# Simultaneous Prediction Intervals: An Application to Forecasting US and Canadian Mortality

Johnny Siu-Hang Li* and Wai-Sum Chan

**Abstract**

In recent years, there has been a new wave of work that is focused on the forecasting of uncertainty in mortality projections. Such work aims to forecast a range of possible outcomes along with associated probabilities, instead of a single prediction that will almost surely be wrong. Conventionally, isolated (point-wise) prediction intervals are used to quantify the uncertainty in future mortality rates and other demographic quantities such as life expectancy. A pointwise interval reflects uncertainty in a variable at a single time point, but it does not account for any dynamic property of the time-series. As a result, in situations when the path or trajectory of future mortality rates is important, a band of pointwise intervals might lead to invalid inference. To improve the communication of uncertainty, a simultaneous prediction band may be used. The primary objective of this paper is to demonstrate how simultaneous prediction bands can be created for prevalent stochastic models. The illustrations in this paper are based on mortality data from the general populations of US and Canada.

**Keywords:** Bayesian methods; Longevity risk; The Cairns-Blake-Dowd model.

## 1 Introduction

When it comes to product pricing and reserving, actuaries often need life tables that include a forecast of future longevity improvement. However, the production of such

*Corresponding author. Address: Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1. Email: shli@uwaterloo.ca

tables is not straightforward. Tuljapurkar (2005) describes the challenge of forecasting mortality as "a bumpy road to Shangri-La," because the demographic future of any human population is a result of complex and only partially understood mechanisms, and is highly uncertain. Indeed, recent mortality data have unfolded significant deviations between the actual experience and the assumptions that actuaries made in the past (see, e.g., Continuous Mortality Investigation Bureau, 1999, 2002).

In recent years, actuaries have been understandably concerned about error in the mortality assumptions they make. Part of their response is a new wave of work that is focused on the forecasting of uncertainty in longevity improvement, rather than producing a single mortality projection that will almost surely be wrong. This goal is accomplished by using stochastic mortality models, for example, the Lee-Carter model (Lee and Carter, 1992) and its variants (e.g., Renshaw and Haberman, 2003, 2006; Delwarde et al., 2007; Li et al., 2009), which are fitted to historic data. The resulting models have uncertainty embedded within them, as reflected in historical change. The use of such a stochastic approach is now highly regarded by leading actuarial organizations (see, e.g., Continuous Mortality Investigation Bureau, 2004).

Given a fitted stochastic mortality model, we can express the uncertainty associated with future death rates in terms of confidence or prediction intervals. The interval estimates are crucially important to life insurers and annuity providers, since they provide guidance on how to determine appropriate margins for adverse deviations. They have also drawn considerable attention among academics. For example, in a study of Canadian insured lives mortality, Li et al. (2007) derived, on the basis of the original Lee-Carter model, approximate formulas that allow actuaries to calculate confidence intervals for future age-specific central death rates with spreadsheet software.

Another significant development in this research area is the concept of mortality fan charts, proposed by Blake et al. (2008) and Dowd et al. (2010a). These charts are highly parallel to the well known inflation fan charts, which have been produced periodically by the Bank of England since 1996. A mortality fan chart depicts prediction intervals at different levels of confidence simultaneously. In particular, it shows the central 10% prediction interval with the heaviest shading, surrounded by the 20%, 30%, ..., 90% prediction intervals with progressively lighter shading. We can therefore interpret the degree of shading as the likelihood of the outcome – the darker the shading, the more likely the outcome. As Blake et al. (2008) mention, mortality fan charts have a wide range of applications, including pricing, hedging and setting capital requirements.

In previous research, including the aforementioned studies, the prediction intervals derived are mostly isolated pointwise prediction intervals. By pointwise we mean that the interval reflects uncertainty in a quantity at a single point of time, but it does not account for any dynamic property of the time-series. However, in actuarial practice, rather than a single death rate at a particular time point, what practitioners need is the entire trajectory of mortality rates for the birth cohort in question. Specifically, of their interest would be questions like "Within what bounds would the trajectory of cohort mortality rates likely to remain with a certain degree of confidence?" Such questions are equally important to investors in mortality-linked securities, particularly those that are path dependent.

From a statistical viewpoint, a band of pointwise intervals might lead to invalid inference concerning the time trajectory, and in particular to misinterpretation of dynamic uncertainty. Assuming that the model is correct, a $100\gamma\%$ pointwise confidence interval should cover $100\gamma\%$ of the random quantity at a certain time point. However, unless all trajectories develop very orderly, a band of $100\gamma\%$ pointwise confidence intervals will cover less than $100\gamma\%$ of the trajectories of the random quantity. In other words, we may understate the actual uncertainty asscoiated with a random mortality trajectory if a band of isolated pointwise confidence intervals is used.

The communication of uncertainty can be further improved by considering a band with a prescribed probability of covering the whole time trajectory. Such a band, which is referred to as a time-simultaneous prediction band, is the primary focus of this paper. Interval forecasting from a time-simultaneous perspective has been applied to economic variables such as GDP (Parigi and Schlitzer, 1995) and unemployment rate (Kolsrud, 2007). It has also been used in biostatistics to construct prediction bands for survival functions from models with covariates (Nair, 1984; Scheike and Zhang, 2003).

In this study we extend the numerical methods proposed by Kolsrud (2007) to construct time-simultaneous prediction bands for mortality forecasts generated from prevalent stochastic mortality models. These methods start with a learning sample of time trajectories generated from a dynamic stochastic model. Suppose that the trajectories are $S$ steps in length. Then, a trajectory can be seen as a point in an $S$-dimensional space, and the learning sample becomes a 'cloud' of points. A time-simultaneous band is developed by constructing geometrically a high-dimensional box that contains a prescribed fraction of points in the cloud.

The rest of this paper is organized as follows: Section 2 describes the data we use for the purpose of illustration; Section 3 provides a formal mathematical definition of a time-simultaneous prediction band; Section 4 presents the numerical methods for constructing time-simultaneous prediction bands; illustrations are also provided in this section; finally, Section 5 concludes the paper.

## 2 Data

We use historic mortality data for US and Canadian (unisex) populations to illustrate the methods we propose. The required data, death counts and exposures-to-risk, are obtained from the Human Mortality Database (2010). We consider data from age 60 to 99 and from year 1951 to 2004. Note that the methods we propose do not require a specific choice of a sample age range and a sample period.

## 3 Definitions

Let $y_t$ be a general single time-series variable, which can be either stationary or non-stationary. Suppose that the forecast originates at time $T$ and that the forecast horizon (the period for which the forecast is prepared) is $S$ years. We define a pointwise prediction interval for $y_{T+s}$, $s = 1, 2, \ldots, S$, as follows:

**Definition 1.** $PI_s = [l_s, h_s]$ *is a pointwise prediction interval for* $y_{T+s}$ *with coverage probability* $0 < 1 - \alpha \leq 1$ *if*

$$\Pr(l_s \leq y_{T+s} \leq h_s) = 1 - \alpha.$$

Note that a pointwise prediction interval treats the time-series random variable at different time points in isolation.

