Canadian Institute of Actuaries | Institut canadien des actuaires

Research paper

Common Mortality Trend Model and Mortality Prediction

Liqun Diao, Yechao Meng, Chengguo Weng and Tony Wirjanto, University of Waterloo

December 2021

Document rp221138

Ce document est disponible en français

Abstract

In the paper, we propose a framework to allow populations at disparate mortality development stages to be contemporaneously taken into account through a bivariate-population mortality system with an underlying process of common mortality trend (CMT). As one of the main contributions of the paper, the CMT model is used as a base learner in a computationally friendly framework to effectively borrow information from multiple populations for the mortality prediction of a target population through a well-established procedure of model averaging. Empirical studies with the Human Mortality Database confirm that the proposed CMT-based prediction framework yields a substantial improvement in prediction performance over a set of benchmark prediction models. Furthermore, these empirical studies reveal that a key parameter introduced in the CMT model is able to characterize the development stage of one population relative to the other in the CMT model.

Keywords: Mortality Prediction; Bivariate-Population Model; Human Mortality Database; Model Averaging; Borrowing Information.

1 Introduction

Longevity risk, attributed to the increase in human life expectancy, has been recognized as one of the major risks faced by insurers, governments, and individuals. The effect of longevity risk tends to be systematic in nature, and it has created a substantial financial pressure on pension funds and annuity providers. Mortality modelling is a critical component to management of longevity risk. The past decades have witnessed a steady proliferation of various mortality models introduced in the literature. The Lee-Carter model (Lee and Carter, 1992) has become the benchmark model for mortality modelling and forecasting of a single population. This model decomposes age-specific mortality rates over a certain time period into the mean age-specific mortality rates, the mortality trend, the amount of mortality changes at a given age, and a noise term.

As an important prototype of various coherent multi-population mortality forecasting models, the Augmented Common Factor (ACF) model proposed by Li and Lee (2005) extends the Lee-Carter model to the case of multiple populations. Coherent multi-population mortality forecasting models were designed to avoid unrealistic crossovers or divergence in anticipated future mortality across countries or between genders, which could arise from applying a single-population model to each population separately. The main idea behind coherent forecasting is that mortality forecasts for populations with similar mortality developments are not expected to diverge substantially even though structural differences remain across populations. Further development of coherent multi-population mortality models includes Li and Hardy (2011); Russolillo et al. (2011); Li (2013); Wang and Yang (2013); Zhou et al. (2014); Kleinow (2015) among others. Another strand of potent literature on multi-population mortality modelling is based on an age-period-cohort (APC) structure of mortality rates, a time-honoured methodology in epidemiology. The representative literature from this strand includes Dowd et al. (2011); Janssen et al. (2013). The basic idea of the APC-structure-based models is to describe a mathematical relationship between death rates and attained age, and that between calendar period and birth cohort. Furthermore, a functional data approach has also been adopted recently for multi-population mortality modelling; for example, Hyndman et al. (2013); Shang and Hyndman (2017). Moreover, Li et al. (2015) proposed bivariate-population generalizations for single-population models M1-M3 and M5-M8 from Cairns et al. (2009).

It has been widely documented that coherent mortality forecasting models are generally superior to individual mortality forecasting models, but multi-population models have so far been applied only to a group of populations in similar mortality development stages in the sense that they have similar trajectories or patterns of mortality development; see, for example, Li and Lee (2005); Li et al. (2017).

In this paper, we propose a framework that describes a bivariate-population mortality system which allows two populations at disparate mortality development stages to be contemporaneously taken into account. The model is developed based on the rationale that various populations enjoy certain commonality in their respective trajectories of evolving over time, which has been widely documented in empirical analysis of mortality data. According to Zhou et al. (2014), some populations dominate while other populations follow its mortality dynamics. For instance, a developing country may demonstrate a similar mortality

improvement pattern in the recent decade to what a developed country has experienced earlier in 1990s. Additionally, as mentioned in Section 5.4 of Li et al. (2015), for two populations, the period effect estimates, which capture the mortality developing trend in time, usually roughly co-move with one another, although they are located at different absolute levels. Due to this similarity in the trajectories and patterns of their mortality rates, populations at different development stages can still bring in useful information to enhance the performance of mortality prediction, with a well-designed mortality model that is capable to capture both the co-movement and the delay in time.

In our framework, an underlying process of common mortality trend (CMT) is specified for two populations at possibly different stages in their journeys of the mortality development to characterize the commonality in the evolution of human societies. For the purpose of illustration, we focus on two specific formulations, denoted as CMT-ACF and CMT-CBD (Cairns-Blake-Dowd) models respectively corresponding to the ACF model and the two-population version of CBD model in Cairns et al. (2006) (i.e., model M5 in Cairns et al. (2009)). As one of the main contributions of the paper, we exploit the bivariate-population CMT model as a base learner and develop a computationally friendly framework which effectively borrow information from multiple populations for the prediction of a target population. The framework is based on an innovative application of the model averaging idea from statistics, where bivariate-population models are developed between the target population and each population from a candidate pool, and a subset of these bivariate-population models are selected and aggregated to predict mortality for the target population.

We apply our model averaging prediction method to mortality data of 24 populations of both genders from the Human Mortality Database (HMD), which is available on the website https://www.mortality.org/ in two separate empirical studies. In the first empirical study, we take the CMT-ACF model as the base learner and apply our method for predicting mortality rates of ages 0-100. In the second empirical study, we take the CMT-CBD model as the base learner because we aim at predicting mortality rates only for senior ages 55-90 and it is well known that the CBD model has the superiority in prediction accuracy for senior ages compared with other models. Both empirical studies confirm that our proposed CMT-based method yields substantially better prediction results than several benchmark models considered in our study. Besides, the framework also provides useful insights in terms of interpreting current mortality data through the introduction of parameter, denoted as Δt. This parameter, designed to characterize the "stage of evolution" of one population to the other in the CMT process, can capture the relative mortality level of the target population in most cases under our empirical studies. The degree of correlation between the Δt value and the mortality level depends on the degree of dominance by the CMT process over the population-specific effects in the CMT model. When the CMT process dominates the mortality development, we expect a strong negative correlation between the Δt value and the mortality level of a target population and vice versa. We design a Relative Scale (RS) measure to assess the degree of dominance by the common trend process in a CMT model. We observe a strong negative correlation between the Δt value and mortality level (proxied by the aggregate logarithmic mortality rates over all the involved ages and the time period of training data) over populations with a small RS measure.

In particular, the RS measure is small for all the CMT-CBD models involved in the second empirical study, and a Pearson correlation coefficient between the Δt value and the mortality level is observed as high as -0.896 and -0.971 for female populations and male populations respectively. Therefore, the Δt parameter in the CMT-CBD model provides a simple but informative index to describe how many years one population is in advance of the other in their mortality development.

The rest of the paper proceeds as follows. Section 2 provides a general setup of our bivariate-population CMT model and describes two specific cases of the model. Section 3 introduces various model averaging procedures which we develop for mortality prediction. Section 4 presents empirical studies of our proposed mortality prediction method and compares with benchmark models considered in our study by analyzing data from the HMD. Finally, Section 5 provides concluding remarks and some discussions on future research.

2 Bivariate-population CMT Model

2.1 General Model Setup

$\eta_j(x,t)$

$$t = 0, 1, \ldots, T,\ x = x_1, \ldots, x_n$$

$$\eta_1(x,t) = F(x,t) + f_1(x,t), +\epsilon_1(x,t), \tag{1}$$
$$\eta_2(x,t) = F(x, t-\Delta t) + f_2(x,t) + \epsilon_2(x,t), \tag{2}$$

$t = 0, 1, \cdots, T,\ x = x_1, \ldots, x_n,\ F(x,t)$

Δt

$f_1(x,t)$ $f_2(x,t)$

$\epsilon_1(x,t)$ $\epsilon_2(x,t)$

Some further elaboration of the CMT model is given below:

$F(x,\cdot)$ (1) and (2) highlights the salient feature of the CMT model and amounts to the key assumption of the model. The underlying CMT process drives the mortality development of both populations while they may be at different stages in their own journeys.

Δt

Δt Δt

Δt

Δt

c) The CMT model specified in equations (1) and (2)

$f_1(x,t)$

$f_2(x,t)$

$F(x,\cdot)$

Δt

$\eta_j(x,t),\ F(x,t)$

$f_j(x,t)$ Δt

Δt

Li and Lee (2005) and the CBD model (Model M5 in Cairns et al., 2009)

$\eta_j(x,t),\ F(x,t)$

$f_j(x,t)$

e) In the literature, multivariate time series models, such as the VAR model or VECM model, have been introduced to model the joint evolution of mortality for multiple populations over time; see, e.g., Zhou et al. (2014). The cross-correlation between two populations' overall mortality improvement is implicitly modelled and captured by corresponding parameters in a VAR/VECM model. In contrast, our CMT model captures the co-movement in a different way. It explicitly decomposes the overall mortality improvement into an underlying common process shared by both population and a specific process for each population. Meanwhile, we provide a transparent view on which population is at a more developed stage of morality than the other in the model through the parameter Δt. Based on our empirical study (see Section 4), the calibrated value of Δt characterizes the relative position in mortality development across populations well.

