The Two-Factor Cairns- Blake- Dowd Mortality Model for Forecasting Egyptian Mortality Rates with StMoMo in R

submitted by

Dr. Mohamed Ahmed Abdel Naby

Lecturer,

Department of mathematics, Insurance &Statistics,

Sadat Academy for Management Sciences- Egypt

E- mail: mohamed24sa@yahoo.com

Dr. Dina Talaat Hamdy Neel

Lecturer,

of Mathematics & Insurance

Faculty of Business Administration

Hours University- Egypt

E- mail: dtalaat@horus.edu.eg.

Raya International Journal of Business Sciences

volume (3), Issue (11), october2024

https://www.rijcs.org/

Publisher

Raya Higher Institute of Management and Foreign Trade in New Damietta

The Two-Factor Cairns- Blake- Dowd Mortality Model for Forecasting Egyptian Mortality Rates with StMoMo in R

submitted by

Dr. Mohamed Ahmed Abdel Naby

Lecturer,

Department of mathematics, Insurance &Statistics,

Sadat Academy for Management Sciences- Egypt

E- mail: mohamed24sa@yahoo.com

Dr. Dina Talaat Hamdy Neel

Lecturer,

of Mathematics & Insurance

Faculty of Business Administration

Hours University- Egypt

E- mail: dtalaat@horus.edu.eg.

ABSTRACT

Forecasting mortality rates is crucial for insurance companies to accurately pricing policies and manage risks.accurate forecasted mortality rates enable insurance companies to minimize losses, set reasonable premiums, introduce new policies, and manage longevity risks effectively.

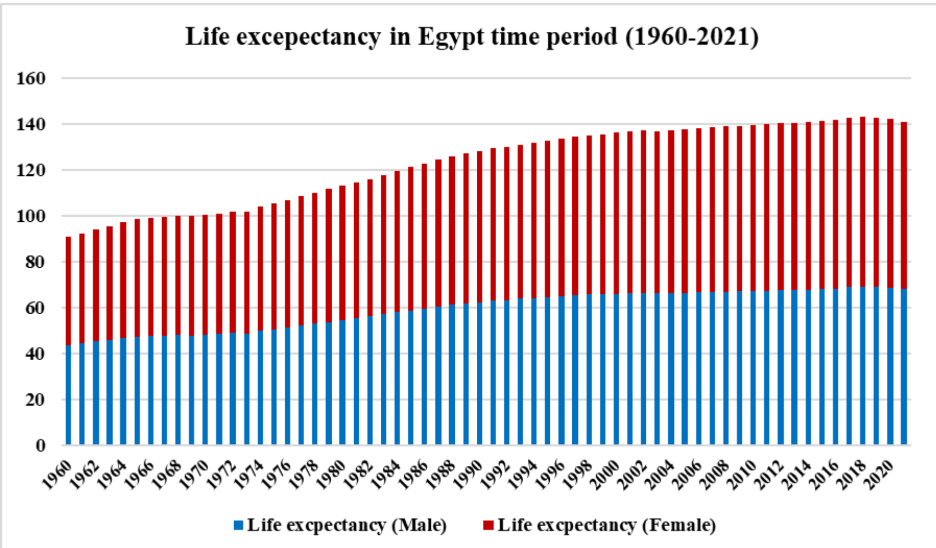
Improved forecasting accuracy not only benefits insurance companies but also impacts pension funds, life insurers, and social securities, making it a vital aspect of actuarial science and demographic statistics. based on this, this research provides the two factor Cairns-Blake-Dowd (CBD) model to forecasting Egyptian mortality rates using R package StMoMo. Estimated of CBD parameters model by maximizing the model log- likelihood, and the multivariate random walk with drift will be used

to forecast the CBD model's parameters. The model applied on both males and females data for the period (1960-2023). The forecast is for period (2024-2043). The main finding is that CBD model is useful for specific age ranges (elderly ages) and time periods. And there is an increasing in life expectancy for both males and females in Egypt.

Key words: Two factor model, Cairns- Blake- Dowd model, log likelihood, multivariate random walk with drift, StMoMo package.

1. Introduction

Official studies show an increase in life expectancy for both males and females in Egypt, for time period (1960-2021). as shown by next figure:



Source:

https://data.worldbank.org/indicator/SP.DYN.LE00.FE.IN?locations=EG

Figure (1)

Life expectancy in Egypt time period (1960-2021)

This figure shows increasing in life expectancy in Egypt for both Male and Female during the period 1960 to 2022. We can refer this increase to medical health improvements.

This increase poses a challenge to governments, private pension plans and life insurance companies because of its impact on costs of pension and medical care.

Actuaries and demographers pay attention to the problems caused by an ageing population and longevity, so they give considerable focus to development statistical techniques for modelling and forecasting mortality rates.

Forecasting mortality rates is crucial for insurance companies to accurately pricing policies and manage risks. accurate forecasted mortality rates enable insurance companies to minimize losses, set reasonable premiums, introduce new policies, and manage longevity risks effectively **Mitchell, Brockett, Mendoza-Arriaga, & Muthuraman, (2013**). Improved forecasting accuracy not only benefits insurance companies but also impacts pension funds, life insurers, and social securities, making it a vital aspect of actuarial science and demographic statistics.

It is necessary to have a model to measure the impact of changes in the mortality rates of a society, which may be a result of changes in the (health, economic or social level)، the Cairns-Blake-Dowd (CBD) models **Li & O'Hare (2017)** were proposed to better capture mortality at ages 55 and over. These CBD models model the logit of the death probabilities at older ages as a linear or quadratic function of age, thereby treating the intercept and slope parameters across years as stochastic processes.