The values of $y_{T+s}$ for $s = 1, \ldots, S$ constitute a trajectory $\mathbf{y} = (y_{T+1}, \ldots, y_{T+S})$. Unless all trajectories develop very orderly, the probability that a trajectory lies completely inside all $S$ pointwise prediction intervals $PI_s$, $s = 1, \ldots, S$, would be less than $1 - \alpha$. Therefore, in situations where the entire trajectory of the time-series variable is important, a band of pointwise prediction intervals would not be sufficient in communicating the underlying uncertainty. This motivates us to consider a time-simultaneous prediction band, which is defined as follows:

**Definition 2.** $\mathbf{PB} = [\mathbf{l}, \mathbf{h}] = ([l_s, h_s])_{s=1}^{S}$ *is a time-simultaneous prediction band with coverage probability* $0 < 1 - \alpha \leq 1$ *for a random trajectory* $\mathbf{y}$ *if*

$$\Pr(\mathbf{y} \in \mathbf{PB}) = \Pr\left(\bigcap_{s=1}^{S}(l_s \leq y_{T+s} \leq h_s)\right) = 1 - \alpha.$$

The definitions of a pointwise prediction interval and a time-simultaneous prediction band we provide above will be used throughout the rest of this article.

# 4 Mortality Models

In this section we provide a brief review of two mortality models which we use to illustrate the concept of time-simultaneous prediction bands.

## 4.1 The Cairns-Blake-Dowd Model

Cairns et al. (2006) propose a two-factor stochastic mortality model, which is then called the Cairns-Blake-Dowd model. Mathematically, the model can be expressed as

$$\ln\left(\frac{q_{x,t}}{1 - q_{x,t}}\right) = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}), \tag{1}$$

where $q_{x,t}$ is the realized single-year death probability at age $x$ and time $t$, $\bar{x}$ is the average age over the age range we consider, and $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ are period indexes. In particular, we may consider $\kappa_t^{(1)}$ as an indicator of the overall mortality level at time $t$ and $\kappa_t^{(2)}$ as an indicator of the steepness of the mortality curve (in logit scale) at time $t$. This model has no identifiability problems, and therefore parameter constraints are not required. The model can be estimated by the method of maximum likelihood (see the Appendix). The maximum likelihood estimates of $\kappa_t^{(1)}$ and $\kappa_t^{(1)}$, $t = 1950, \ldots, 2004$, are shown graphically in Figures 1 and 2.

Note that family of Cairns-Blake-Dowd models is based on $q_{x,t}$ rather than $m_{x,t}$. To obtain central death rates, we can use the following relation:

$$m_{x,t} = -\ln(1 - q_{x,t}), \tag{2}$$

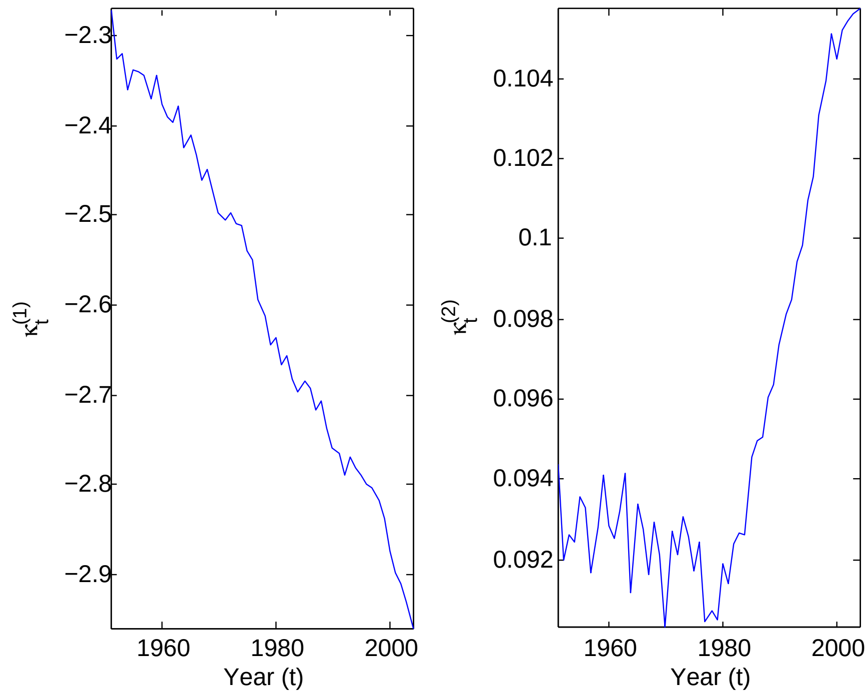



Figure 1: Maximum likelihood estimates of parameters in the Cairns-Blake-Dowd model, Canadian population.

which holds if we assume that the force of mortality is constant over each year of integer age and over each calendar year.

After fitting equation (1) to historic death probabilities, the period indexes $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ are modeled by a bivariate random walk with drift, that is,

$$\kappa_{t+1} = \kappa_t + \mu + CZ(t+1), \tag{3}$$

where $\kappa_t = (\kappa_t^{(1)}, \kappa_t^{(2)})'$, $\mu = (\mu_1, \mu_2)'$ is a constant $2 \times 1$ vector, $C$ is a constant $2 \times 2$ upper triangular matrix, and $Z(t)$ is a 2-dimensional standard normal random vector.

Consider the cohort of individuals who are aged $x$ at the forecast origin $T$. The best estimate of the death probability for this birth cohort at time $T+s$ is given by the following equation:

$$\ln\left(\frac{\hat{q}_{x+s,T+s}}{1-\hat{q}_{x+s,T+s}}\right) = \kappa_T^{(1)}(s) + \kappa_T^{(2)}(s)(x+s-\bar{x}),$$

where $\kappa_T^{(1)}(s) = \kappa_T^{(1)} + s\mu_1$ and $\kappa_T^{(2)}(s) = \kappa_T^{(2)} + s\mu_2$ are the minimum sqaure error (MMSE) forecasts of $\kappa_{T+s}^{(1)}$ and $\kappa_{T+s}^{(2)}$, respectively. In practice when $\mu_1$ and $\mu_2$ are not known, we replace them with their estimates.