2.2 CMT-ACF Model

2.2.1 Specification of CMT-ACF Model

$t = 0, 1, \ldots, T,\ x = x_1, \ldots, x_n.$ $\eta_j(x,t) = \log m_j(x,t)$

$$F(x,t) = B(x)K(t),$$
$$f_j(x,t) = a_j(x) + b_j(x)k_j(t),$$

in the CMT model in equations (1) and (2), we have the following model specification:

$$\log m_1(x,t) = a_1(x) + B(x)K(t) + b_1(x)k_1(t) + \epsilon_1(x,t), \quad (3)$$
$$\log m_2(x,t) = a_2(x) + B(x)K(t - \Delta t) + b_2(x)k_2(t) + \epsilon_2(x,t). \quad (4)$$

The above specification yields a similar model to a bivariate-population ACF model in Li and Lee (2005) Δt

(3) and (4)

$b_j(x)$

$k_j(t)$

$$\sum_{l=1}^{n} B(x_l) = 1,\ \sum_{t\in\mathbb{S}} K(t) = 0,\ \sum_{l=1}^{n} b_j(x_l) = 1,\qquad \sum_{t=0}^{T} k_j(t) = 0$$

are imposed to avoid issues of unindentifiability for the model, where

$$\mathbb{S} = \begin{cases} \{-\Delta t, -\Delta t + 1, \ldots T\} & \text{if } \Delta t \geq 0; \\ \{0, 1, \ldots, T - \Delta t\}, & \text{otherwise.} \end{cases}$$

2.2.2 Calibration and Extrapolation of CMT-ACF Model

Δt 2.4

Δt

Δt Δt Δt Δt

$E_j(x,t)$

$$\frac{E_1(x,t)m_1(x,t) + E_2(x,t)m_2(x,t)}{E_1(x,t) + E_2(x,t)}.$$

Δt Δt

Δt Δt

Δt

Δt

Δt

Δt

Δt

Δt

Δt

- over (*T, T* — Δt], it only contains data from the reference (*j* = 2).

1

Δt

Δt Δt Δt Δt Δt

$$R_1(x,t) = [\log m_1(x,t) - B(x)K(t)]$$

$R_2(x,t) = [\log m_2(x,t) - B(x)K(t - \Delta t)]$ $a_j(x)$

$$a_j(x) = \frac{1}{T+1}\sum_{t=0}^{T} R_j(x,t), \quad j = 1, 2, \quad \text{and } x = x_1, \ldots, x_n. \qquad (5)$$

4. For each *j* = 1, 2, we apply a SVD procedure further to $R_j(x,t) - a_j(x)$ to get $b_j(x)$ and $k_j(t)$.

5. Fit each of the obtained sequences *K(t)* and $k_j(t)$ with an AutoRegressive Integrated Moving Average (ARIMA) model. In our study, we apply the **auto.arima** function from the R package **forecast** to each sequence.

6. For prediction, we follow the conventional extrapolation paradigm. We obtain predictions for *K(t)* and $k_1(t)$ over future years t and then use equation (3) to obtain forecasts by taking the white noise terms as zero. The prediction of $k_1(t)$ is obtained by projecting its established time series model into the future. The prediction of *K(t)* is obtained either from the calibration step or by an extrapolating procedure, depending on value of Δt. If $\Delta t \geq 0$, the aggregated data from Step 1 span over the time horizon $[-\Delta t, T]$ and in this case, the prediction of *K(t)* for any future year $t > T$ should be obtained via extrapolating the established time series model. In contrast, if $\Delta t < 0$, the aggregated data from Step 1 span over the time horizon $[0, T - \Delta t]$, where we note $T - \Delta t > T$. In this case, we directly input the calibrate values of *K(t)* obtained in Step 2 into equation (3) for mortality forecasts over the period $[T, T - \Delta t]$, and apply the extrapolation procedure for prediction beyond time $T - \Delta t$.

2.3 CMT-CBD Model

$q_j(x,t)$

$t = 0, 1, \ldots, T,\ x = x_1, \ldots, x_n$

(1) and (2) with

$$\eta_j(x,t) = \operatorname{logit} q_j(x,t) \equiv \log\left[\frac{q_j(x,t)}{1-q_j(x,t)}\right],$$
$$F(x,t) = K(t),$$
$$f_j(x,t) = (x-\bar{x})k_j(t),$$

$$\log\left[\frac{q_j(x,t)}{1-q_j(x,t)}\right] = K(t) + (x-\bar{x})k_1(t) + \epsilon_1(x,t), \quad (6)$$
$$\log\left[\frac{q_j(x,t)}{1-q_j(x,t)}\right] = K(t-\Delta t) + (x-\bar{x})k_2(t) + \epsilon_2(x,t), \quad (7)$$

where $j = 1, 2,\ t = 0, 1, \cdots, T,\ x = x_1, \ldots, x_n$, and $\bar{x} = \sum_{l=1}^{n} x_l/n$. Each marginal of the above model has the same structure as the original CBD model in Cairns et al. (2006), or model M5 in Cairns et al. (2009) with $B(x) = 1$ and $b_j(x) = x - \bar{x}$ and no static age function nor cohort effect. We refer to the specification in equations (6) and (7) as a CMT- CBD model. This specification extends the CBD model to the bivariate-population scenario where the age-period term $K(t)$ is shared by both populations but with a difference in their development stages.

2.3.2 Calibration and Extrapolation of CMT-CBD Model

Δt

2.4. The model extrapolation can be obtained by following the same Steps 5 and 6 for the CMT-ACF model as described in Section 2.2.2. Below we describe the procedure of calibrating the CMT-CBD model with a fixed Δt. It involves the following steps:

$d_j(x,t)$ $E_j(x,t)$

$q_j(x,t),$ $d_j(x,t) \sim Bin\big[E_j(x,t), q_j(x,t)\big]$

$d_j(x,t)$ $E_j(x,t)$

$$\ell = \sum_{j=1}^{2}\sum_{l=1}^{n}\sum_{t\in\mathbb{S}} \ell\big(d_j(x_l,t), q_j(x_l,t)\big) \qquad (8)$$

where

$$\ell\big(d_j(x,t), q_j(x,t)\big) \quad \propto \quad d_j(x,t)\log\big[q_j(x,t)\big] + \big[E_j(x,t) - d_j(x,t)\big]\log\big[1-q_j(x,t)\big].$$

3. From equations (6) and (7), we have the following two expressions:

$$q_1(x,t) = \text{expit}\,\big[K(t) + (x-\bar{x})k_1(t)\big], \qquad (9)$$
$$q_2(x,t) = \text{expit}\,\big[K(t-\Delta t) + (x-\bar{x})k_2(t)\big], \qquad (10)$$

4. We calibrate the $K(t)$ and $k_j(t)$ sequences by plugging the forms of $q_j(x,t)$ in (9) and (10) into the log-likelihood (8) and then maximizing it with respect to $K(t)$and $k_j(t)$ using the Newton-Raphson iterative procedure.

2.4 Choice of Δt

Δt

Δt

Δt

Δt Δt

$\eta_j(x,t)$

$\log m_j(x,t)$ in the CMT-ACF model and $\text{logit} q_j(x,t)$ in the CMT-CBD model, using validation data from both the target and the reference populations.

In the analysis of HMD in Section 4 Δt

Δt 4, the CMT model is applied to many pairs out of 30 populations and we universally adopt the set {-10, -9, . . . , 9, 10} as the candidate set for Δt to make the prediction procedure fully driven by data instead of our subjective judgment on the development stages of populations.

3 Model Averaging Method for Prediction

As an application, we fit the CMT model with HMD and study its potentials in enhancing the performance of mortality prediction. The HMD contains data of human mortality rates for 47 countries or regions.

The proposed CMT model in Section 2

$(n-1)\ \Delta t$

Δt

In view of the computational hurdle in calibrating a multi-population CMT model, we instead propose a model averaging procedure for borrowing information. The model averaging procedure allows us to exploit the proposed bivariate-population CMT models as the building blocks and form an enhanced prediction rules with information borrowed from multiple populations. More specifically, given a target population and a pool of reference populations, we build bivariate-population CMT models between the target population and each reference population in the pool. Then, we obtain extrapolative results (i.e., prediction) on future mortality rates for the target population out of each bivariate-population model and form a final prediction by averaging the extrapolative results from these bivariate-population CMT models.

The primary idea of model averaging is to reduce the prediction uncertainty by aggregating predictive results from multiple predictive models. This idea has been widely adopted in the machine learning community. The following are several specific strategies in utilizing the model averaging idea. The last strategy, RankAvg, provides an adaptive selection of individual predictive rules used in the model averaging and is expected to have a robust performance over different numerical settings. The other three strategies are proposed for comparison.

- **Simple Average (SimAvg):** This is the most naive strategy where we take all accessible information into account. Each reference population in the pool is used to build a bivariate-population CMT model with the given target population.

The final mortality prediction for the target population is just the simple average of predictions from all of these bivariate-population models.