Many studies focusing on forecasting mortality by CBD model and others on importance of mortality forecasting for elderly ages specifically. **Redzwan & Ramli**

(2022) conducted bibliometric analysis on mortality forecasting and it is important for quantifying mortality and longevity risks in the insurance and actuarial fields. And the research process had three main steps. first, searching the Scopus database for relevant publications, second, applying inclusion/ exclusion criteria to select the final set of 138 documents, third, conducting the bibliometric analysis using the two approaches. The study finds that publications on mortality forecasting models have been increasing, with fifty percent published between 2016-2021, indicating growing interest in this field.

Where the study of **Odhiambo, Weke, & Ngare, (2021)** aims at integrating the original Cairns- Blake- Dowd (CBD) formulation by the introduction of an artificial Recurrent Neural Networks short- term long memory, when forecasting future evolution of the $k_t^{(2)}$ parameter thus overcoming the challenges showed by the traditional ARIMA (p, d, q) time series process. The result of this study is that this approach has shown high efficiency when modeling mortality patterns and forecasted values as opposed.

Apicella, Dacorogna, Di Lorenzo, & Sibillo, (2019) proposes a new dynamic corrective methodology to improve the predictive accuracy of the Cairns-Blake-Dowd (CBD) mortality projection model. In this study Various out-of-sample validation methods, including both a static and a dynamic approach to capture changes in the underlying data. The data is the Italian and French females mortality data.

Zulkifle, Yusof, & Nor, (2019) compares between two mortality forecasting model Lee Carter (LC) & Cairns Blake Dowd (CBD). using Malaysian data from the Department of Statistics Malaysia, covering the years 1980-2017, with a focus on ages 55–80 years. This study shows that LC uses the log of the central rate of mortality, whereas CBD uses the logit of the odd mortality as the variable that is dependent. Analysis of comparisons was performed using Akaike information criteria (AIC) and Bayesian information criterion (BIC). The results of this study indicate that both models show a good fit and forecast accuracy over various age range and sample period phases.

Since **Safitri, Mardiyati, & Malik, (2019)** applied the Cairns-Blake-Dowd mortality forecasting model to forecast Indonesian mortality rate data from 1950 to 1955 to 2010 -2015 with age group 0,1-4,5-9 and so on up to 80-84 years old. The least-squares method was used to estimate the parameters of the CBD model and bivariate random walk with drift to obtain the forecasted values of the parameters for the next three periods. The final results show that the forecasted Indonesian mortality rates for the next three periods show a downward trend. And the study introduced by **Andres, Millossovich & Vladimir, (2018)** provides The R package StMoMo tools, which define a family of generalized age-period-cohort stochastic mortality models for fitting stochastic mortality models, assessing the goodness of fit of these models, and projecting stochastic mortality models, including the CBD model. This study explains the capabilities of the StMoMo package by comparing several stochastic mortality models applied to England and Wales populations.

According **(Janssen, 2018)** Mortality forecasts are essential for predicting population aging, determining pension sustainability, setting life insurance premiums, and planning for changing societal needs. The study shows recent pension reforms have linked retirement age and payments to anticipated mortality and life expectancy. And the field of mortality forecasting is growing and advancing due to the increasing societal relevance of accurate forecasts.

Maccheroni & Nocito, (2017) compares the performance of the Lee-Carter and Cairns-Blake-Dowd mortality forecasting models on Italian death rates using backtesting analysis. Three different backtest approaches to evaluate short-run and medium-length forecasts of the models. The data used in this study is the Italian mortality from National Statistics Institute database for the period 1975 to 2014. This study finds that The Cairns-Blake-Dowd (CBD) model forecasts are more reliable for ages above 75.

From the insight of **Chan, Li, J. S.-H., & Li, J. (2014)** the CBD mortality model provides indexes that can be used to indicate and model longevity risk, unlike other extrapolative mortality models. This is done by fitting historical data to stochastic mortality models to identify parameters that vary during time. Focusing on the CBD model parameters as longevity risk indexes, and modeling them using multivariate time series models, we study the joint prediction region of mortality indexes as a longevity risk metric. The main finding is that the CBD model parameters are suitable as indexes to indicate longevity risk levels over time.

The study of **Sweeting, (2011)** extended the Cairns-Blake-Dowd mortality model to account for trend changes, resulting in greater long-term uncertainty in life expectancy projections. This study used statistical techniques to identify points of statistically significant variations in the trends of the two factors in the Cairns-Blake-Dowd mortality model. The frequency that follows each trend varies was then applied to forecast the frequency of upcoming modifications, and the sizes of historical changes were used to project the sizes of future changes. The results showed that the two factors in the Cairns-Blake-Dowd mortality model do not follow a random walk, but rather a random fluctuation around a trend that changes periodically. In addition, modeling mortality in this manner leads to greater long-term uncertainty in projections of future life expectancy.

2. **Research objective**:

The objective of this research summarized as follow:

1. Apply the two factor model Cairns-Blake-Dowd (CBD) model to identify and estimate both time index and cohort index to the Egyptian population (Male & Female).
2. Using **R** package **StMoMo**, which makes use of the two factor models unifying form to present computational tools for implementing many stochastic mortality models, assessing their goodness of fit and performing mortality forecasting.

3. **Research importance**

The importance of this study stems from the importance of mortality rates. Fertility, migration, and mortality are the three socioeconomic variables which influence the rate of population growth.