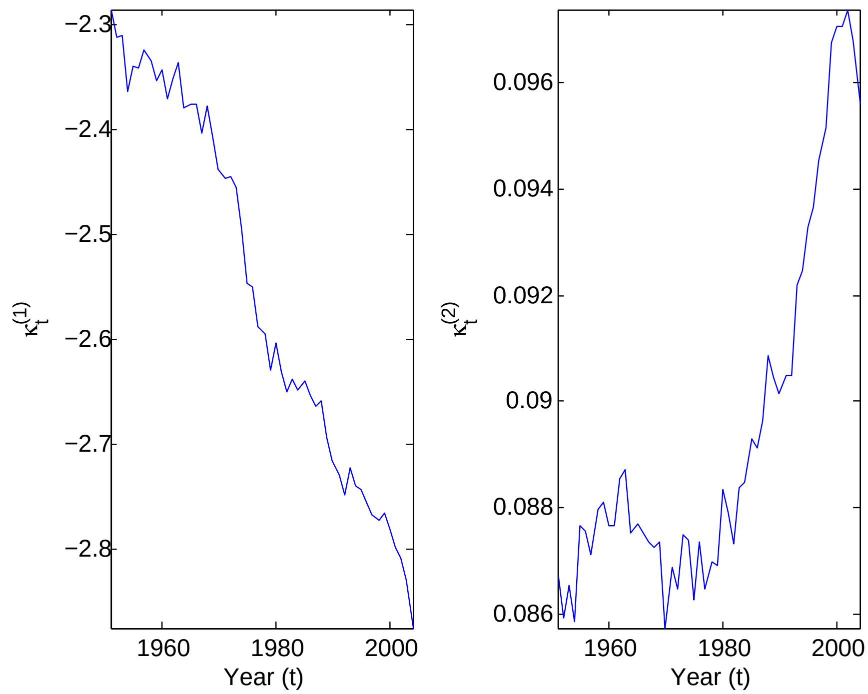



Figure 2: Maximum likelihood estimates of parameters in the Cairns-Blake-Dowd model, US population.

## 4.2 The Cairns-Blake-Dowd Model with a Cohort Effect

The original Cairns-Blake-Dowd model is a purely period effect model. It does not incorporate cohort effects, that is, the observed phenomenon that people born in certain years have experienced more rapid improvement in people born in other years. In a report by the Continuous Mortality Investigation Bureau (2002), it was noted that cohort effects are highly significant in the mortality experience of UK male pensioners and UK male insured lives. Therefore, in some situations, a model with a cohort effect is needed for adequate fit.

To model cohort effects, we may consider the following generalization of the Cairns-Blake-Dowd model:

$$\ln\left(\frac{q_{x,t}}{1-q_{x,t}}\right) = \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}) + \kappa_t^{(3)}((x-\bar{x})^2 - \hat{\sigma}_x^2) + \gamma_{t-x}^{(4)}, \tag{4}$$

where $\kappa_t^{(1)}$, $\kappa_t^{(2)}$, and $\kappa_t^{(3)}$ are period risk factors, $\gamma_{t-x}^{(4)}$ is a cohort risk factor, and $\hat{\sigma}_x^2$ is the mean of $(x-\bar{x})^2$ over the age range we consider.[1]

[1] This model is labeled as Model M7 in Cairns et al. (2009).

This generalization is different from the original Cairns-Blake-Dowd model in two ways. First, it contains a cohort risk factor $\gamma^{(4)}_{t-x}$ that is explicitly linked to the year of birth, $t-x$. The response to $\gamma^{(4)}_{t-x}$ is constant over age. Second, it includes a quadratic term $\kappa^{(3)}_t((x-\bar{x})^2-\hat{\sigma}^2_x)$ to capture the potential curvature in the relationship between $\ln\left(\frac{q_{x,t}}{1-q_{x,t}}\right)$ and $x$.

As with the Lee-Carter model, there is an identifiability problem. In more detail, if $\kappa^{(1)}_t$, $\kappa^{(2)}_t$, $\kappa^{(3)}_t$, and $\gamma^{(4)}_{t-x}$ are model parameters, then it can be shown that

$$\begin{aligned}\tilde{\kappa}^{(1)}_t &= \kappa^{(1)}_t + \phi_1 + \phi_2 t + \phi_3 t^2 + \phi_3\hat{\sigma}^2_x,\\ \tilde{\kappa}^{(2)}_t &= \kappa^{(2)}_t - \phi_2 - 2\phi_3 t,\\ \tilde{\kappa}^{(3)}_t &= \kappa^{(3)}_t + \phi_3,\end{aligned}$$

and

$$\tilde{\gamma}^{(4)}_{t-x} = \gamma^{(4)}_{t-x} - \phi_1 - \phi_2(t-x-\bar{x}) - \phi_3(t-x-\bar{x})^2$$

are also parameters of the model. We use the following constraints to stipulate parameter uniqueness:

$$\begin{aligned}&\sum_{x,t}\gamma^{(4)}_{t-x} = 0,\\ &\sum_{x,t}(t-x)\gamma^{(4)}_{t-x} = 0,\\ &\sum_{x,t}(t-x)^2\gamma^{(4)}_{t-x} = 0.\end{aligned}$$

The use of these constraints is equivalent to setting $\phi_1$, $\phi_2$, and $\phi_3$ to 0, thereby ensuring the fitted $\gamma^{(4)}_{t-x}$ will fluctuate around zero and will have no discernible linear trend or quadratic curvature. This model can also be estimated by the method of maximum likelihood, which is described in the Appendix. The parameter estimates are displayed graphically in Figures 3 and 4.

Having fitted equation (4) to historic data, the period indexes are modeled by a trivariate random walk with drift:

$$\kappa_{t+1} = \kappa_t + \mu + CZ(t+1), \tag{5}$$

where $\kappa_t = (\kappa^{(1)}_t, \kappa^{(2)}_t, \kappa^{(3)}_t)'$, $\mu = (\mu_1, \mu_2, \mu_3)'$ is a constant $3\times 1$ vector, $C$ is a constant $3\times 3$ upper triangular matrix, and $Z(t)$ is a 3-dimensional standard normal random vector.

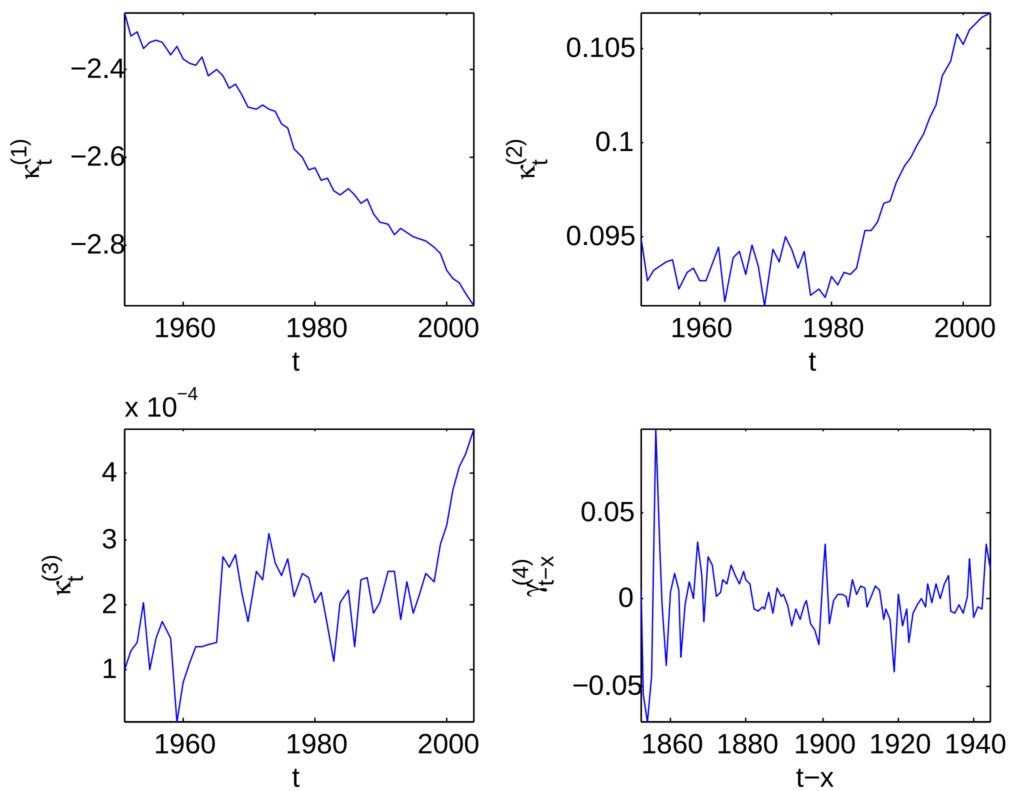



Figure 3: Maximum likelihood estimates of parameters in the generalized Cairns-Blake-Dowd model, Canadian population.