- **Average Based on Geographic Information (GeoAvg):** We use geographic proximity as exogenous information to pre-specify groups. For a given target population, the final prediction is the average of predictions from the CMT models that are calibrated with reference populations within the same geographic group.
- **Average Based on Clustering Results (KmeansAvg):** This is a data-driven strategy using, for example, cluster analysis, to find populations with similar mortality characteristics (see Hatzopoulos and Haberman (2013)). We apply a clustering algorithm to group a given set of populations. The final prediction for the target population is obtained as the average of the predictions from the bivariate CMT models built between the target and each of the reference population located in the same cluster as the target stays.

$$\eta_1(x,t)$$

(2.4) for the choice of Δt. We next take the average of predictions from the CMT models that yield smallest SSEs on the same validation data set and use it as the prediction for the mortality of the target population. The prediction results certainly depend on the number of top CMT models that we use in the average, which is decided by a validation procedure.

4 Empirical Analysis

In this section, we empirically evaluate the performance of our proposed prediction methods via analysis of the HMD and compare them to several classic benchmark models. In Section 4.1, we describe the choice of populations we use in our analysis. In Section 4.2, we conduct a study on the performance of model averaging based on the CMT-ACF models and consider the mortality prediction for a full range of ages 0-100. In Section 4.3, we switch our focus of analysis to the mortality prediction for senior age group, ages 55-90, and study the model averaging strategies based on the CMT-CBD models since the CBD model is known for its superior performance in characterizing the mortality development for seniors. In both studies, we provide the predictions for the combined target (total population) and the gender-specific target (female and male populations).

4.1 Data Description

We start from Diao et al. (2020)

$$\Delta t$$

According to Brainerd and Cutler (2005), demographic disasters in the form of sharply rising death rates happened among several member countries of the former Soviet Union. As we have noticed that populations may demonstrate different types of development patterns over time sometimes and lead to violations of the assumption that commonality trend exists among populations across time, we want to conduct pre-analysis for the guidance of population choices.

We first illustrate the age-aggregated logarithmic mortality sequence from each of the 30 populations on the left panel of Figure 1. We then conduct a k-means cluster analysis to detect structural dissimilarity among sequences of the age-aggregated logarithmic mortality. The k-means cluster procedure is designed to exclude the effect of mortality level on the dissimilarity among different populations. Below are the details of the k-means procedure:

$$D_{i,j}$$

$$\mathrm{diff}_{i,j}(t) = \sum_x \log_i(x,t) - \sum_x \log_j(x,t)$$

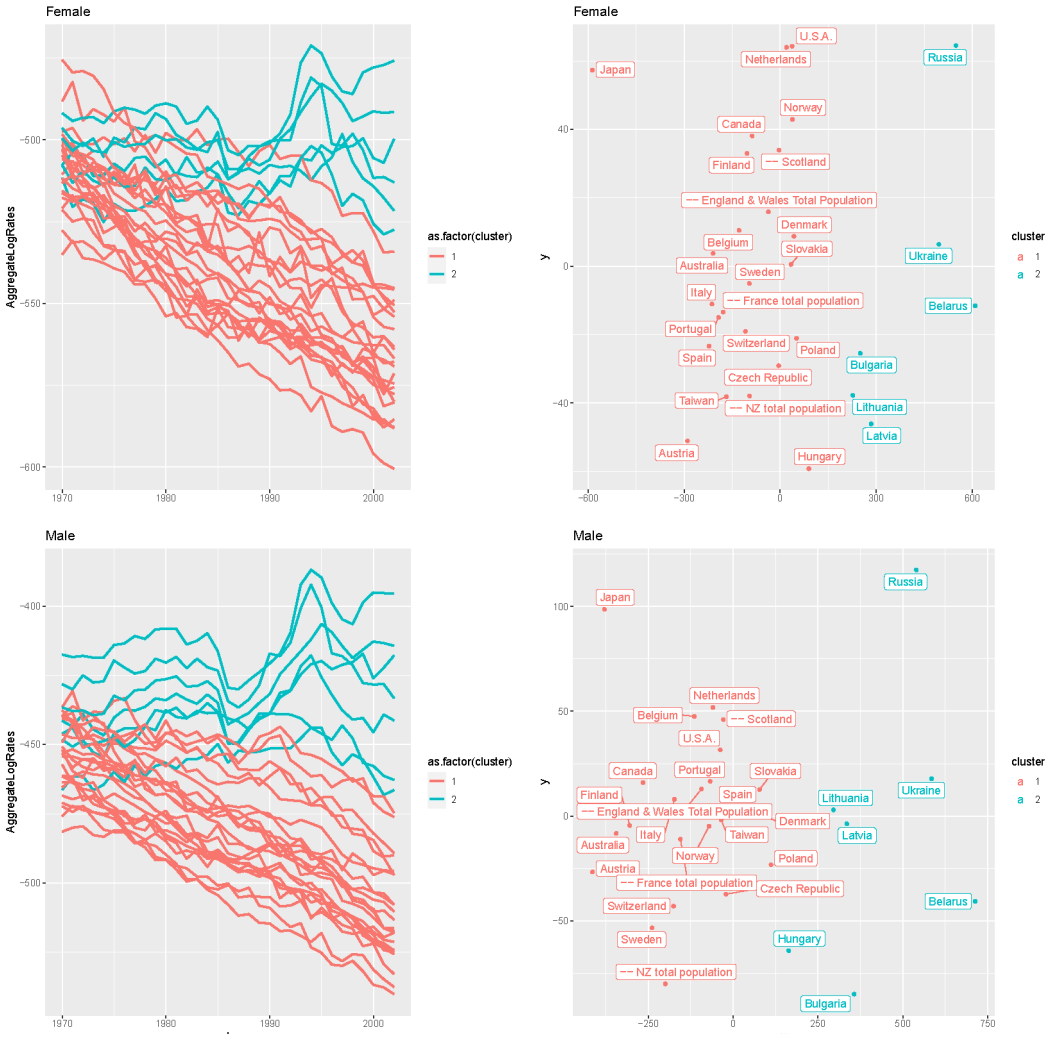
2. We then use multi-dimensional scaling (MDS) as a tool to transform the matrix D into 2-dimension. According to Cox and Cox (2008)

$$D_{i,j}$$

3. A k-means cluster analysis is then conducted on the transformed output of MDS.

It is worth noting that the variance is used in the above k-means procedure so that the disparity in mortality level would not contribute to the dissimilarity across different populations. The results of clustering are illustrated on the right panel of Figure 1. From Figure 1, we can see that most of the former member countries from the Soviet Union have a different mortality development pattern compared with the others, and they lie in a separate cluster (in blue) far away from the others (in red). Since the “outlier” populations share little commonality with the remaining populations, we exclude data from Belarus, Bulgaria, Latvia, Lithuania, Russia and Ukraine from our study, leaving us only with mortality data available from 1970 to 2010 from 24 populations as listed in Table 1, with their geographic group pre-specified as in Diao et al. (2020).

Figure 1: Clustering results of age-aggregated logarithmic mortality rates based on MDS (from top to bottom: female and male populations)



4.2 Empirical Study with the CMT-ACF Model

In this empirical study, we investigate the performance of the model averaging method with the CMT-ACF models as the building blocks. We study CMT-ACF models under various model averaging strategies and compare their performance with the classic ACF models in predicting the mortality rates of ages 0-100.

4.2.1 Prediction Models

In this empirical study, we take each of the 24 populations in Table 1 as the target and obtain its predicted value through one of the model averaging strategies as proposed in Section 3. More specifically, for a given target, we develop a bivariate CMT model with each of the remaining 23 populations, resulting 23 prediction rules. The model averaging method takes the predicted values generated from all or a subset of these 23 prediction rules and compute an

average as the final prediction. We repeat the model formulation, model averaging and prediction for each of the 24 populations in Table 1. We repeat the analysis over the 24 populations for each gender (female and male) separately.

Table 1: 24 populations from HMD and their pre-specified geographic groups

Target Population	Geographic Group	Target Population	Geographic Group
Australia	Oceania	Netherlands	West Europe
Austria	West Europe	New Zealand	Oceania
Japan	Asia	Norway	Scandinavia
Belgium	West Europe	Poland	East Europe
Scotland	Great Britain	Portugal	South Europe
Canada	North America	U.S.A.	North America
Czech Republic	East Europe	Slovakia	East Europe
Denmark	Scandinavia	Spain	South Europe
Finland	Scandinavia	Sweden	Scandinavia
France	West Europe	Switzerland	West Europe
Hungary	East Europe	Taiwan	Asia
Italy	South Europe	England & Wales	Great Britain

The four proposed averaging strategies in Section 3 are now embedded with CMT-ACF model and labeled as follows:

- **CMT-ACF-SimAvg:** The final prediction is a simple average of predicted values from all of the 23 bivariate CMT-ACF models.
- **CMT-ACF-GeoAvg:** The final prediction is an average of predicted values from the CMT-ACF models built with the reference from the same geographic region as the target; see Table 1 for the geographic grouping information.
- **CMT-ACF-KmeansAvg**: A *k*-means clustering algorithm (with 8 clusters and 1,000 independent random initializations) is applied to the logarithmic mortality rates of the 30 populations. This yields 8 clusters of populations. The final prediction is an average of the predicted values from the CMT-ACF models with the reference population from the same cluster as the target population.
- **CMT-ACF-RankAvg:** The final prediction is an average of predicted values from the *s* "top-ranked" CMT models. The value *s* is determined in the following way:
    1. We rank the 23 bivariate-population CMT models according to the validation SSEs for logarithmic mortality rates in an ascending order;
    2. We form a prediction rule by averaging over the first *u* CMT modes from the

ordered list, for u = 1, 2,..., 23.