Where mortality rates are defined from the actuarial point of view as the probability of someone dying during a certain period. We can say that mortality rates are considered one of the most important demographic indicators in society, because of their importance in developing development plans in society, and a basis for calculating life insurance annuities. The stochastic nature of decrements must be taken into consideration in actuarial calculations for life annuity portfolios. This includes the theoretically diversifiable risk of unexpected changes surrounding the life table and the systematic longevity risk resulting from the unknown underlying life table.

Also, for social insurance assessing old-age pensions in an equitable manner helps to uphold the principle of justice and ensure equitable distribution between all ages.

4. Research methodology

The methodology of this research is as follow:

4.1. Cairns- Blake- Dowd CBD mortality model.

4.2. Forecasting Egyptian mortality rates

4.1. Cairns- Blake- Dowd CBD mortality model

This section illustrates the CBD model for forecasting mortality by two factors. *The first* influences the mortality rate uniformly across all age groups, but *the second* modifies the dynamics of the mortality rate differently at later ages than at younger ages. This model can be expressed as follow **(Safitri et al., 2019)** :

$$\text{logit}(q_{x,t}) = k_t^{(1)} + k_t^{(2)}(x - \bar{x}), \qquad (1)$$

Where:

- $q_{x,t}$ is the mortality rate.
- $k_t^{(1)}$:represent the intercept.
- $k_t^{(2)}$: represent the slope.
- $\bar{x}$: is the mean age range under consideration and it is calculated by $\sum \frac{x_i}{n_x}$
- x is the age group $x = x_1, \dots, x_p$
- t is the period $t = t_1, \dots, t_q$
- $k_t^{(1)}, k_t^{(2)}$: the period mortality index (mean and variance of mortality rates for a period).

For each year t , this defines a Gompertz law of mortality with $k_t^{(1)}$ as intercept, and $k_t^{(2)}$ as the slope. The sets of coefficients $\left\{k_t^{(1)}, t = 1, \dots, n_t\right\}$ and $\left\{k_t^{(2)}, t = 1, \dots, n_t\right\}$ then describe how the Gompertz line changes in time. The idea is to forecast the mortality by projecting the period effects $k_t^{(1)}, k_t^{(2)}$ applying a bivariate random walk with drift. A linear function of age is the log of the force of mortality.

- **Model fitting**

Estimation of CBD stochastic mortality model parameter obtained by maximizing the model log- likelihood **(Andres et al., 2018)**, which given by:

$$\mathcal{L}(d_{xt,}\hat{d}_{xt}) = \sum_x \sum_t \omega_{xt} \{d_{xt} \log \hat{d}_{xt} - \hat{d}_{xt} - \log d_{xt}!\} \qquad (2)$$

ω_{xt} is weight taking the value 0 if particular (x, t) data cell is omitted or 1 if the cell is included, and the expected number of deaths predicted by the model, with g^{-1} denoting the inverse of the link function g

$$\hat{d}_{xt} = E_{xt} g^{-1}\left(\alpha_x + \sum_{i=1}^{N} \beta_x^{(i)} k_t^{(i)} + \beta_x^{(0)} \gamma_{t-x}\right) \qquad (3)$$

- **Goodness of fit analysis**

This is analyzed by inspecting the residuals of the fitted model. Since regular trends in the residuals indicate the inability of the model to describe all the features of the data accurately. The residuals of scaled deviation are defined as **(Andres et al., 2018)**:

$$r_{xt} = sign(d_{xt} - \hat{d}_{xt}) \sqrt{\frac{dev(x,t)}{\hat{\phi}}}, \qquad \hat{\phi} = \frac{D(d_{xt}, \hat{d}_{xt})}{K - \nu} \qquad (4)$$

Since

$$dev(x,t) = 2d_{xt} \log\left(\frac{d_{xt}}{\hat{d}_{xt}}\right) - (d_{xt} - \hat{d}_{xt}) \qquad (5)$$

And random component is the entire models deviation of the model, defined as:

$$D(d_{xt}, \hat{d}_{xt}) = \sum_{x} \sum_{t} \omega_{xt} dev(x,t) \qquad (6)$$

Since

- $K = \sum_x \sum_t \omega_{xt}$: is the number of observations in the data.
- ν: is the effective number of parameters in the model.

- **Forecasting and simulation with CBD model**

According to CBD model the dynamics of mortality are driven by:

- The period indexes $k_t^{(i)}$, $i = 1, \ldots, N$
- The cohort index γ_{t-x}

So, Modeling these indicators using time series techniques is necessary for the forecasting and simulation of mortality rates. For modelling the period index use a multivariate random walk with drift **(Andres et al., 2018)** assuming that

$$k_t = \delta + k_{t-1} + \xi_t^k, \quad k_t = \begin{pmatrix} k_t^{(1)} \\ \vdots \\ k_t^{(N)} \end{pmatrix}, \quad \xi_t^k \sim N(0, \Sigma) \qquad (7)$$

Where:

δ: is an N- dimension vector of drift parameters.