For the cohort of individuals who are aged $x$ at the forecast origin $T$, the best estimate of the death probability at time $T+s$ can be calculated with the following equation:

$$\begin{aligned} \ln\left(\frac{\hat{q}_{x+s,T+s}}{1-\hat{q}_{x+s,T+s}}\right) &= \kappa_T^{(1)}(s) + \kappa_T^{(2)}(s)(x+s-\bar{x}) \\ &\quad + \kappa_T^{(3)}(s)((x+s-\bar{x})^2 - \hat{\sigma}_x^2) + \gamma_{T-x}^{(4)}, \end{aligned} \tag{6}$$

where $\kappa_T^{(i)}(m) = \kappa_T^{(i)} + s\mu_i$, $i = 1, 2, 3$, is the MMSE forecast of $\kappa_{T+s}^{(i)}$.[2] We replace the unknown parameters with their estimates in actual calculations.

[2]The birth cohort used in our illustrations is involved in the data sample, so there is no need to extrapolate $\gamma_{t-x}^{(4)}$. To make a forecast for individuals who were born in later years, a process for $\gamma_{t-x}^{(4)}$ is needed. As Dowd et al. (2010b) suggest, $\gamma_{t-x}^{(4)}$, which has no long-term trend, may be modeled by an AR(1) process.

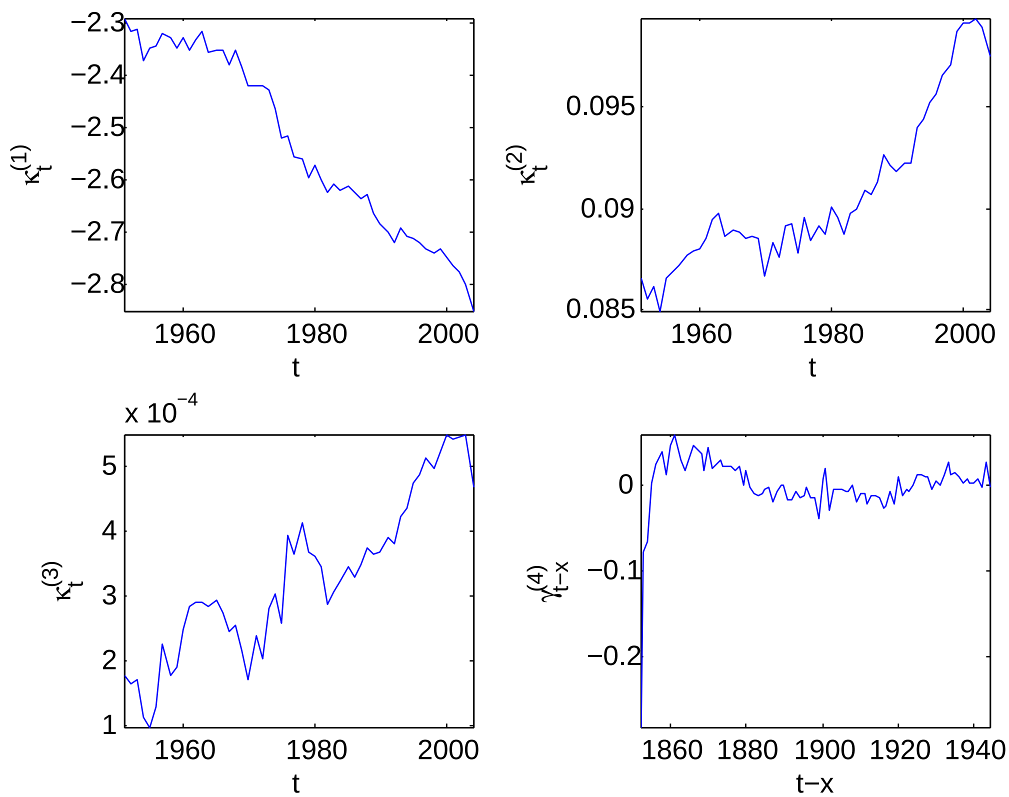



Figure 4: Maximum likelihood estimates of parameters in the generalized Cairns-Blake-Dowd model, US population.

# 5 Time-Simultaneous Prediction bands

Consider again the cohort of individuals who are aged $x$ at the forecast origin $T$. Assuming that the forecast horizon is $S$, the trajectory of interest would be $\mathbf{m} = (m_{x+1,T+1}, \ldots, m_{x+S,T+S})$. From the stochastic components of the mortality models, we can simulate realizations of this trajectory. We let

$$\mathbf{M} = \{\mathbf{m}^{(n)}\}_{n=1}^{N} = \{(m_{x+1,T+1}^{(n)}, \ldots, m_{x+S,T+S}^{(n)})\}_{n=1}^{N}$$

be a sample of $N$ simulated trajectories. This sample is called the learning sample, which can be used to derive prediction bands.

Prediction intervals provided in previous research are mostly pointwise prediction intervals. They are calculated by considering each death rate in isolation. In particular, for each $s = 1, \ldots, S$, the sample $\{m_{x+s,T+s}^{(n)}\}_{n=1}^{N}$ is ordered. Let $PI_s = [l_s, h_s]$ be a pointwise prediction interval for $m_{x+s,T+s}$ with coverage probability $1-\alpha$. Then the limits $l_s$ and $h_s$ are set to the $\lfloor \alpha/2 \rfloor$th lowest and highest values in the sample, respectively. Note that, in general, the fraction of trajectories that are completely inside all $S$ pointwise prediction intervals $PI_s$, $s = 1, \ldots, S$, is less than $1 - \alpha$.

Our goal is to construct a time-simultaneous prediction band such that $\lceil (1 - \alpha)N \rceil$ of the trajectories in $\mathbf{M}$ are completely inside the band at every time point, while $\lfloor \alpha N \rfloor$ are outside the band at one or more points of time. By construction the band will then contain a randomly selected trajectory $\mathbf{m}^{(n)}$ in the sample $\mathbf{M}$ with probability (no less than) $1 - \alpha$.[3]

## 5.1 Adjusted Intervals

Kolsrud (2007) proposes a simple and yet intuitive method, which he calls 'adjusted intervals,' to construct a time-simultaneous prediction band from a learning sample. The idea behind this method is to widen the pointwise intervals (with pointwise coverage probability $1 - \alpha$) uniformly until the band of intervals has simultaneous a coverage of $1 - \alpha$. We can implement this method with the following algorithm:

1. For each $s =, 1, \ldots, S$, widen the interval uniformly to include the nearest sample point above and the nearest sample point below.