3. We calculate the validation SSE for each of the 23 prediction rules formed in the preceding step and take s as the one that yields the smallest validation SSE.

Remark 1. *(a) It is also worth noting that the KmeansAvg strategy is an application of the simple k-means clustering method to the logarithmic mortality rates. Unlike to the k-means clustering based on a MDS procedure, the k-means clustering results in the KmeansAvg strategies are affected by the mortality level of each population. In other words, similarity in mortality level is one of the factors (in addition to the development patterns of mortality) that are conducive to assigning two populations into the same cluster.*

(b) Among the above four averaging strategies, we expect GeoAvg and KmeansAvg generally not to perform as well as SimAvg and RankAvg in the current empirical study for the following reasons. First, our preanalysis in Section 4.1 has excluded those "outlier" populations from the analysis. Those 24 populations remaining in the analysis show similar patterns of mortality development and it tends to be the case that a CMT model is viable for each pair of them. Second, both the GeoAvg and the KmeansAvg strategies only borrow information from a small subset of the population pool, and do not make use of any information out of the subsets that could be helpful. This contradicts the core idea of our CMT model to allow borrowing information from populations at disparate mortality development stages in addition to those at similar development stages.

For comparison, we consider the following benchmark methods:

- **Lee-Carter:** The Lee-Carter model is fitted to each population separately. The sequence of kt obtained from a SVD procedure is fitted using the auto.arima function from the R package forecast to search for a suitable ARIMA model.
- **ACF·AIO:** The ACF model is fitted to the 24 populations jointly.
- **ACF·GeoInfo:** The ACF model is fitted to each geographic groups in Table 1.
- **ACF·kmeans:** The ACF model is fitted to each cluster from the k-means clustering algorithm.

These benchmarks are calibrated using training set (1970-2002) and the resulting models are extrapolated to the testing period (2003-2010) for mortality prediction.

4.2.2 Prediction Performance

Evaluation Based on Test SSEs

$e(t)$

$$e(t) = \sum_{x=0}^{100} [\log m_1(x,t) - \log \widehat{m_1}(x,t)]^2, \quad t = 2003, \ldots, 2010, \tag{11}$$

$m_1(x,t)$ $\widehat{m_1}(x,t)$

$\sum_{t=2003}^{2010} e(t)$

2. Generally speaking, Table 2 shows that our proposed CMT-based prediction models, in particular CMT-ACF·RankAvg and CMT-ACF·SimAvg, stably generate smaller test SSEs and substantially outperform the benchmark models. The performance of CMT- ACF·GeoAvg and CMT-ACF·KmeansAvg, though not performing as well as the other two CMT models, is competitive with the benchmark methods.

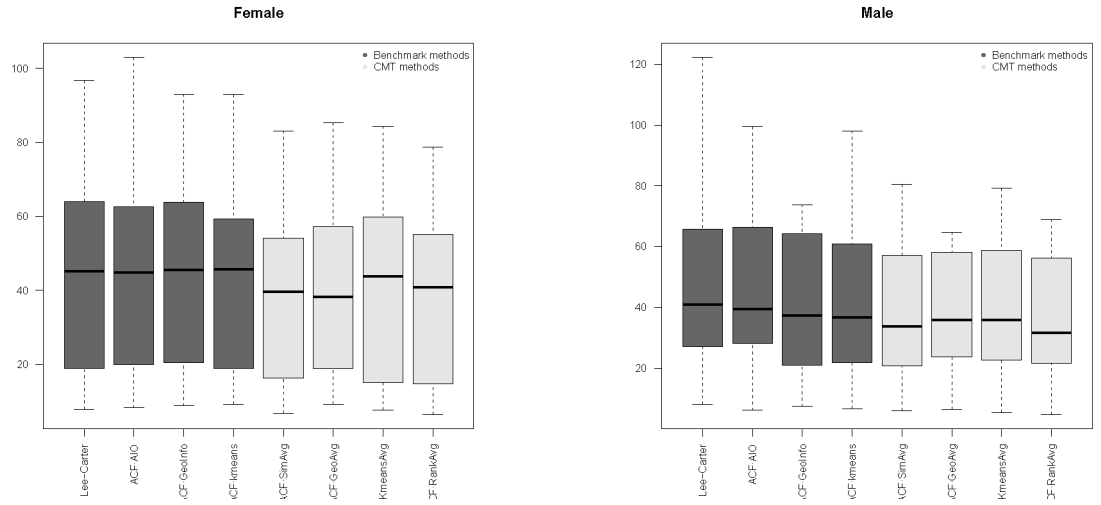
Table 2: Summary statistics of test SSEs of 24 female populations and 24 male populations comparing the prediction performance of CMT-ACF models versus benchmark methods

	1st Quartile	Median	Mean	3rd Quartile
Female Population				
Lee-Carter	19.24	45.11	44.04	63.86
ACF·AIO	20.49	44.91	43.76	62.09
ACF·GeoInfo	20.95	45.49	42.96	63.77
ACF·kmeans	18.99	45.77	41.63	59.19
CMT-ACF·SimAvg	16.59	39.65	37.78	54.00
CMT-ACF·GeoAvg	19.43	38.21	38.95	56.67
CMT-ACF·KmeansAvg	15.11	43.79	39.51	59.47
CMT-ACF·RankAvg	15.04	40.84	37.34	54.73

Male Population				
Lee-Carter	28.28	40.85	49.07	65.29
ACF‧AIO	28.28	39.54	46.90	65.76
ACF·GeoInfo	23.75	37.43	41.41	64.02
ACF·kmeans	24.35	36.70	42.31	58.82
CMT-ACF·SimAvg	23.00	33.67	36.66	56.69
CMT-ACF·GeoAvg	26.74	35.76	37.97	57.63
CMT-ACF·KmeansAvg	24.80	35.82	39.69	58.38
CMT-ACF·RankAvg	24.40	31.61	35.58	56.11

Figure 2 shows the boxplots of test SSEs of 24 populations obtained under various prediction models. Figure 2 contains two subfigures corresponding to the results for the female and male populations from left to right, respectively. In each subfigure, the order of the boxplots follows four classic benchmark models (in dark grey), Lee-Carter, ACF·AIO, ACF·GeoInfo, and ACF·kmeans, and four CMT models (in light grey), CMT- ACF·SimAvg, CMT-ACF·GeoAvg, CMT-ACF·KmeansAvg, and CMT-ACF·RankAvg. As shown in Figure 2, the CMT-based models consistently yield smaller median, fewer outliers, and smaller variation of the resulting test SSEs. The improvement is more substantial for CMT-ACF·RankAvg and CMT-ACF·SimAvg. The results reinforce our conclusion that the CMT-based models are capable to achieve superior performance in mortality forecasting. Furthermore, the relative underperformance of CMT-ACF·GeoAvg and CMT- ACF·KmeansAvg compared with CMT-ACF·RankAvg and CMT-ACF·SimAvg reinforces our comments in Remark 1.

Figure 2: Boxplots of test SSEs of 24 female populations and 24 male populations (from left to right) comparing the prediction performance of CMT-ACF models versus benchmark ACF models



Evaluation Based on One-sided Diebold-Mariano Tests

We also conduct formal hypothesis tests to determine if the test SSE of a given model is statistically significantly smaller than that of another. The one-sided Diebold-Mariano (DM) test is regarded as one commonly used statistical hypothesis testing method in the field of forecast comparison; see Diebold and Mariano (1995) and Harvey et al. (1997)

$$\{e(t), t = 2003, \ldots, 2010\}$$

Table 3 reports the number of wins for one model from the group of CMT models over another from the group of benchmark models. Each cell of the table contains two integers recording the comparison results of the model in the row versus the one in the column. The first integer is the number of wins of the model in the row, and the second integer is the number of wins of the model in the column. For example, "(23, 1)" in the first column and the fourth row in the panel of Female Population means that the CMT-ACF·RankAvg model wins for 23 times over the Lee-Carter model and the Lee-Carter model wins the CMT-ACF·RankAvg model only once among all of the 24 comparisons for female populations. As Table 3 clearly indicates, the CMT-based models perform significantly better than benchmark models in terms of the number of wins. The superiority of CMT-ACF·SimAvg and CMT-ACF·RankAvg is evident when these models are compared with all of the benchmark models considered in our study since these two models are capable to win frequently among all 24 comparisons in all data scopes (female or male population). Although less convincing results are obtained when the CMT-ACF·GeoAvg and CMT-ACF·KmeansAvg are applied, they are sufficiently competitive to the benchmark methods. Furthermore, the outperformance of CMT-ACF-SimAvg and CMT-ACF·RankAvg over CMT-ACF·GeoAvg and CMT-ACF·KmeansAvg confirms again on our previous comments regarding their relative performance in Remark 1.