Σ: is the $N \times N$ variance- covariance matrix of the multivariate white noise ξ_t^k

For the cohort index $\gamma_c \equiv \gamma_{t-x}$, assume that it follows a univariate ARIMA(p, q, d) process with drift, which is independent of the period index k_t, so that

$$\Delta^d \gamma_c = \delta_0 + \phi_1 \Delta^d \gamma_{c-1} + \cdots + \phi_p \Delta^d \gamma_{c-p} + \epsilon_c + \delta_1 \epsilon_{c-1} + \cdots + \delta_q \epsilon_{c-q} \qquad (8)$$

Since:

- Δ: is the difference operator.
- δ_0: is the drift parameter.
- $\phi_1, \cdots, \phi_p$: are the autoregressive coefficients with $\phi \neq 0$,
- $\delta_1, \cdots, \delta_q$: are the moving average coefficients with $\delta_q \neq 0$
- ϵ_c: a gaussian white noise process with variance σ_ϵ

The time series models (7) and (8) can be used to obtain forecasted values of the period index $\dot{k}_{tn+s} := \left(\dot{k}_{tn+s}^{(1)}, \cdots, \dot{k}_{tn+s}^{(N)}\right)'$. Cohort index $\dot{\gamma}_{tn+s-x_1}, s = 1, \cdots, h$, respectively to derive forecasted values of the predictor

$$\dot{\eta}_{x,tn+s} = \alpha_x + \sum_{i=1}^{N} \beta_x^{(i)} \dot{k}_{tn+s}^{(i)} + \beta_x^{(0)} \dot{\gamma}_{tn+s-x}^{(i)} \qquad (9)$$

which can in turn be used to obtain forecasted age-specific central mortality rates,

$\dot{\mu}_{x,tn+s}$

$\dot{q}_{x,tn+s}$ age-specific one-year mortality rates.

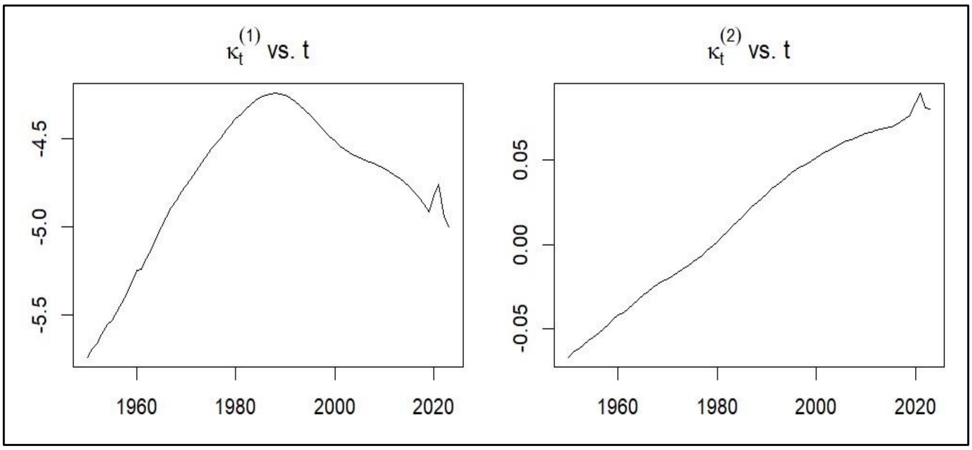
4.2. Forecasting Egyptian mortality rates

The data used is the Egyptian mortality rates for both males and females from 1950-2023, for ages $(0-84)$. The mortality data are from the United Nations https://population.un.org/wpp/Download/Standard/Mortality/.

In this research, age group 0 years old will be called x_1, 1-4 years old with x_2, 5-9 years old with x_3, and so on. The age group used in the parameter estimation process will influence the mortality rates, since the value of $(x - \bar{x})$ has a significant impact on $(q_{x,t})$.

- **Model fitting**

StMoMo package relies on function **gnm** to estimate the parameters of GAPC mortality models. The figures (2) and (3) respectively is the result of applying the function function "**fitStMoMo**" for plotting the fitted parameters on both male and female Egyptian mortality rates as follow:



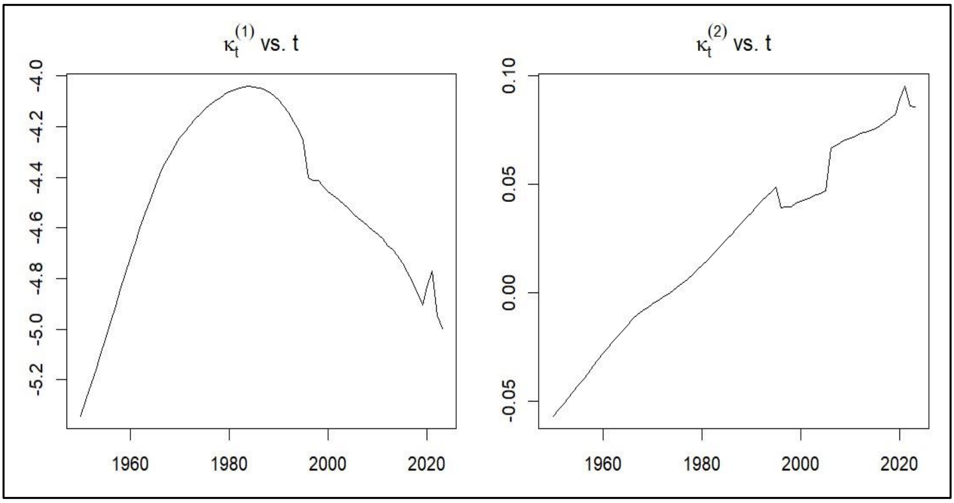
Source: R program output.

Figure (2)

CBD model parameters fitted to the Egyptian male data

This figure represents two time series for $k_t^{(1)}$ and $k_t^{(2)}$ respectively. *For the* $\boldsymbol{k_t^{(1)}}$ ***vs t*** *graph* the value of $k_t^{(1)}$ initially increase until around the early 1970, there is a peak around the late 1980, from that point the value begins to decline with a noticeable decrease until around 2020, There's a slight uptick towards the end of the series, suggesting a potential change in trend.