2. Check the simultaneous coverage of all intervals in the learning sample $\mathbf{M}$.

3. If the simultaneous coverage is less than the prescribed level $1 - \alpha$, go to Step (1). Otherwise, terminate the algorithm. The resulting band of intervals would contain no less than $1 - \alpha$ of the trajectories in the learning sample.[4]

Following the algorithm above, we construct time-simultaneous prediction bands for the death rates associated with the birth cohort in question, on the basis of the learning samples generated from the Cairns-Blake-Dowd model and its generalization. The resulting time-simultaneous prediction bands are shown in Figures 5 and 6. Also shown in the figures are the pointwise confidence intervals and the mean forecasts, which are obtained by averaging the values in the learning sample for each $s = 1, \ldots, S$.

[3]In a finite sample the fractions are discrete, and may only be approximately equal to a prescribed level.

[4]The final coverage in the sample might be larger than $1 - \alpha$, because each uniform widening of all intervals includes at least two new trajectories.

We observe from Figures 5 and 6 that the time-simultaneous prediction bands are significantly wider than the corresponding pointwise intervals. In particular, the pointwise intervals (with pointwise coverage of 95%) can only capture 68-69% of the trajectories in the learning sample. It is clear from this illustration that pointwise confidence intervals may seriously understate the uncertainty associated with random trajectories.

It is also interesting to note that the prediction band derived from the generalized model with a cohort effect is considerably wider than that from the original model. This is because the generalized version, which contains more parameters (stochastic factors), has a less stringent model structure, imposing less restrictions to the dynamics of future mortality rates.

## 5.2 Chebyshev Bands

Another method proposed by Kolsrud (2007) is called 'Chebyshev bands.' To explain Chebyshev bands, we need to define the envelope of a sample:

**Definition 3.** *The envelope of a (sub)sample is the tightest band that contains all trajectories in the (sub)sample.*

As an example, the envelope of the learning sample $\mathbf{M}$ can be expressed as $([\min_n m^{(n)}_{x+s,T+s}, \max_n m^{(n)}_{x+s,T+s}])^S_{s=1}$. The idea behind Chebyshev bands is that we construct a time-simultaneous prediction band as the envelope of a subsample $\mathbf{M}^*$ that contains $\lceil(1-\alpha)N\rceil$ trajectories with the shortest distance to the mean trajectory

$$\bar{\mathbf{m}} = (\bar{m}_{x+1,T+1}, \ldots, \bar{m}_{x+S,T+S}),$$

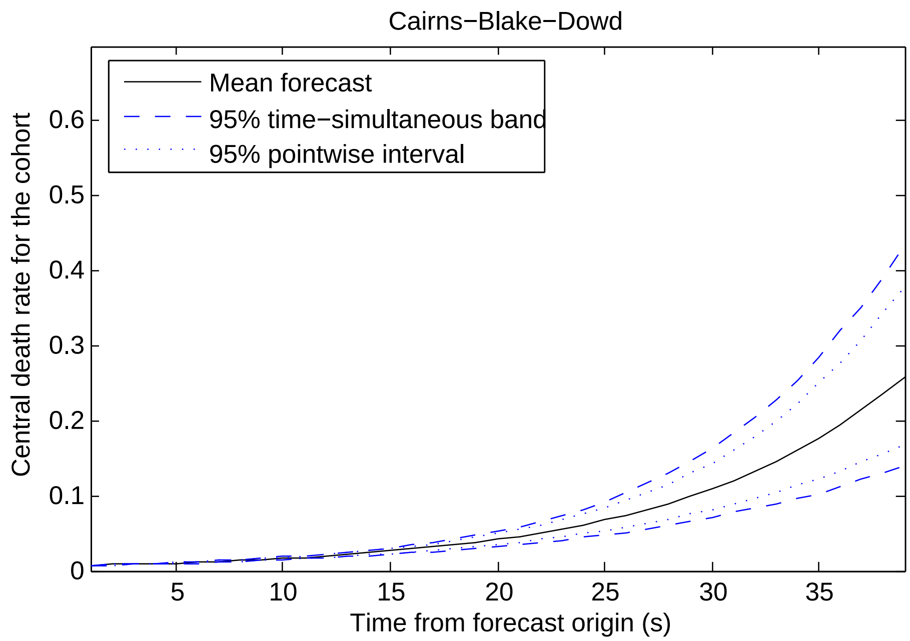
where $\bar{m}_{x+s,T+s} = \frac{1}{N}\sum_{n=1}^{N} m^{(n)}_{x+s,T+s}$ is the pointwise mean $s$ steps beyond the forecast origin.

Kolsrud (2007) suggests that we may measure the distance to the mean trajectory with the weighted Chebyshev distance, which can be expressed as:

$$\max_{s=1,\ldots,S}\left(\frac{|m_{x+s,T+s} - \bar{m}_{x+s,T+s}|}{\sigma_s}\right),$$

where

$$\sigma_s = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(m_{x+s,T+s} - \bar{m}_{x+s,T+s}\right)^2}$$



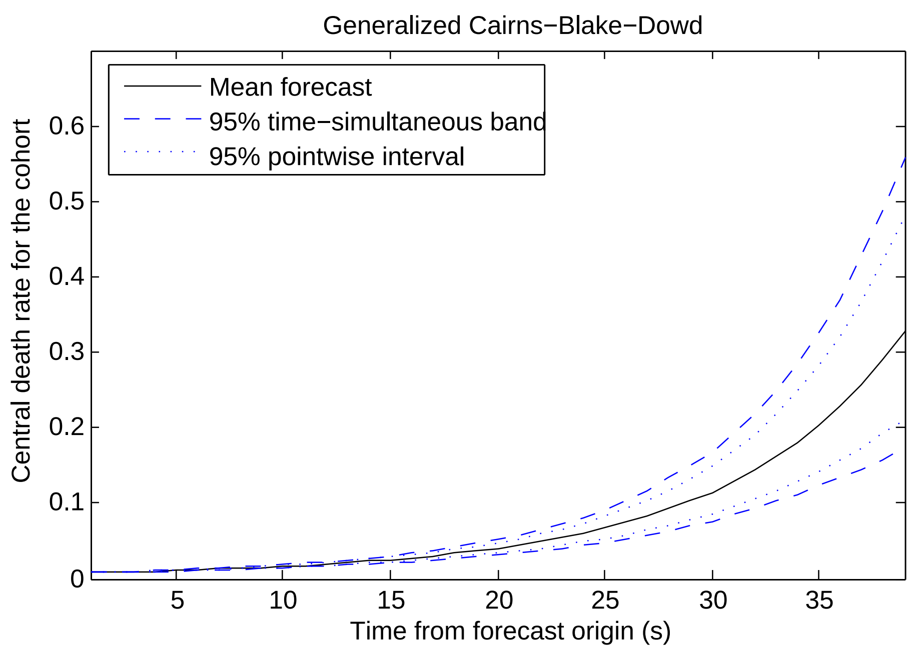




Figure 5: Forecasts of the central death rate ($m_{60+s,2004+s}$) for the cohort aged 60 in year 2004, Canadian population. The time-simultaneous prediction bands are constructed by the method of adjusted intervals.