Table 3: Number of wins for comparisons between the CMT-ACF model and the benchmark ACF model based on a pairs of one-sided DM tests: In each cell, the first integer indicates the number of wins of the model in the row over the model in the column out of 30 comparisons and the second integer is the number of wins of the model in the column over the one in the row.

	Lee-Carter	ACF-AIO	ACF-GeoInfo	ACF-kmeans
Female Population				
CMT-ACF·SimAvg	(23, 1)	(22, 1)	(17, 0)	(16, 3)
CMT-ACF·GeoAvg	(13, 0)	(15, 1)	(13, 3)	(12, 4)
CMT-ACF·KmeansAvg	(15, 1)	(15, 2)	(15, 3)	(11, 3)
CMT-ACF·RankAvg	(21, 2)	(19, 2)	(16, 1)	(14, 2)
Male Population				
CMT-ACF·SimAvg	(24, 0)	(20, 0)	(16, 4)	(15, 4)
CMT-ACF·GeoAvg	(16, 3)	(13, 3)	(10, 9)	(11, 7)
CMT-ACF·KmeansAvg	(19, 0)	(17, 0)	(10, 5)	(10, 9)
CMT-ACF·RankAvg	(24, 0)	(21, 0)	(14, 3)	(17, 3)

Δt

Δt

Δt 4 . As discussed in Section 2

Δt

Δt

Δt 4

Δt

Δt

Δt

Target	Female	Male	Target	Female	Male
Australia	0.13	0	New Zealand	-1.70	-0.87
Austria	-1.22	-1.91	Norway	-2.00	-0.83
Belgium	0.26	0.13	Poland	0.17	0.70
Canada	3.09	0.83	Portugal	-0.09	-0.39
Czech Republic	-1.35	-1.96	Slovakia	0.57	-0.61
Denmark	-0.83	0	Spain	0.48	-0.70
Finland	-0.39	-0.78	Sweden	1.65	0.39
France	0.74	0.74	Switzerland	-1.30	-1.61
Hungary	-1.52	-1.74	Taiwan	-1.52	-1.65
Italy	0.04	1.35	England & Wales	1.91	2.91
Japan	2.83	4.61	Scotland	-1.70	-1.96
Netherlands	0.17	1.43	U.S.A.	1.57	1.91

Δt

Δt

4 and 5

Δt

Δt

Δt

Δt

Δt

(1)-(2) Δt

$RS(i,j)$

$RS(i,j)$

where the mean operator is applied for the corresponding sequence over the time period involved in the CMT model, and

$$\Delta k_i(t) = k_i(t+1) - k_i(t), \;\; \Delta k_j(t) = k_j(t+1) - k_j(t), \;\; \Delta K(t) = K(t+1) - K(t).$$

$RS(i,j)$

$RS(i,j)$

3 $RS(i,j)$

Δt

3 are symmetrical since we have $RS(i,j) = RS(j,i)$ by the definition of the *RS* measure.

Δt

3, $\gamma \in \{0.25, 0.5, 1, \infty\}$

$RS(i,j)$

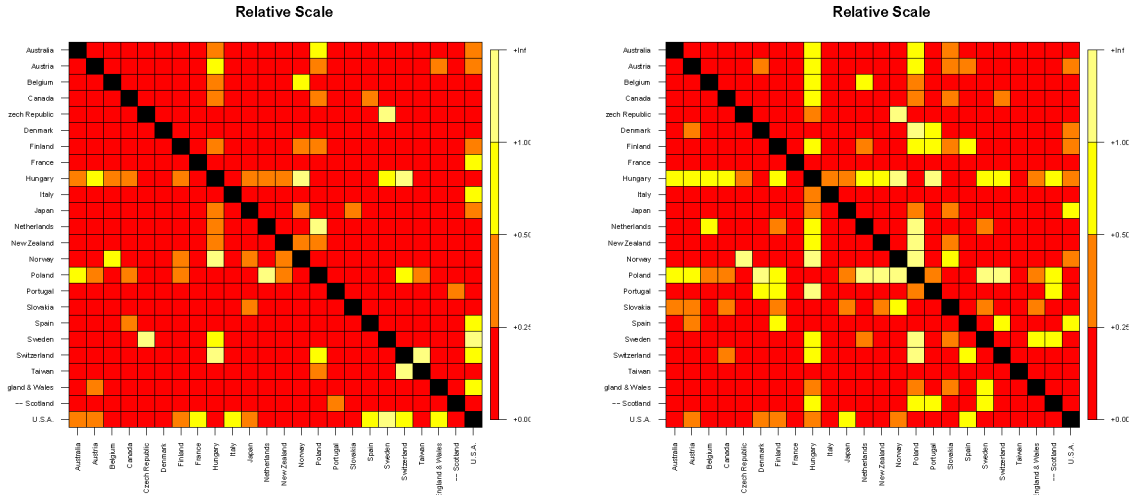
$RS(i,j)$

Δt

Δt 4 and 5,

Δt

Figure 3: Values of Relative Scale based on CMT-ACF models. Left panel: female data. Right panel: male data.



4.3 Empirical Study with the CMT-CBD Model

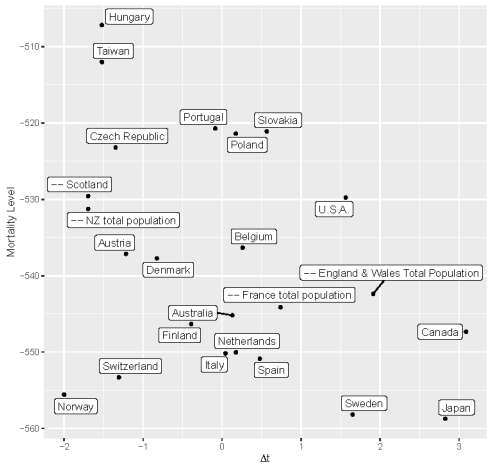
In this empirical study, we investigate the prediction performance of the model averaging method for seniors with an age ranging from 55 to 90. The CBD-type models are known as an improvement of the Lee-Carter/APC-type models for modelling and predicting the mortality of a senior group because of their relatively simple log-linear structure of the mortality curve and parsimonious age effects, see Cairns et al. (2009), Cairns et al. (2011). We use the CMT-CBD models as building blocks in the model averaging method considering grouping strategies proposed in Section 3, and compare their prediction performance with the classic CBD model (Cairns et al., 2006), also known as model M5 in Cairns et al. (2009).

4.3.1 Prediction Models

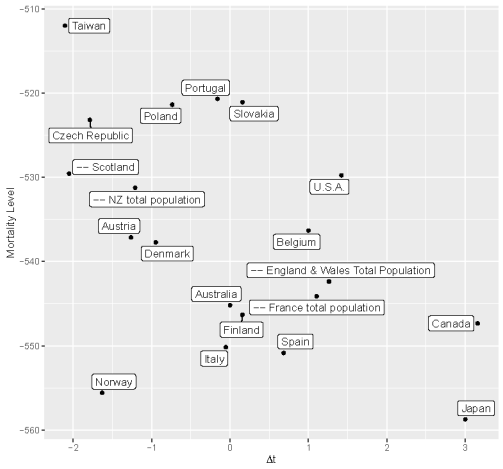
We consider the same four model averaging strategies as described in Section 3, which are now used with CMT-CBD models and labelled as follows:

- **CMT-CBD·SimAvg:** The final prediction is an average of predicted values from all the 23 bivariate-population CMT-CBD models.
- **CMT-CBD·GeoAvg:** The final prediction is an average of predicted values from the CMT-CBD models built with a reference from the same geographic group as listed in Table 1.
- **CMT-CBD·KmeansAvg:** A k-means cluster analysis with 8 clusters and 1,000 independent initialization is done as that for the **CMT-ACF·KmeansAvg** method in Section 4.2.1. The final prediction is an average of the predicted values from the CMT-CBD models built with a reference from the same cluster.
- **CMT-CBD·RankAvg:** The final prediction is an average of predicted values from some "top-ranked" CMT-CBD models which are chosen in the same manner as those selected for the **CMT-ACF·RankAvg** method described in Section 4.2.1.