But ***for*** $\boldsymbol{k_t^{(2)}}$ ***vs t*** *graph* the series starts near -0-.05 and shows a consistent upward trend, there is a smooth rise over the period, particularly accelerating after 2000, with a slight peak just before 2020. This graph shows a more straightforward growth pattern, indicating steady increase in $k_t^{(2)}$ over time, with only a minor deviation near the end. The parameters of the CBD model adapted to the female population are shown in the following figure:



source: R program output.

Figure (3)

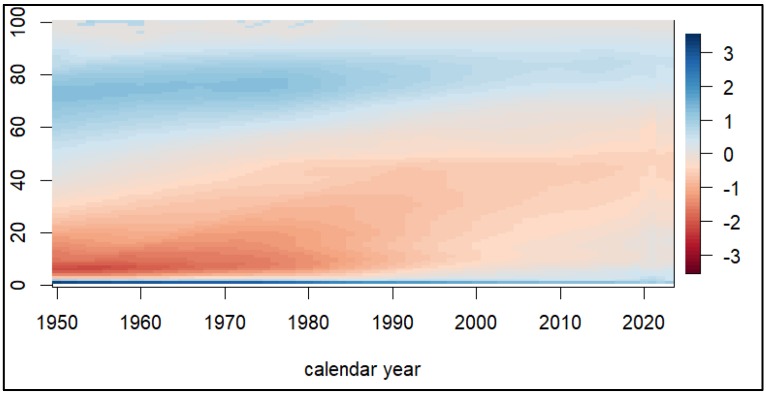
CBD model parameters fitted to the Egyptian female data

As previous figure, this figure represents two time series. From the graph $\boldsymbol{k_t^{(1)}}$ $\boldsymbol{vs\ t}$ starting around -5.2 in early 1960s, $k_t^{(1)}$ follows an upward trend until approximately 1990, reaching its peak close to -4.0. after 1990 noticeable decline until around 2020, when a slight upward change occurs, and this indicates that $k_t^{(1)}$ follows a cyclical or inverted U- shaped pattern, rising sharply from 1960 to 1990 and declining from 1990 onward.

Since, for the graph $\boldsymbol{k_t^{(2)}}$ $\boldsymbol{vs\ t}$ the series starts slightly below zero (around -0.05) in the 1960s, increases gradually until the 1980s, and shows accelerated growth into the 2000s, after 2000 the slope becomes steeper, there is a notable jump around 2010 followed by continued growth and slight dip near 2020, the general upward trend indicates a system that is improving or increasing, with slight irregularities in the later stages after 2010.

- **Goodness of fit**

Apply **residuals** function to a fitted stochastic mortality model of the class "fitStMoMo" to get standardised deviance residuals. Figures (4) and (5) respectively shows the colour map (heat map) of the deviance residuals for male & female.



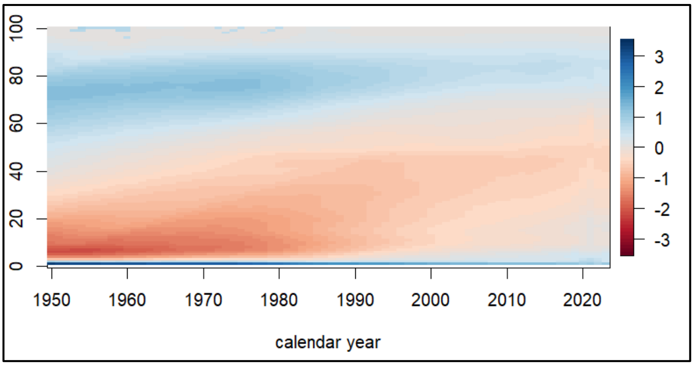
source: R program output.

Figure (4)

Heat map of deviance residuals for Male data

This heatmap shows deviance residuals over time, where the x-axis represents calendar years (from 1950 to 2020) and the y-axis likely represents age. There is a clear shift in residuals over the period, for (1950-1980) there is a negative residual at ages (0-20). From (1980- 2020) the residuals at younger ages start to turn positive indicating a change in pattern.

But for age specific patterns, for ages 20 and below there is a strong negative trend which converts to a positive trend around the year 2000. For ages above 40 the residuals are generally positive over the entire period, although the intensity of this positivity decreases slightly over time.



source: R program output.

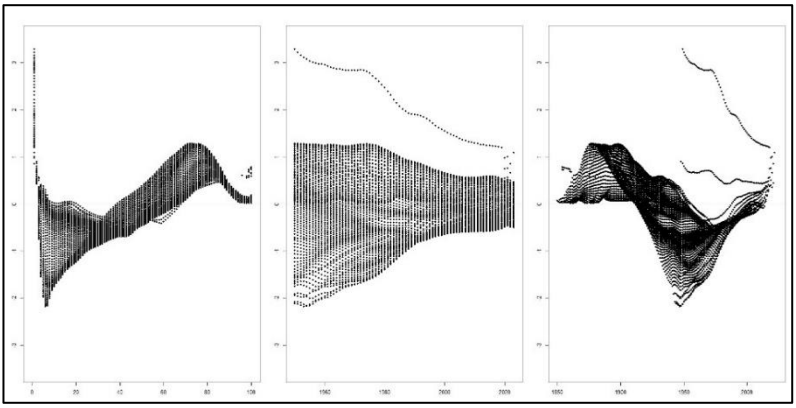
Figure (5)

Heat map of deviance residuals for female data

Like the previous heatmap, this heatmap shows deviance residuals. For the period (1950-1980) there is a concentration of negative residuals particularly at younger ages (0-40). This indicates that the model or data trend was overestimating the actual values (if residuals are observed errors). For period (1980-2000) residuals begin shift, especially for younger age groups (below 20). The negative residuals weaken, and positive residuals (blue) start emerging, especially at older ages (above

40). For period (2000-2020) Positive residuals become more dominant across various age groups, especially between ages 40 and 80. Younger ages also start showing positive residuals, suggesting underestimation by the model or changes in the underlying trend.

