td>
</tr>
<tr>
<td>$\Delta m_{t}^{\text{UK}}(84)$</td>
<td>0 1 1 1 1</td>
<td></td>
<td>$\Delta m_{t}^{\text{US}}(65)$</td>
<td>0 1 1 0 0</td>
<td></td>
</tr>
</tbody>
</table>

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 2-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 U.K. and U.S. 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 U.K. and U.S. average mortality improvement rates.

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 $t$:

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

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

### 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. 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, for example Athanasopoulos et al., 2009; Capistrán et al., 2010; Borges et al., 2013; Syntetos et al., 2016; Shang and Hyndman, 2017).

---

6Since 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).
 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.

For \( t \in [1, T] \),

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

We can then link these two vectors by the equation

\[
y_t = S b_t,
\]

where \( S \) is a “summing matrix” of dimension \( 25 \times 22 \), which aggregates age-country-specific mortality rates to construct the LDIV. It is given by

\[
S = \begin{pmatrix}
1 & 1 & 1 & \ldots & \ldots & 1 & 1 & 1 \\
1 & 1 & \ldots & 1 & 0 & 0 & \ldots & 0 \\
0 & 0 & \ldots & 0 & 1 & 1 & \ldots & 1 \\
\end{pmatrix}
\begin{pmatrix}
I_{22}
\end{pmatrix},
\]

where \( I_{22} \) denotes a \( 22 \times 22 \) identity matrix. The aggregation constraints (5)–(7) are reflected by the first three rows of the matrix \( S \).

Let \( \hat{y}_{T+h} \) be a vector of \( h \)-step-ahead base forecasts of all series in the hierarchy, and \( \hat{b}_{T+h} \) 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

\[
\hat{y}_{T+h} = S P \hat{y}_{T+h},
\]


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

$$P = (0_{22 \times 3}, I_{22}),$$

(10)

where $0_{22 \times 3}$ is a zero matrix of dimension $22 \times 3$. 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

$$\hat{y}_{T+h} = S\beta_{T+h} + \epsilon_{T+h},$$

(11)

where $\beta_{T+h} = \mathbb{E}[b_{T+h}|y_1, y_2, \ldots, y_T]$ is the unknown mean of bottom-level base forecasts, and $\epsilon_{T+h}$ is the error term which has a mean vector zero and a covariance matrix $\Sigma_h$. As shown by Hyndman et al. (2011), if $\Sigma_h$ is known, we can find the generalized least squares (GLS) estimator of $\beta_{T+h}$ as

$$\hat{\beta}_{T+h} = (S^\prime \Sigma_h^\dagger S)^{-1} S^\prime \Sigma_h^\dagger \hat{y}_{T+h},$$

(12)

where $\Sigma_h^\dagger$ is the Moore-Penrose generalized inverse of $\Sigma_h$ as $\Sigma_h$ is likely to be singular.

We then obtain the reconciled forecasts by

$$\hat{y}_{T+h} = S\hat{\beta}_{T+h} = S(S^\prime \Sigma_h^\dagger S)^{-1} S^\prime \Sigma_h^\dagger \hat{y}_{T+h}.$$  

(13)

This implies that $P = (S'\Sigma_h^\dagger S)^{-1} S^\prime \Sigma_h^\dagger$. Since $SPS = S$, 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), $\Sigma_h$ is generally not known nor identifiable. Wickramasuriya et al. (2018) provided an alternative estimation of $P$ by minimizing the trace of $\text{VAR}[y_{t+h} - \hat{y}_{t+h}|y_1, y_2, \ldots, y_t]$, which is the covariance matrix of the in-sample reconciled forecast errors. This method is referred to as the MinT reconciliation approach.

Let $W_h$ be a positive definite covariance matrix of the $h$-step-ahead in-sample base forecast errors, that is,

$$W_h = \mathbb{E}[\hat{e}_{t+h} \hat{e}_{t+h}'|y_1, y_2, \ldots, y_t],$$

(14)

where $\hat{e}_{t+h} = y_{t+h} - \hat{y}_{t+h}$. One can verify that

$$\text{VAR}[y_{t+h} - \hat{y}_{t+h}|y_1, y_2, \ldots, y_t] = SPW_h P' S.$$  

(15)

The optimal reconciliation matrix is then given by

$$P = (S'W_h^{-1}S)^{-1}S'W_h^{-1}.$$  

(16)

---

7 In our analysis, although the individual series themselves are mostly non-stationary, the in-sample forecast errors we used to compute the $W_h$ matrix are in fact stationary. Consequently, the MinT approach is valid in our study.

8 For a detailed proof see Wickramasuriya et al. (2018), Appendix A.1.
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 $P$. However, in the GLS solution, $\Sigma_h$ 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

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. (2018). Both works aim to adjust probabilistic forecasts to ensure that the aggregation constraints are met. In particular, Jeon et al. (2018) are the first to utilize information on full probabilistic distributions of all levels to produce the reconciled probabilistic forecasts. Moreover, Jeon et al. (2018) provided a generalization of the MinT approach.

Following Jeon et al. (2018), we first define the following terms:

- Let $f(y_{T+h}|y_1, y_2, \ldots, y_T)$ be the probabilistic distribution of $h$-step-ahead forecasts;
- Let $\hat{y}_{T+h}^i$ denote the $i^{th}$ sample of base forecasts generated from $f(y_{T+h}|y_1, y_2, \ldots, y_T)$;
- A sample of size $N$ generated from $f(y_{T+h}|y_1, y_2, \ldots, y_T)$ is denoted by $\hat{Y}$, where $\hat{Y} = (\hat{y}_{T+h}^1, \hat{y}_{T+h}^2, \ldots, \hat{y}_{T+h}^N)$.