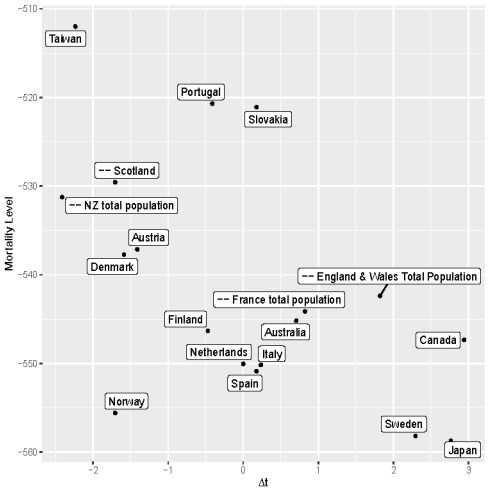
Δt



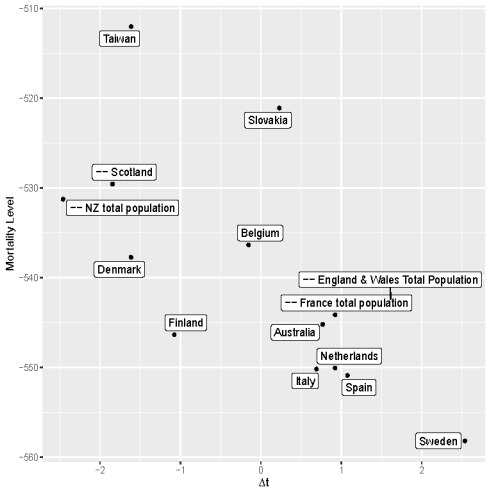
(a) $\gamma = \infty$, Corr = - 0.397



(b) $\gamma = 1$, Corr = - 0.506

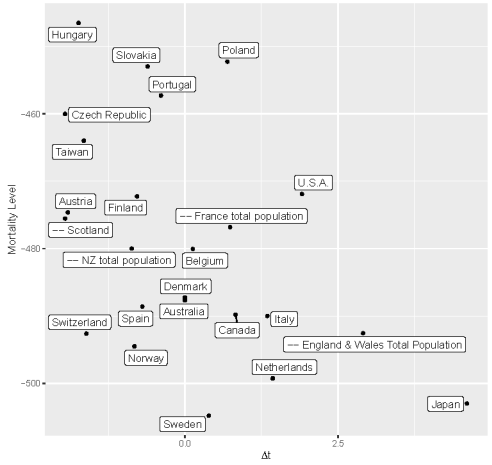


(c) $\gamma = 0.5$, Corr = - 0.560

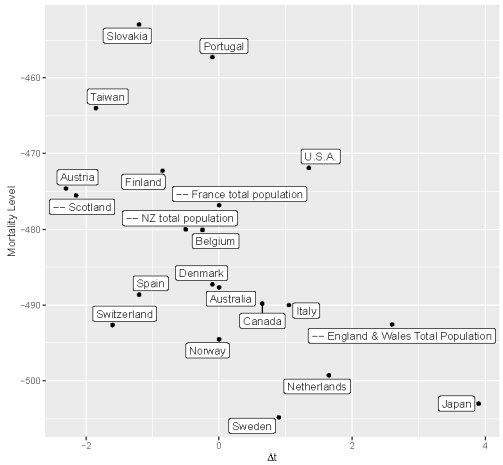


(d) $\gamma = 0.25$, Corr = - 0.671

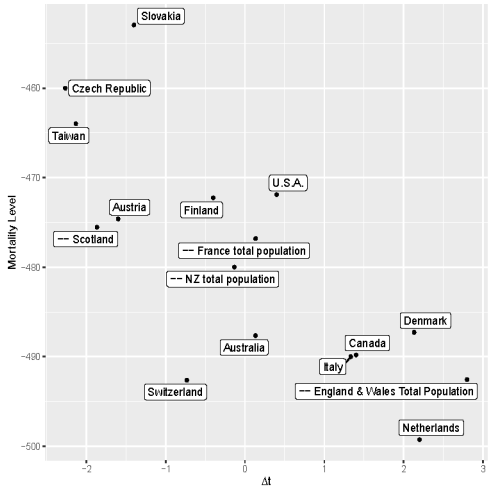
Figure 5: Male Data: Relationship between the mean value of Δt and the relative mortality level based on CMT-ACF models with different threshold values of *γ*



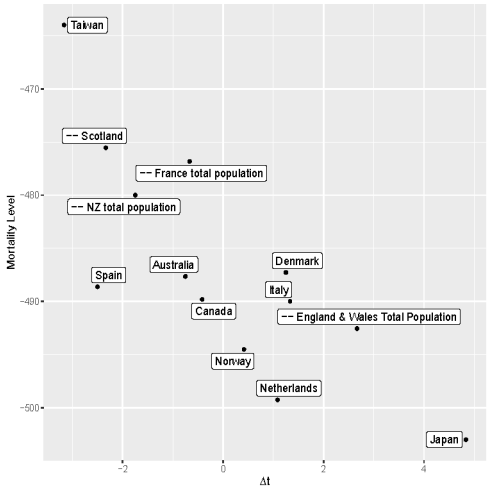
(a) *γ* = co, Corr = - 0.478



(b) *γ* = 1, Corr = - 0.555



(c) *γ* = 0.5, Corr = - 0.772



(d) *γ* = 0.25, Corr = - 0.793

Remark 2. *Regarding the above four averaging strategies applied to the CMT-ACF models in the empirical study of the preceding section, we have observed that KmeansAvg and GeoAvg do not perform as well as SimAvg and RankAvg, and we have explained a reason for this in Remark 1. We still expect RankAvg to work well for the CMT-CBD models in the current empirical analysis since this particular averaging strategy is fully data-driven and adaptive. However, the reason for the outperformance of SimAvg strategy for the CMT- ACF models stated in Remark 1 no longer firmly stands for the current empirical study with the CMT-CBD models. As we have stated in Remark 1, the primary reason for the SimAvg to outperform the other strategies in our previous empirical analysis with the CMT- ACF models is the preanalysis we had conducted in Section 4.1 that excluded several former Soviet Union populations for their disparate mortality develop patterns. The dissimilarity in the preanalysis is based on all of the ages 0-100. In contrast, in the current empirical study, we only focus on senior ages 55-90 and the k-means clustering in the preanalysis may not remain the same if we focus on the senior ages 55-90 only. In short, there is no transparent reason that a joint CBD model of a target population with each of the other 23 populations would bring more information than noise to enhance the prediction of mortality for the target population, which creates a risk of using the simple averaging over all of the involved populations.*

For comparison, we consider a classic CBD model.

- **CBD:** The single-population CBD model is fitted to the training mortality data of each individual population independently.

4.3.2 Prediction Performance

Evaluation Based on Test SSEs

As in the empirical study for the CMT-ACF model, we also use test SSEs as a measure of prediction accuracy. To compare the prediction performance of the prediction models listed in the preceding subsection, we calculate the test SSE as the summation of squares between the difference between the logit function of the probability $q(x,t)$ and its predicted counterpart over the age range 55-90, and the test period, years 2003-2010; i.e.,

$$e(t) = \sum_{x=55}^{90} \left[\operatorname{logit} q_1(x,t) - \operatorname{logit} \widehat{q_1}(x,t)\right]^2, \quad t = 2003, \ldots, 2010, \tag{13}$$

where $q_1(x,t)$ is the probability that an age-*x* individual from the target population dies between year *t* and *t* +1 given the individual is alive at time t, and $\widehat{q_1}(x,t)$ is its prediction.

Table 5 summarizes the prediction performance of the various models described in Section 4.3.1. Table 5 shows that the improvement of the CMT-CBD models over the classic CBD model is not as significant as the CMT-ACF models over benchmark ACF models as reported in Table 2. The CMT-CBD·RankAvg still performs better than the CBD model. Other averaging methods demonstrate a less satisfactory performance. This observation is in accordance with our previous comments in Remark 2.

Table 5: Summary statistics of test SSEs of 24 female populations and 24 male populations comparing the prediction performance of CMT-CBD models versus the CBD model

	1st Quartile	Median	Mean	3rd Quartile
Female Population				
CBD	3.61	5.87	7.37	11.67
CMT-CBD·SimAvg	5.13	7.96	10.28	13.32
CMT-CBD·GeoAvg	4.60	6.50	8.76	11.95
CMT-CBD·KmeansAvg	3.97	5.76	7.61	11.70
CMT-CBD·RankAvg	3.18	5.76	6.94	10.32
Male Population				
CBD	2.32	2.88	3.08	3.63
CMT-CBD·SimAvg	2.08	3.04	5.95	4.38
CMT-CBD·GeoAvg	2.12	2.60	2.84	3.69
CMT-CBD·KmeansAvg	2.22	2.54	2.84	3.56
CMT-CBD·RankAvg	2.04	2.67	2.85	3.38

Evaluation Based on One-sided Diebold-Mariano (DM) Tests

To conduct a formal comparison in prediction performance between the CMT-CBD models and the CBD model, we also conduct the DM tests in the same manner as we have done for the empirical study with the CMT-AFT models in Section 4.2.2. The results in Table 6 suggest that CMT-CBD-RankAvg performs significantly better than the benchmark CBD model for both the gender-specific populations. CMT-CBD·GeoAvg significantly outperforms the CBD model for male populations and CMT-CBD·KmeansAvg significantly outperforms the CBD model for female populations. Among the CMT-CBD models, the simple averaging strategy only performs slightly better than the CBD model for both female and male population. Nevertheless, the well-designed “Rank and Average” method produces superior performance.

Table 6: Number of wins for comparisons between a CMT-CBD model and the CBD model based on a pairs of one-sided DM tests: In each cell, the first integer indicates the number of wins of the model in the row over the model in the column out of 24 comparisons and the second integer is the number of wins of the model in the column over the one in the row.

	CBD	
	Female	Male
CMT-CBD·SimAvg	(10, 7)	(12, 9)
CMT-CBD·GeoAvg	(7, 8)	(12, 6)
CMT-CBD·KmeansAvg	(11, 5)	(6, 5)
CMT-CBD·RankAvg	(11, 3)	(10, 5)

4.3.3 Interpretation of Δt Values in the CMT-CBD Models

For each target population, we have 23 Δt values resulted from the 23 bivariate CMT- CBD models that we have calibrated between the target and a reference from the rest 23 populations. We calculate the mean of the 23 Δt values for each target population and report them in Table 7.