But for age specific pattern ages (0-20) there's a strong negative residual trend, which indicates overestimation. However, over time, this trend shifts to more neutral or slightly positive, showing a possible correction or adaptation of the model to new data. Ages (40-80) Positive residuals dominate this range, particularly in the later years. This could suggest a consistent underestimation of the actual values by the model for these age groups. For ages (80 and above) The residuals tend to be more neutral to slightly positive, indicating better model fit or a balance between overestimation and underestimation. Figures (6) and figure (7) respectively reveals a few age-related patterns for both male and female that show the absence of a quadratic age term in the CBD, which may be required to represent the death rates' frequently observed curvature in a logit scale.



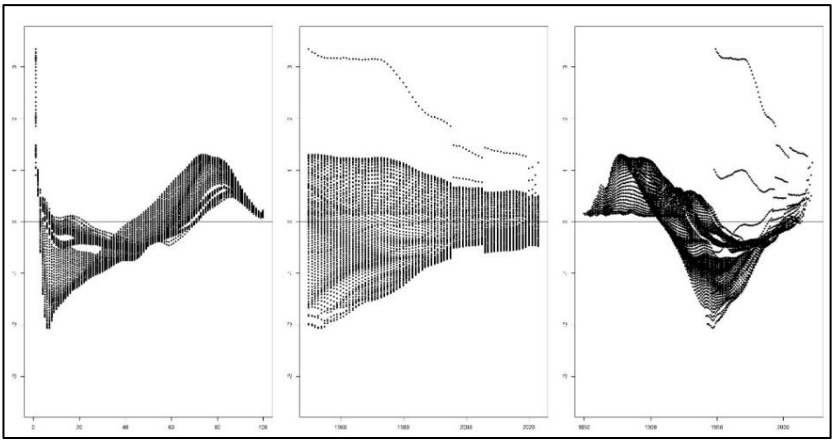
source: R program output.

Figure (6)

Scatter plots of deviance residuals for CBD fitted to the Egyptian males

- *Left panel: residuals vs. age* : There is a high spread of residuals at younger ages indicating that the model has more difficulty predicting mortality at younger ages. As age increases, the residuals appear to cluster more tightly around 0, meaning the model fits better for older ages. The residuals are larger and more dispersed at younger ages, where mortality rates are generally harder to predict with the CBD model. At older ages, the residuals become smaller and more centered around zero, indicating a better fit.
- *Middle Panel: Residuals vs. Time:* The model fits better in certain historical periods (mid-20th century), while it struggles more in earlier and more recent periods, where mortality trends are likely more volatile or subject to major demographic changes.

- *Right panel Residuals across age and time:* This combined view highlights that the CBD model performs best for older ages in mid-20th century data. However, it struggles at younger ages and in more recent years, where mortality dynamics may be influenced by factors not captured by the model.



source: R program output.

Figure (7)

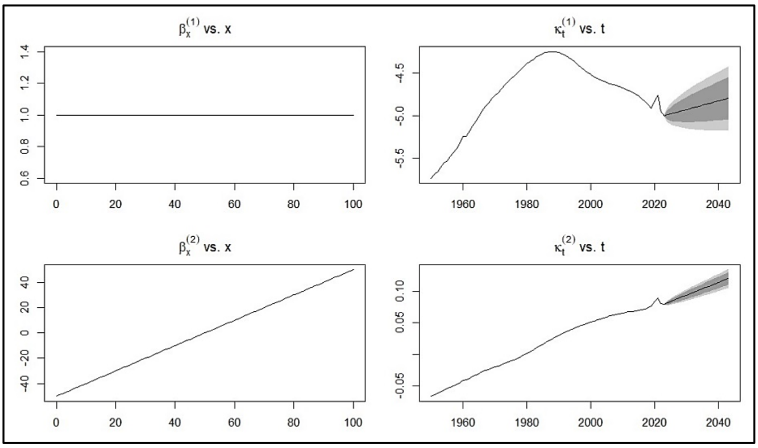
Scatter plots of deviance residuals for CBD fitted to the Egyptian female

- *Left panel residuals vs. age:* The residuals show higher dispersion at younger ages, particularly below 20 years. So we can say that the model fits best between ages 40 and 90, while it struggles with predictions at younger ages and very old ages.
- *Middle Panel Residuals vs. Time:* The model performs well for the period from 1960 to 2000 but has difficulty fitting earlier periods and more recent years (post-2000).
- *Right panel Residuals across age and time:* This plot highlights that the CBD model is effective for predicting mortality rates for middle-aged to older adults during the

mid-20th century, but it struggles at younger ages, very old ages, and in recent years due to possibly unmodeled shifts in mortality trends.