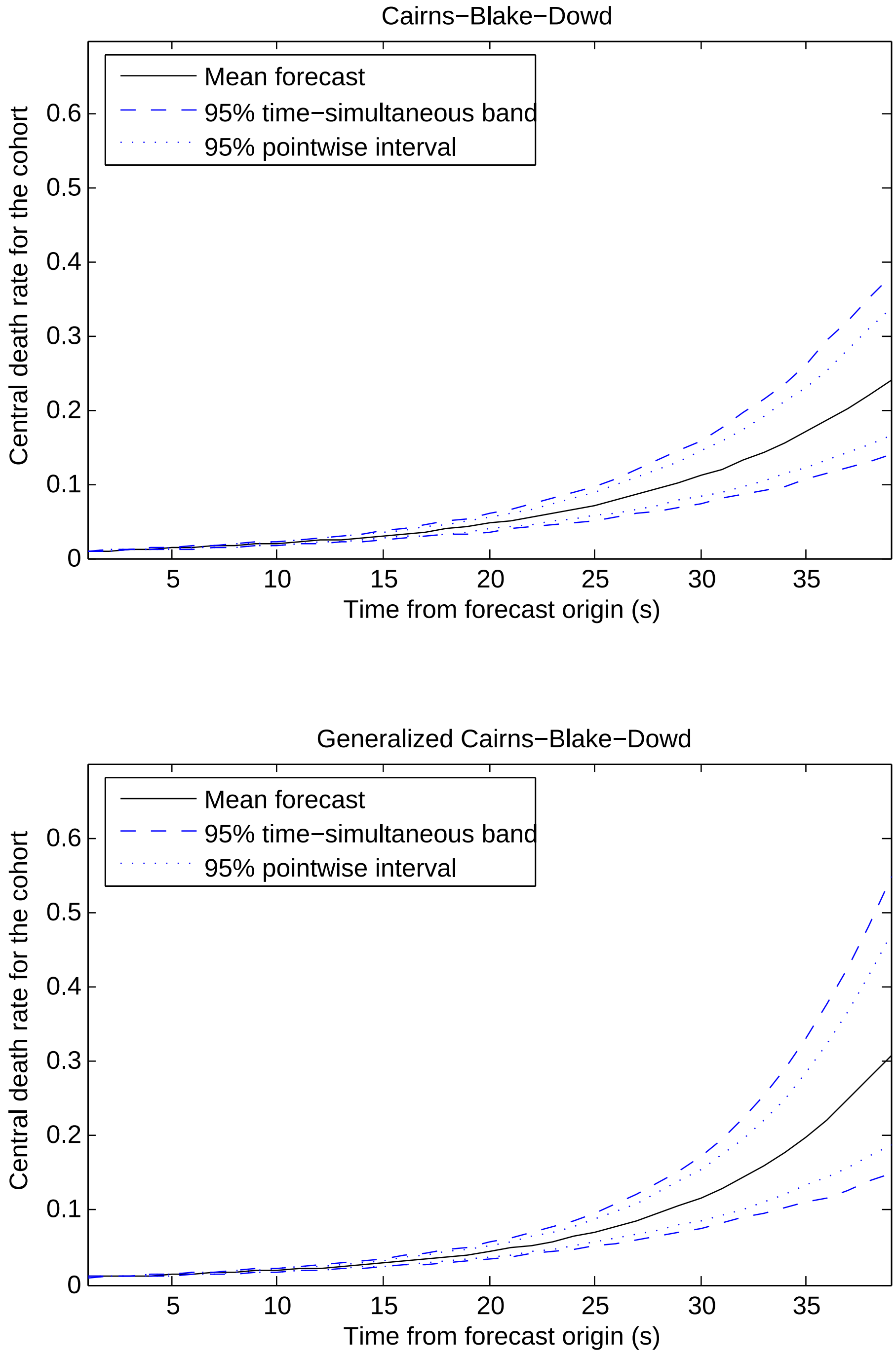



Figure 6: Forecasts of the central death rate ($m_{60+s,2004+s}$) for the cohort aged 60 in year 2004, US population. The time-simultaneous prediction bands are constructed by the method of adjusted intervals.

is the pointwise standard deviation $s$ steps beyond the forecast origin. Contrary to most geometric or probabilistic measures of distance, the weighted Chebyshev distance relates directly to the $S$-dimensional rectangular shape of a band. The distances are weighted by the pointwise standard deviations to take possible heteroskedasticity in the learning sample into account. This issue is important in our application because the volatility of the simulated death rates increases with both age and the distance from the forecast origin.

Using the method described above, we construct time-simultaneous prediction bands from the learning samples that are based on the two mortality models we consider. The resulting prediction bands are displayed in Figures 7 and 8. For the reader's reference, we also display the corresponding pointwise confidence intervals, mean forecasts and envelopes of the learning samples.

As with those constructed by the other numerical method, the time-simultaneous prediction bands in Figures 7 and 8 are significantly wider than the corresponding pointwise intervals. We also observe that the generalized Cairns-Blake-Dowd model yields a more conservative prediction band, because, as we mentioned earlier, it is less restrictive than the original version. The widths of the bands constructed by both numerical methods are very close to each other, with an average percentage difference of less than 5%.

# 6 Concluding Remarks

Future mortality rates are difficult to predict, so measures of uncertainty such as prediction intervals are particularly important to users of mortality projections. We have demonstrated that pointwise prediction intervals, which are often provided in previous mortality studies, can significantly understate the uncertainty associated with a random mortality trajectory. The use of a time-simultaneous prediction band is strongly recommended when the user demands a forecast of a whole path of cohort death rates.