We also provide an analysis on the relationship between the calibrated Δt values in the bivariate CMT-CBD models and the mortality development level as we have previously conducted in Section 4.2.3 for the CMT-ACF models. We recall that the degree to which the Δt value reflects the relative mortality level of involved populations in a CMT model depends on how much the common trend process contributes to the mortality development compared with the population-specific effects. In equation 12, we defined the RS measure to capture the degree of dominance by the common trend process over the population-specific effects. We apply the same RS measure in the current empirical study for the CMT-CBD models. The resulting RS values for all of the involved pairs of populations are demonstrated in Figure 6. The figure indicates that all the RS values are quite small. Thus, the common trend process dominates for the development of mortality in all the involved bivariate-population CMT-CBD models, and we should expect a high correlation between the Δt value and the mortality level of a population.

We provide the scatter plots of the mean of age-aggregated logarithmic mortality rates versus the mean of Δt values for all the 24 target populations in Figure 7. Figure 7 exhibits an obvious negative relationship between the mean value of Δt and the mortality level. Populations with lower mortality tend to have a positive mean value of Δt, and those with higher mortality tend to have a negative mean value of Δt.

Table 7: Mean of 23 Δt values for each target population based on the CMT-CBD model

Target	Female	Male	Target	Female	Male
Australia	4.48	4.43	New Zealand	-0.13	2.22
Austria	1.17	-0.83	Norway	1.65	2.61
Belgium	1.65	-1.13	Poland	-7.17	-7.78
Canada	3.91	5.17	Portugal	-0.04	-1.87
Czech Republic	-7.30	-8.39	Slovakia	-6.26	-8.57
Denmark	-6.04	-2.74	Spain	8.26	5.43
Finland	1.48	-2.70	Sweden	0.91	6.43
France	7.65	4.13	Switzerland	7.04	6.57
Hungary	-7.91	-9.48	Taiwan	-4.96	-0.91
Italy	6.00	3.87	England & Wales	-2.39	-0.39
Japan	9.48	8.65	Scotland	-6.04	-5.65
Netherlands	-3.17	-0.30	U.S.A.	-2.26	1.22

The Pearson correlation coefficients between the two variables for the female population and male populations are as high as -0.8958 and -0.9706 respectively, suggesting a very strong negative linear correlation. These observations confirm the intended interpretation of the parameter Δt in our bivariate CMT-CBD models: a positive and large value implies the advance of the target population in mortality development relative to the reference population, and a negative value implies a delay of the target population in mortality development.

Moreover, since the CMT is the major driving force of mortality development for both populations in every bivariate-population system involved in the current empirical study as validated by the observed small RS values, the value of the Δt also quantifies the number of years of difference in mortality development between the target population and the reference population. For instance, the calibrated Δt value between the US male population (as the target) and the Canadian male population (as the reference) is -6, meaning that the Canadian male population is 6 years earlier than the target US male population in terms of mortality development stage. We show the age-aggregated logarithmic mortality sequences of the two populations in Figure 8 for a graphical illustration of the relationship between the two populations.

5 Concluding Remarks

We propose a general and flexible mortality forecasting framework involving building bivariate CMT models based on existing bivariate-population mortality models and averaging these bivariate models through various strategies. An important advantage of our proposed framework is that, by allowing more flexible structure on the timeline, we are able to borrow information not only from across populations but also from across time.

Figure 6: Values of Relative Scale based on CMT-CBD models. Left panel: female data. Right panel: male data.

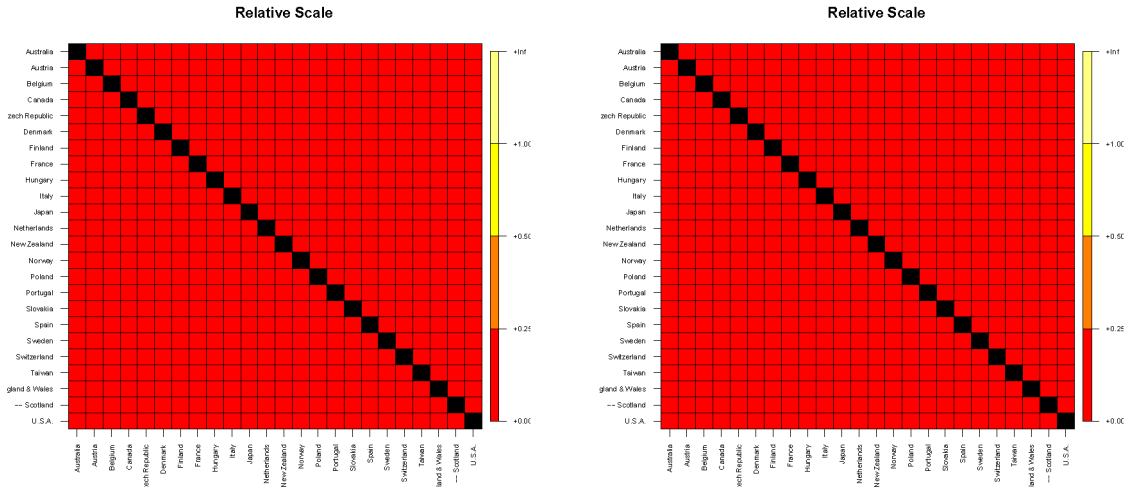
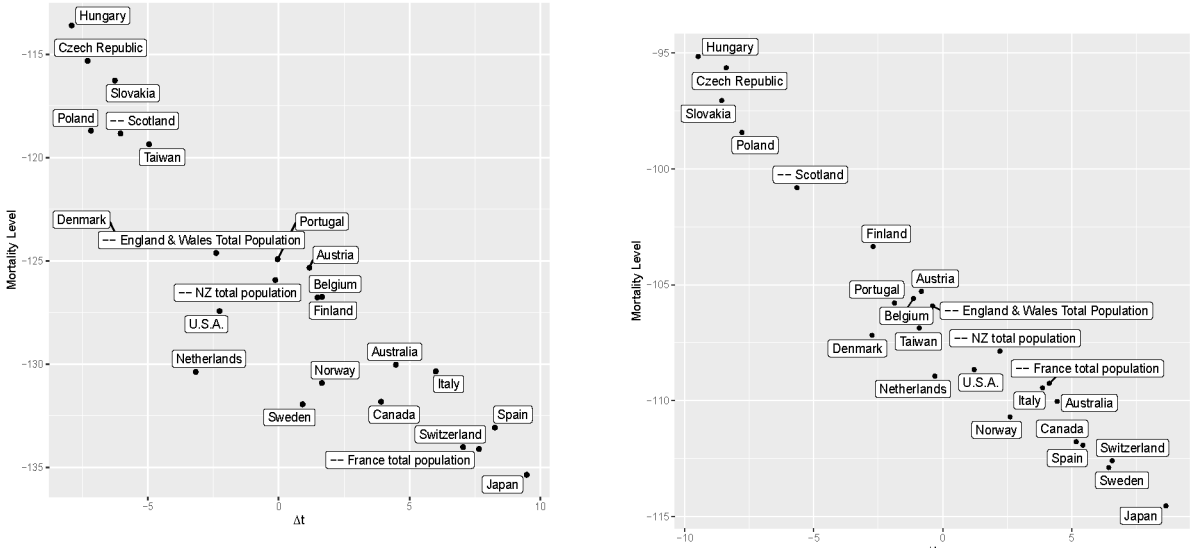