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

$$\tilde{Y} = S P \hat{Y},$$

where $S$ and $P$ are the same as defined in Section 3.110, and $\tilde{Y}$, 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 $\tilde{Y}$. The problem is then down to how to construct unreconciled

9There are several choices of the estimator of $W_h$ proposed in Wickramasuriya et al. (2018), Section 2.4. In this work, we have used the “shrinkage” estimator as described there.

10Note that for the bottom up reconciliation, $P = (0_{22 \times 3}, I_{22})$, which is the same as in Section 3.1.
forecasts $\hat{Y}$. To this end, Jeon et al. (2018) introduced three schemes, namely the “stacked sample”, “permuted sample” and “ranked sample”. The empirical results in Jeon et al. (2018) 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 $\hat{Z}_m$ to be an $N \times 1$ column matrix, which contains base forecasts of size $N$ for the $m^{th}$ series in $y_{T+h}$. For example, $\hat{Z}_1$ represents $N$ samples generated from the distribution $f(LDIV_{T+h}|LDIV_1, LDIV_2, \ldots, LDIV_T)$.

- Arrange the elements in $\hat{Z}_m$ in ascending order to form $\hat{Z}_m^R$.

- Let $\hat{Y}^R = (\hat{Z}_1^R, \hat{Z}_2^R, \ldots, \hat{Z}_{25}^R)'$ be the ranked sample of size $N$ for all series.

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

$$\tilde{Y} = SP\hat{Y}^R. \quad (18)$$

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.\footnote{The MinT reconciliation method is implemented by R package \texttt{hts} (Hyndman et al., 2018).} For comparison, we also include the results from the RMS, the base forecasts, the bottom up approach, and the optimal combination GLS approach.

![Figure 5: Density of the LDIV in 2016](Image)
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.

<table>
<thead>
<tr>
<th>LDIV≥</th>
<th>PRF≥</th>
<th>Estimated probability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Bottom up</td>
</tr>
<tr>
<td>3.40%</td>
<td>0%</td>
<td>29.32%</td>
</tr>
<tr>
<td>3.50%</td>
<td>20%</td>
<td>27.51%</td>
</tr>
<tr>
<td>3.60%</td>
<td>40%</td>
<td>25.76%</td>
</tr>
<tr>
<td>3.70%</td>
<td>60%</td>
<td>24.09%</td>
</tr>
<tr>
<td>3.80%</td>
<td>80%</td>
<td>22.51%</td>
</tr>
<tr>
<td>3.90%</td>
<td>100%</td>
<td>20.98%</td>
</tr>
</tbody>
</table>

| 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 percent 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.
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
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
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, 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 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.

![Figure 10: Indicated spread of the Kortis bond](image-url)

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 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.\(^{13}\)

\(^{12}\)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.
Table 4: Estimated Quantiles of the PRF with additional mortality information

<table>
<thead>
<tr>
<th>LDIV ≥</th>
<th>PRF ≥</th>
<th>Estimated probability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>2010</td>
</tr>
<tr>
<td>3.40%</td>
<td>0%</td>
<td>4.34%</td>
</tr>
<tr>
<td>3.50%</td>
<td>20%</td>
<td>3.58%</td>
</tr>
<tr>
<td>3.60%</td>
<td>40%</td>
<td>2.77%</td>
</tr>
<tr>
<td>3.70%</td>
<td>60%</td>
<td>2.14%</td>
</tr>
<tr>
<td>3.80%</td>
<td>80%</td>
<td>1.63%</td>
</tr>
<tr>
<td>3.90%</td>
<td>100%</td>
<td>1.25%</td>
</tr>
</tbody>
</table>

Expected loss 2.59% 2.78% 1.00% 0.98% 1.47% 1.35%
Conditional expected loss 60% 56% 54% 76% 69% 50%

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.

Figure 11: Historical LDIV up to year 2016
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 purple 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.

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 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 Table 5 and 6 in the Appendix, with bold figures highlighting the method that provides the best forecast performance.

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. (2018) 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.
References


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 Financial LLC.


Appendix

For each individual series, the root mean squared forecast error (RMSFE) is defined as

\[
RMSFE = \sqrt{\frac{1}{h} \sum_{k=1}^{h} (m_{T+k} - \hat{m}_{T+k})^2},
\]

(19)

where \(h\) represents the length of the forecast horizon, \(m_{T+k}\) represents the \(k\)th 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

\[
\text{Average RMSFE} = \frac{1}{n_j} \sum_{i=1}^{n_j} \text{RMSFE}_i,
\]

(20)

where \(n_j\) represents the number of individual series at level \(j\) in the hierarchy, and \(\text{RMSFE}_i\) represents the RMSFE of individual series \(i\) at level \(j\).

Table 5: RMSFE (×100) of out-of-sample forecasts during 2010–2016

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

Table 6: Average RMSFE (×100) of out-of-sample forecasts during 2010–2016

<table>
<thead>
<tr>
<th>Level</th>
<th>Bottom up</th>
<th>Base</th>
<th>GLS</th>
<th>MinT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top</td>
<td>0.38</td>
<td>0.81</td>
<td>0.53</td>
<td>0.34</td>
</tr>
<tr>
<td>Middle</td>
<td>1.02</td>
<td>0.94</td>
<td>1.03</td>
<td>0.92</td>
</tr>
<tr>
<td>Bottom</td>
<td>NA</td>
<td>1.26</td>
<td>1.14</td>
<td>1.05</td>
</tr>
</tbody>
</table>
Han Li
Department of Actuarial Studies and Business Analytics
Macquarie University
Sydney, NSW 2109, Australia
Email: han.li@mq.edu.au

Qihe Tang
School of Risk and Actuarial Studies
UNSW Sydney
Sydney, NSW 2052, Australia
Email: qihe.tang@unsw.edu.au

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