We have introduced two numerical methods for constructing a time-simultaneous prediction band, namely adjusted intervals and Chebyshev bands. These methods can be applied to all stochastic mortality models from which sample paths of future mortality rates can be generated. They do not require knowledge or assumptions about the simultaneous distribution of the random trajectory. Other than cohort

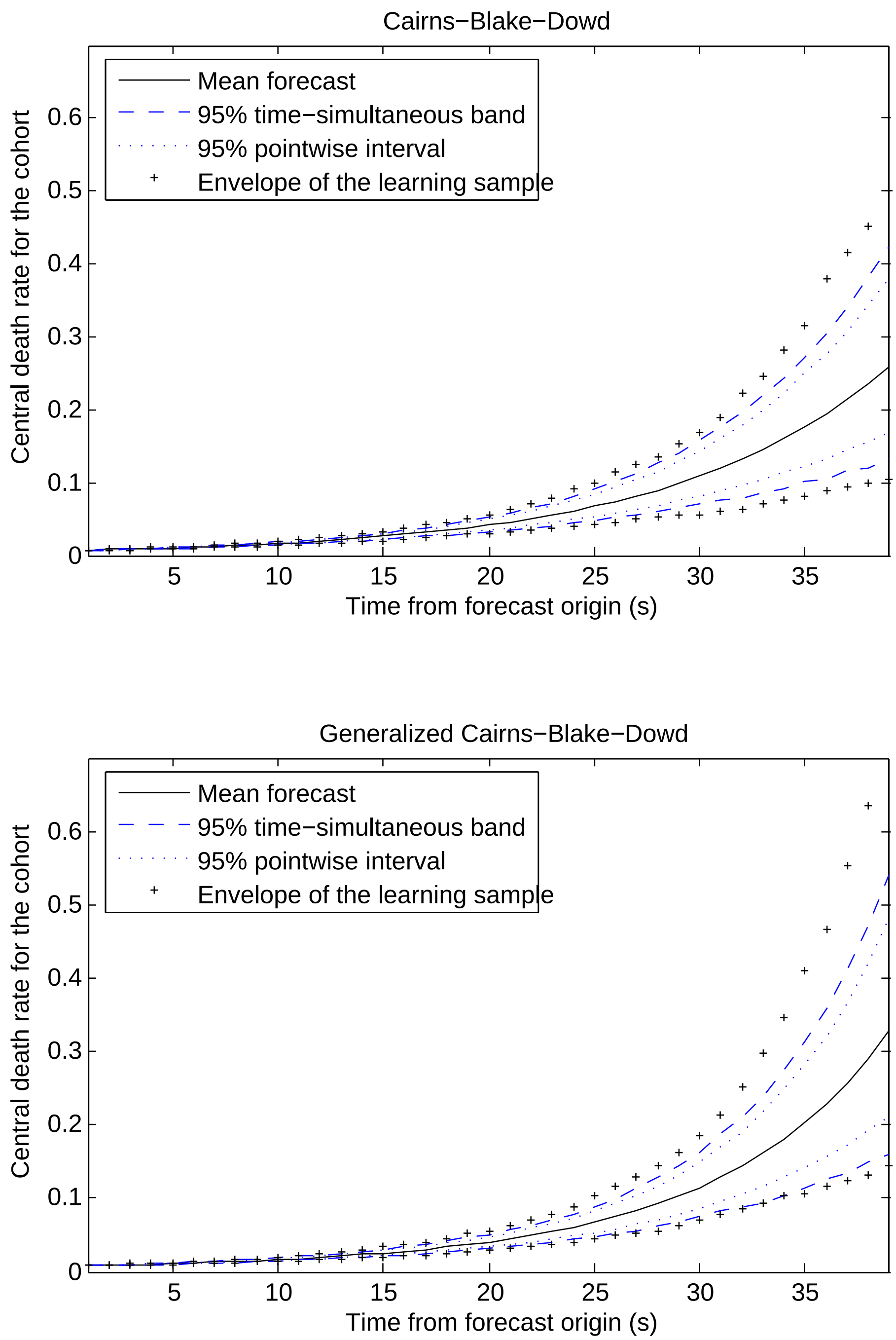
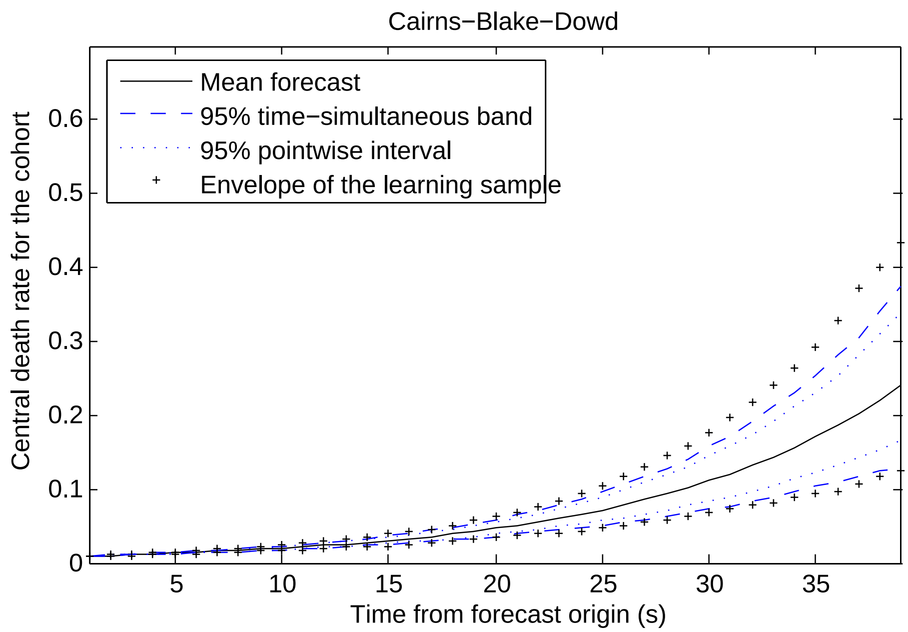



Figure 7: Forecasts of the central death rate ($m_{60+s,2004+s}$) for the cohort aged 60 in year 2004, Canadian population. The time-simultaneous prediction bands are constructed by the method of Chebyshev bands.



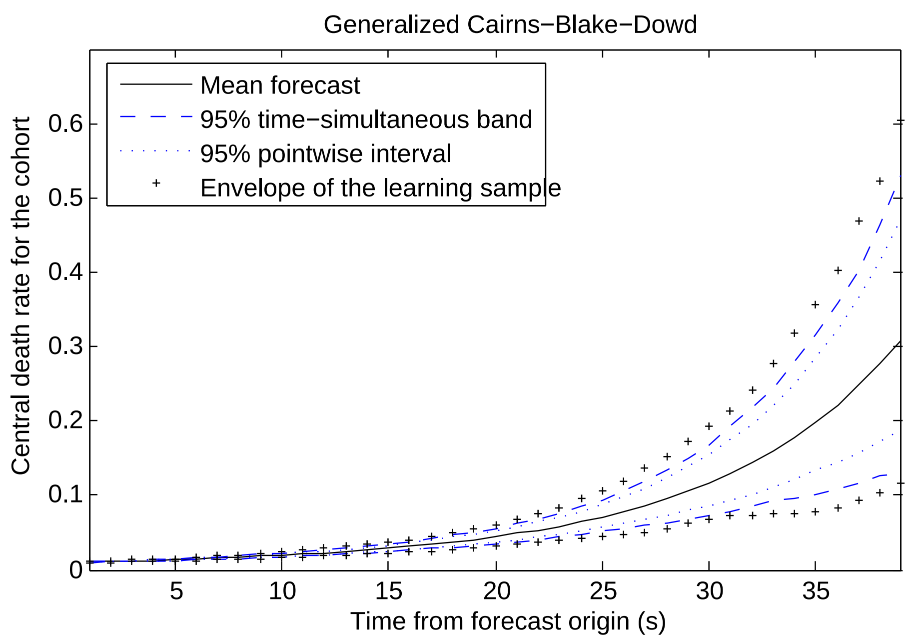




Figure 8: Forecasts of the central death rate ($m_{60+s,2004+s}$) for the cohort aged 60 in year 2004, US population. The time-simultaneous prediction bands are constructed by the method of Chebyshev bands.

mortality rates, they can be applied to various demographic quantities, such as period mortality rates and period life expectancies, by adjusting the definition of the trajectory and the learning sample accordingly.