- **Forecasting and simulation with CBD model**

This section about result of forecasting mortality rate for both males and females for ages (0-100) for time period (2024-2043). In the package **StMoMo** the forecasting of CBD models is implemented via the generic method **forecast**. This function estimates and forecasts the multivariate random walk with drift in equation (7). And uses function **Arima** of package **forecast** to estimate and forecast the ARIMA process for equation (8). Figure (8) represent forecast of male mortality rates as follow:



source: R program output.

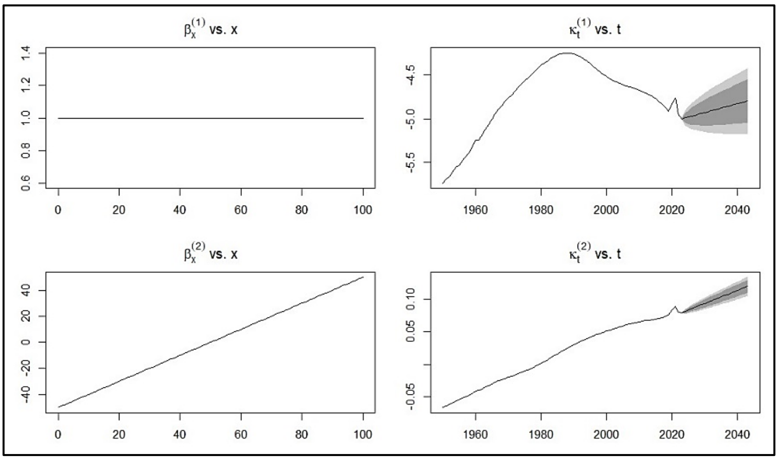
Figure (8)

Forecast male mortality rates for period (2024-2043)

Explanation of this graph on 4 stages:

- *Graph of* $\beta_x^{(1)}$ *vs. x:* $\beta_x^{(1)}$ is constant across all ages x, and the value remains around 1 which indicating that the parameter not changing with age, also that the impact of $k_t^{(1)}$ on mortality rates is uniform during different ages.
- *Graph of* $k_t^{(1)}$ *vs. t:* shows that past behavior is similar to earlier, and there is a rise up until the 1990 s followed by a decline until 2020. Regarding forecasting, the shaded region represents the confidence intervals, suggesting the uncertainty in the future values of $k_t^{(1)}$, where the central trend indicates a slight continued decline moving into the future, with wider confidence intervals as time progresses (indicating greater uncertainty).
- *Graph of* $\beta_x^{(2)}$ *vs. x:* $\beta_x^{(2)}$ increase linearly with age x, indicating that the sensitivity to $k_t^{(2)}$ increases as age increases. This might reflect a model where improvements in mortality rates are more pronounced at older ages. This aligns with longevity improvements where the advantages of developments such as medical advancements become apparent particularly strongly by elder populations.
- *Graph of* $k_t^{(2)}$ *vs. t:* shows consistent growth in $k_t^{(2)}$ with noticeable jumps in recent decades. Although it continues to grow at a smaller rate in the future, the primary tendency will remain present. Where the shaded regions indicate uncertainty, but not are the same for $k_t^{(1)}$, which indicates that the forecast for $k_t^{(2)}$ is more stable.

The next figure (9) represents the result of CBD female forecasting as follow:



source: R program output.

Figure (9)

Forecast female mortality rates for period (2024-2043)

This figure can explain as 4 graphs as figure (8):

- *Graph of* $\beta_x^{(1)}$ *vs. x:* like male graph $\beta_x^{(1)}$ remains constant across the range of ages from 0 to 100. The value is approximately 1, showing that mortality rates for all age groups have a uniform sensitivity to changes in $k_t^{(1)}$, that implies

  varies with time and has a constant impact on death rates in all age groups
- *Graph of* $k_t^{(1)}$ *vs. t*: The trajectory of $k_t^{(1)}$ peaks in the 1990s and declines steadily into the 2020s. since forecasting the shaded area represents the confidence intervals for the forecast. And the central line suggests a

continued decline, but there's notable uncertainty post-2020, as reflected in the broadening confidence intervals. Compared to male data $k_t^{(1)}$ seems somewhat smoother here, and the range of possible outcomes is similar in scale.

- *Graph of $\beta_x^{(2)}$ vs. x:* $\beta_x^{(2)}$ shows a linear increase as age x increases, consistent with the prior graph. This indicates that older individuals are more sensitive to changes in $k_t^{(2)}$ than younger individuals.
- *Graph of $k_t^{(2)}$ vs. t:* shows consistent growth in $k_t^{(2)}$ with noticeable jumps in recent decades. The future forecasting shows a continuation of this upward trend, with the confidence intervals widening as we move further into the future. Although the forecast of $k_t^{(2)}$ suggests ongoing improvements though with a slight moderation in the rate of increase compared to earlier years.

5. Overall summary

This section provide summary for all above stages to apply CBD model :

- **First: heat map**

Male:

 This could indicate changes in a model's fit over time, with better fits (closer to zero residuals) occurring more recently for certain age groups.

 Alternatively, it might suggest structural changes in the data being modeled, such as demographic shifts, improvements in healthcare, or other societal factors influencing the underlying trends.

Female

- *Model Performance*: The model's accuracy seems to have improved over time, particularly for younger ages, but still struggles with certain age groups, especially middle-aged and older individuals.
- *Trend Shifts:* The heatmap suggests that there may have been significant demographic or societal changes over time, leading to shifts in the model's residuals. These could be due to changes in mortality rates, healthcare improvements, lifestyle changes, or other factors influencing the population.