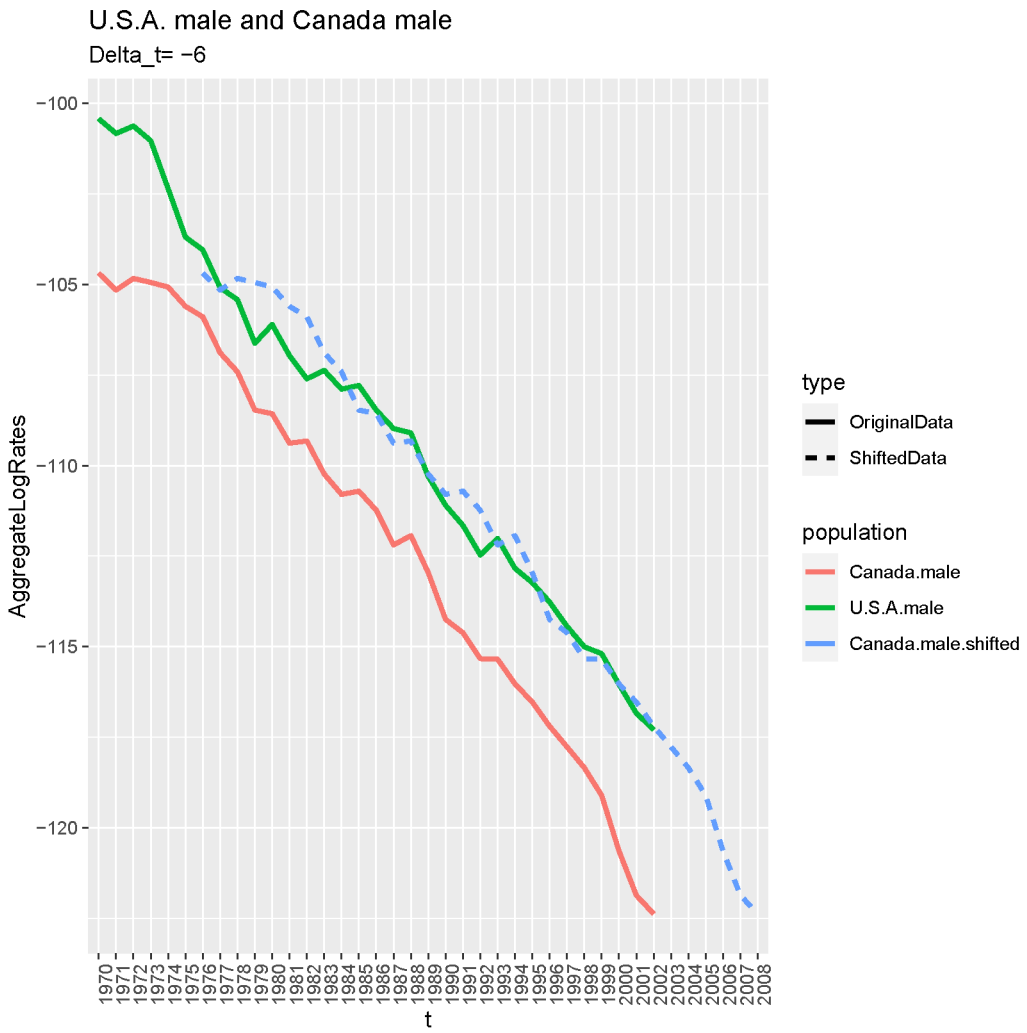
Figure 7: Relationship between the mean value of Δt and the relative mortality level based on CMT-CBD models



(a) Female: Corr = - 0.896

(b) Male: Corr = - 0.971

Figure 8: Age-aggregated logarithmic mortality sequences for the US male population and the Canadian male population



It provides a practical way to enhance both the accuracy and efficiency of mortality forecast. This advantage is further exploited by the introduction of the model averaging idea, which helps efficiently reduce prediction errors and provides predictions with accuracy and robustness in a computational-friendly way. We conclude from our empirical studies that our proposed CMT models using the "Rank and Average" model averaging strategy are capable of achieving a superior prediction performance, regardless of whether the main focus of analysis is paid to the prediction for the whole age groups or just a senior subgroup, in the total population or the gender-specific populations.

Another advantage of our CMT model framework is that we no longer need to be concerned with whether the populations included in the model are in a similar socio-economic condition when we attempt to borrow information from other populations for the purpose of mortality forecast. Our framework is capable of recognizing useful information and screen out irrelevant

noise and thus allowing for a greater degree of flexibility in multi-population mortality modelling than the extant models. In addition, the parameter Δt, designed to capture the "stage of evolution" for each population in the CMT model, is economically sensible as it effectively characterizes the disparity of mortality levels that have been caused by historical events and the socio-economic development, across populations in most cases we have considered in this study. It is consistent with the intuition that a positive and large value of Δt implies a relatively leading position for the target population in mortality development with lower mortality level while a delayed position in mortality development for the target population is revealed by a negative value of Δt. In addition, when the CMT is the major driving force of the mortality development for a bivariate-population systems, the value of the Δt can also quantifies number of years of difference in mortality level of the target population compared to the reference population.

There are many relevant issues spawned by this framework that are worth further exploration. Since the model averaging idea brings in both accuracy and robustness with an affordable increase in computational demand, we would like to extend the model averaging idea to more challenging scenarios where the prediction of mortality rates of a particular age group (i.e., infant, teenage, or senior) becomes the goal of a study with richer candidates in the reference pool (i.e., allowing to borrow information from specific combination in region, gender, and age). To exploit the full potential of the model averaging idea, we will consider more flexible and powerful strategies for the calculation of the final prediction: a weighted average based scheme will be considered as an alternative to the naive simple average strategy. CMT·GeoAvg, CMT·KmeansAvg and CMT·RankAvg select a subset of the developed bivariate CMT models, while other fully data-driven selection methods can also be considered. For instance, Diao et al. (2020) used a so-called "Deletion/Substitution/Addition" algorithm to automatically select a group of populations under the ACF model framework. Developing a selection procedure of the CMT models in the same spirit is another direction warranting further research.

Acknowledgment

The authors acknowledge funding support from the Canadian Institute of Actuaries (No.: CS000215).

References

Elizabeth Brainerd and David M Cutler. Autopsy on an empire: understanding mortality in Russia and the former Soviet Union. *Journal of Economic Perspectives,* 19(1):107-130, 2005.

Andrew JG Cairns, David Blake, and Kevin Dowd. A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration. *Journal of Risk and Insurance*, 73(4):687-718, 2006.

Andrew JG Cairns, David Blake, Kevin Dowd, Guy D Coughlan, David Epstein, Alen Ong, and Igor Balevich. A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal*, 13(1):1-35, 2009.

Andrew JG Cairns, David Blake, Kevin Dowd, Guy D Coughlan, David Epstein, and Marwa Khalaf-Allah. Mortality density forecasts: An analysis of six stochastic mortality models. Insurance: *Mathematics and Economics*, 48(3):355-367, 2011.

Michael AA Cox and Trevor F Cox. Multidimensional scaling. In Handbook of Data Visualization, pages 315-347. Springer, 2008.

Liqun Diao, Yechao Meng, and Chengguo Weng. A DSA algorithm for mortality forecasting. Available at SSRN 3537368, 2020.

Francis X Diebold and Roberto S Mariano. Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3), 1995.

Kevin Dowd, Andrew JG Cairns, David Blake, Guy D Coughlan, and Marwa Khalaf-Allah. A gravity model of mortality rates for two related populations. *North American Actuarial Journal*, 15(2):334-356, 2011.

David Harvey, Stephen Leybourne, and Paul Newbold. Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2):281-291, 1997.

Petros Hatzopoulos and Steven Haberman. Common mortality modeling and coherent forecasts. An empirical analysis of worldwide mortality data. *Insurance: Mathematics and Economics*, 52(2):320-337, 2013.

Rob J Hyndman, Heather Booth, and Farah Yasmeen. Coherent mortality forecasting: the product-ratio method with functional time series models. *Demography*, 50(1):261-283, 2013.

Fanny Janssen, Leo JG van Wissen, and Anton E Kunst. Including the smoking epidemic in internationally coherent mortality projections. *Demography*, 50(4):1341-1362, 2013.

Torsten Kleinow. A common age effect model for the mortality of multiple populations. *Insurance: Mathematics and Economics*, 63:147-152, 2015.

Ronald D Lee and Lawrence R Carter. Modeling and forecasting US mortality. *Journal of the American Statistical Association*, 87(419):659-671, 1992.

Jackie Li. A Poisson common factor model for projecting mortality and life expectancy jointly for females and males. *Population Studies*, 67(1):111-126, 2013.

Johnny Siu-Hang Li and Mary R Hardy. Measuring basis risk in longevity hedges. *North American*

Actuarial Journal, 15(2):177-200, 2011.

Johnny Siu-Hang Li, Rui Zhou, and Mary Hardy. A step-by-step guide to building two- population stochastic mortality models. *Insurance: Mathematics and Economics*, 63:121-134, 2015.

Johnny Siu-Hang Li, Wai-Sum Chan, and Rui Zhou. Semicoherent multipopulation mortality modeling: the impact on longevity risk securitization. *Journal of Risk and Insurance*, 84(3):1025-1065, 2017.

Nan Li and Ronald Lee. Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography*, 42(3):575-594, 2005.

Maria Russolillo, Giuseppe Giordano, and Steven Haberman. Extending the Lee-Carter model: a three-way decomposition. *Scandinavian Actuarial Journal*, 2011(2):96-117, 2011.

Han Lin Shang and Rob J Hyndman. Grouped functional time series forecasting: An application to age-specific mortality rates. *Journal of Computational and Graphical Statistics*, 26(2):330-343, 2017.

Chou-Wen Wang and Sharon S Yang. Pricing survivor derivatives with cohort mortality dependence under the Lee-Carter framework. *Journal of Risk and Insurance*, 80(4):1027-1056, 2013.

Rui Zhou, Yujiao Wang, Kai Kaufhold, Johnny Siu-Hang Li, and Ken Seng Tan. Modeling period effects in multi-population mortality models: Applications to Solvency II. *North American Actuarial Journal*, 18(1):150-167, 2014.

Canadian Institute of Actuaries
360 Albert Street, Suite 1740
Ottawa, ON K1R 7X7
613-236-8196
head.office@cia-ica.ca

cia-ica.ca
seeingbeyondrisk.ca

The Canadian Institute of Actuaries (CIA) is the qualifying and governing body of the actuarial profession in Canada. We develop and uphold rigorous standards, share our risk management expertise, and advance actuarial science for the financial well-being of society. Our more than 6,000 members apply their knowledge of math, statistics, data analytics, and business in providing services and advice of the highest quality to help ensure the financial security of all Canadians.