The problem of model risk has not been taken into account in this study. Denuit (2009) proposes handling model risk by considering a set of different mortality projection models. Given the available data, each model is assigned a weight that is determined by a model selection criterion such as BIC (Schwarz, 1978). One avenue for future research is to explore how this method can be integrated into the construction of time-simultaneous prediction bands.

# Acknowledgments

This paper was presented in the Fourth Living to 100 International Symposium, organized by the Society of Actuaries in January 2011. We sincerely thank the discussant, Mr. Thomas Edwards, and other participants of the symposium for their stimulating discussions.

This work was supported by a grant from the Natural Science and Engineering Research Council of Canada and by a direct grant from the Chinese University of Hong Kong (Project No. 2070401).

# References

Cairns, A.J.G. (2000). A Discussion of Parameter and Model Uncertainty in Insurance. *Insurance: Mathematics and Economics*, **27**, 313-330.

Cairns, A.J.G., Blake, D. and Dowd, K. (2006). A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration. *Journal of Risk and Insurance*, **73**, 687-718.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., Ong, A., and Balevich, I. (2009). A Quantitative Comparison of Stochastic Mortality Models Using Data from England and Wales and the United States. *North American Actuarial Journal*, **13**, 1-35.

Continuous Mortality Investigation Bureau (1999). Standard Tables of Mortality Based on the 1991-94 Experiences. CMI Report no. 17. London: Institute of Actuaries and

Faculty of Actuaries.

Continuous Mortality Investigation Bureau (2002). An Interim Basis for Adjusting the "92" Series Mortality Projections for Cohort Effects. CMI Working Paper 1. London: Institute of Actuaries and Faculty of Actuaries.

Continuous Mortality Investigation Bureau (2004). Projecting Future Mortality: A Discussion Paper. CMI Working Paper no. 3. London: Institute of Actuaries and Faculty of Actuaries.

Delwarde, A., Denuit, M. and Partrat, C. (2007). Negative Binomial Version of the Lee-Carter Model for Mortality Forecasting. *Applied Stochastic Models in Business and Industry*, **23**, 385-401.

Denuit, M. (2009). *Dynamic Life Tables: Construction and Applications.* Keynote Lecture in the International Association of Actuaries Life Colloquium, Munich, Germany.

Dowd, K., Blake, D. and Cairns, A.J.G. (2010a). Facing up to Uncertain Life Expectancy: The Longevity Fan Charts. *Demography*, **47**, 67-78.

Dowd, K., Cairns, A.J.G., Blake, D., Coughlan, G.D., Epstein, D. and Khalaf-Allah, M. (2010b). Backtesting Stochastic Mortality Models: An Ex-Post Evaluation of Multi-Period-Ahead Density Forecasts. Pensions Institute Discussion Paper PI-0803. Available at http://www.pensions-institute.org/workingpapers/wp0803.pdf.

Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute of Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on 1 Jan 2010).

Kolsrud, D. (2007). Time-Simultaneous Prediction Band for a Time Series. *Journal of Forecasting*, **26**, 171-188.

Lee, R. and Carter, L. (1992). Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association*, **87**, 659-671.

Li, J.S.H., Hardy, M.R. and Tan, K.S. (2007). Report on Mortality Improvement Scales for Canadian Insured Lives. Available at http://www.soa.org/files/pdf/cia-mortality-rpt.pdf.

Li, J.S.H., Hardy, M.R. and Tan, K.S. (2009). Uncertainty in Mortality Forecasting: An Extension to the Classical Lee-Carter Approach. *ASTIN Bulletin*, **39**, 137-164.

Nair, V.N. (1984). Confidence Bands for Survival Functions with Censored Data: A Comparative Study. *Technometrics*, **26**, 265-275.

Parigi, G. and Schlitzer, G. (1995). Quarterly Forecasts of the Italian Business Cycle by Means of Monthly Economic Indicators. *Journal of Forecasting*, **14**, 117-141.

Renshaw, A.E. and Haberman, S. (2003). Lee-Carter Mortality Forecasting with Age-specific Enhancement. *Insurance: Mathematics and Economics*, **33**, 255-272.

Renshaw, A.E. and Haberman, S. (2006). A Cohort-Based Extension to the Lee-Carter Model for Mortality Reduction Factors. *Insurance: Mathematics and Economics*, **38**, 556-570.

Scheike, T.H. and Zhang, M-.J. (2003). Extensions and Applications of the Cox-Aalen Survival Model. *Biometrics*, **59**, 1036-1045.

Schwarz, G. (1978). Estimating the Dimension of a Model, *Annals of Statistics*, **6**, 461-464.

Tuljapurkar, S. (2005). Future Mortality: A Bumpy Road to Shangri-La? *Science of Aging Knowledge Environment*, **14**, pe9.

# Appendix

### Maximum Likelihood Estimation for the Stochastic Mortality Models

All models in this paper can be fitted by the method of maximum likelihood. Let us define $D_{x,t}$ by the number of deaths at age $x$ and in year $t$, and $E_{x,t}$ by the corresponding exposures to the risk of death. To construct the likelihood function, we treat $D_{x,t}$ as independent Poisson responses, that is,

$$D_{x,t} \sim \text{Poisson}(E_{x,t} m_{x,t}),$$

where $E_{x,t} m_{x,t}$ is the expected number of deaths at age $x$ and in year $t$. This gives the following log-likelihood, which is applicable to all three models:

$$l = \sum_{x,t} (D_{x,t} \ln(E_{x,t} m_{x,t}) - E_{x,t} m_{x,t} - \ln(D_{x,t}!)), \tag{7}$$

where $D_{x,t}!$ stands for $D_{x,t}$ factorial. The summation is taken over all $x$ in the sample age range and all $t$ in the sample period.

For the Cairns-Blake-Dowd model and its generalization, which are based on $q_{x,t}$ rather than $m_{x,t}$, we use the following relation:

$$m_{x,t} = -\ln(1 - q_{x,t}), \tag{8}$$

which holds if we assume that the force of mortality is constant over each year of integer age and over each calendar year. The required likelihood functions can be obtained by substituting the model equations into equation (8) and then into equation (7).

Parameter estimates can be obtained by maximizing the corresponding likelihood function. The maximization can be accomplished by an iterative Newton-Raphson method, in which parameters are updated one at a time. The updating of a typical parameter $\theta$ proceeds according to

$$u(\theta) = \theta - \frac{\partial l / \partial \theta}{\partial^2 l / \partial \theta^2},$$

where $u(\theta)$ is the updated value of $\theta$ in the iteration. The parameter constraints (if any) are applied at the end of each iteration.