- **Second: Scatter plot**

Male:

- The CBD model is well-suited to older populations, as evidenced by the smaller residuals for those age groups.
- Younger ages and certain historical periods (especially the distant past and recent years) show more variability in residuals, indicating that the model doesn't capture those trends as well.
- To improve the model, you may need to explore additional factors or modifications, especially for younger ages and more recent mortality trends where significant changes in life expectancy, healthcare advances, or population dynamics are at play.

Female:

- The residual plots suggest that the CBD model fits well for mid-range ages (40-90) and for data from the mid-20th century (1960-2000).
- The model struggles with predicting mortality for younger individuals, the very old, and periods of rapid change (both in earlier historical data and post-2000).
- The residuals post-2000 could indicate that mortality trends are evolving in ways not captured by the traditional CBD model, potentially signaling the need for adjustments or additional factors to account for recent improvements in healthcare and longevity.

- **<u>Third: Forecasting</u>**

Male:

- Uniform sensitivity in $\beta_x^{(1)}$ Mortality rates across all ages are equally sensitive to changes in $k_t^{(1)}$
- Increasing sensitivity in $\beta_x^{(2)}$ Mortality increasing are more pronounced at older ages as shown by $\beta_x^{(2)}$.
- Forecasted decline in $k_t^{(1)}$ Although the past shows a peak and subsequent decline, the forecast predicts further decline, but with substantial uncertainty.
- Continued increase in $k_t^{(2)}$ the future trend for $k_t^{(2)}$ continues to show improvements, with a smaller range of uncertainty compared to $k_t^{(1)}$.

Female:

- No more changes from male forecasting, Uniform sensitivity in $\beta_x^{(1)}$ Mortality rates across all ages are equally sensitive to changes in $k_t^{(1)}$.
- Increasing sensitivity in $\beta_x^{(2)}$ older populations continue to show a greater sensitivity to changes in $k_t^{(2)}$.
- Forecasted decline in $k_t^{(1)}$ The continuing decline shows that there is significant uncertainty around the predicted slowing from the underlying component reflected by this parameter after 2020.
- Continued increase in $k_t^{(2)}$ the future trend for $k_t^{(2)}$ continues to show improvements, with a smaller range of uncertainty compared to $k_t^{(1)}$.

6. Conclusion

The CBD model is useful for specific age ranges (elderly ages) and time periods, but its limitations at the age extremes and in more recent years highlight areas where improvements or alternative modeling approaches may be need for better mortality forecasting.

Also, the model shows a consistent model performance and changes in data fit over time. This reflects the continuous modifications aimed at increasing accuracy, specifically across different age groups, indicating of demographic or health-related transformations over the decades.

7. Recommendation

1. Expanding CBD model to modeling age specific factors to capture younger age behavior accurately.

2. Improving model by adding exogenous variables such as: socioeconomic factors which could help in explain mortality for specific periods or unusual residual patterns.

References

- Andres, V., Millossovich, P., & Vladimir, K. (2018). StMoMo: Stochastic Mortality Modeling in R. *Journal of Statistical Software*, *84*(3), 1-38
- Apicella, G., Dacorogna, M., Di Lorenzo, E., & Sibillo, M. (2019). Improving the Forecast of Longevity by Combining Models*, North American Actuarial Journal*, 23(2), 298-319
- Chan, W.-S., Li, J. S.-H., & Li, J. (2014). The CBD Mortality Indexes: Modeling and Applications. *North American Actuarial Journal*, *18*(1), 38–58.
- Janssen, F. (2018). Advances in Mortality Forecasting: Introduction. *Genus* 74(1),21.
- Li, H., & O'Hare, C. (2017). Semi-parametric extensions of the Cairns–Blake–Dowd model: a one-dimensional kernel smoothing approach. *Insurance: Mathematics and Economics*, 77(3), 166–176.
- Maccheroni, C., & Nocito, S. (2017). Backtesting the Lee–Carter and the Cairns–Blake–Dowd Stochastic Mortality Models on Italian Death Rates. *Risks*, *5*(3), 34.

- Mitchell, D., Brockett, P., Mendoza-Arriaga, R., & Muthuraman, K. (2013). Modeling and forecasting mortality rates. *Insurance: Mathematics and Economics,* 52(2), 275–285.
- Odhiambo, J., Weke, P., & Ngare, P. (2021). A Deep learning integrated cairns-blake-dowd (CBD) sytematic mortality risk model. *Journal of Risk and Financial Management*, *14*(6), 259
- Redzwan, N., & Ramli, R. (2022). A Bibliometric Analysis of Research on Stochastic Mortality Modelling and Forecasting. *Risks*, *10*(10), 191.
- Safitri, Y. R., Mardiyati, S., & Malik, M. (2019). The Cairns-Blake-Dowd model to forecast Indonesian mortality rates. In *AIP Conference Proceedings* (Vol. 2168, No. 1). AIP Publishing.
- Sweeting, P. J. (2011). A trend-change extension of the Cairns-Blake-Dowd model. *Annals of Actuarial Science*, *5*(2), 143-162.
- United Nations. Retrieved from https://population.un.org/wpp/Download/Standard/Mortality/
- World bank group. Retrieved from https://data.worldbank.org/indicator/SP.DYN.LE00.FE.IN?locations=EG

- Zulkifle, H., Yusof, F., & Nor, S. R. M. (2019). Comparison of Lee Carter Model and Cairns, Blake and Dowd Model in Forecasting Malaysian Higher Age Mortality. *Matematika*, 65–77.