The Performance of The Lee-Carter Model on Heterogeneous Adult Mortality Data in a Limited Data Situation

Angela Chukwu^{1*} & Oluwaseun Adegbilero-Iwari^{1, 2}
1. Department of Statistics, University of Ibadan, Ibadan, Nigeria
2. College of Medicine and Health Sciences, Afe-Babalola University, Ado-Ekiti, Nigeria;
*E-mail of the corresponding author: seuneniola01@gmail.com

Abstract

Adult mortality in developing Countries remains one of the greatest challenges for monitoring, analyzing and projecting the health situation of a large proportion of the world's population. Because of limited resources, many developing Countries, especially in sub-Saharan Africa lack vital registration systems that could reliably and continuously collect information on adult mortality. However, periodic reviews of mortality rate and pattern using mortality models could show new trends and may provide information for the planning of a Country's health care delivery services and a host of other programs.

Previous studies have found that the Lee-Carter model works well with homogenous mortality data. In this current study, we want to see the performance of the model on heterogenous adult mortality data in a limited data situation. A modified version of the Lee-Carter method is used to model sex-combined adult mortality data of Nigerians aged 15-84 years for the time periods 1990, 2000 and 2012. The model's parameters are estimated using the approach proposed by Lee and Carter (1992) based on the singular value decomposition technique, while the mortality index is predicted using the approach developed by Li *et al.* (2002). Our results show that on the overall, the model follows the mortality pattern very well for most of the ages despite the heterogonous nature of the data used. Forecast values of the mortality index show a gradual decline in mortality from 2013-2025 in Nigeria.

KEYWORDS: Adult mortality; Age; Lee-Carter model; Nigeria.

1. Introduction

Whether it is the *eye* or the *toe*, every part of the human body is designed to play a particular function or the other. Oftentimes, we do not realize the importance of the '*less significant part*' until it brings a form of discomfort we can't bear. The workforce of any nation forms an integral part of the adult population. Therefore, the latter should not be overlooked or treated as though they do not '*really*' matter. For several years, emphasis has been placed on infant and child mortality while adult mortality has been given little prominence especially in a developing country like Nigeria.

The level of adult mortality varies markedly across the world, both between countries and between regions and social groups within particular countries. At the global level, adult mortality is said to have reduced especially in low income countries (WHS 2014). However, estimates of life expectancy show that Nigeria is one of the nine Countries in Africa with life expectancy below 55 years. Adult mortality is higher in Nigeria compared to her neighbors such as Benin, Ghana and Cameroon (WHS 2014).

With a wide availability of mortality history data, the Lee-carter model has been applied and has given successful results for various Countries. For instance the U.S. (Lee and Carter 1992); (Preston and Wang 2009), Canada: (Lee and Nault 1993); Li and Chan (2007), Chile (Lee and Rofman 1994), Japan (Wilmoth 1998), Taiwan (Wang & Liu 2010), Nordic countries (Koissi, Shapiro & Högnäs 2004), Belgium (Brouhns *et al.* 2002), Australia (Booth et al. 2002), the U.K. (Renshaw and Haberman 2003), Sweden: (Lundström & JanQvist 2004); Wang (2007). B & and Scherp (2007) and Chukwu & Oladipupo (2012) discovered that the Lee-carter model works well on homogenous mortality data.

1.1 Study Objectives

The main objective of this study is to see the performance of the model on heterogenous adult mortality data. Heterogeneity in this study implies that the data used for the study is sex-combined and the age groups have varying mortality trends. Others objectives are:

- 1. To determine past time trends in the general pattern of adult mortality across the age groups
- 2. To assess the sensitivity of the hazard rate at age x to the time trend.
- 3. To forecast for future expectation of the mortality trend

2. Data and Methods

The data used for this study was extracted from the global health observatory of the World Health Organization. It covers the age-specific mortality rate records of persons aged 15 to 84 years for the time periods 1990, 2000 and 2012 in Nigeria. Methods used for analysis include the Singular value decomposition technique using the principal component analysis option. Forecast of the mortality index was done by applying the random walk with drift model. The analysis for this study was done using the *biplot* ad-in in Microsoft Excel 10.0.

2.1 The Lee-Carter Model

The Lee-Carter model is a simple bilinear model in the variables x (age) and t (calendar year) defined as; $\ln m_{xt} = \hat{a}_x + \hat{b}_x \hat{k}_t + \varepsilon_{xt}$

Where;

 m_{xt} : is the matrix of the observed age-specific death rate at age x during year t. It is obtained from observed deaths divided by population exposed to risk.

 \hat{a}_x : represents the logarithm of the geometric mean of the empirical mortality rates, averaged over the historical data. It describes the average shape of the age profile.

 \hat{k}_t : represents the underlying time trend for the general mortality. It captures the main time trend on the logarithmic scale in mortality rates at all ages. It is also referred to as the mortality index.

 \hat{b}_x : represents the sensitivity of the hazard rate at age x to the time trend. It indicates the relative pace of change in mortality by age as \hat{k}_t varies. It modifies the main time trend according to whether change at a particular age is faster or slower than the main trend.

 ε_{xt} : is the residual term at age x and time t. It reflects the age specific influences not captured by the model. Each of ε_{xt} is independent and identically distributed $\varepsilon_{xt} \sim N(0, \sigma^2)$.

2.1.1 Merits of the Model

- It produces an excellent fit to mortality trends;
- > It is parsimonious in the number of parameters used;
- > It linearizes mortality trends and thereby adds confidence to extrapolations;
- > It produces sensible estimates of forecast of uncertainty and
- > The model outperforms other models with respect to its prediction errors.

Despite its advantages over other extrapolative methods, the Lee-Carter model also shares the fundamental weaknesses of extrapolation: historical patterns may not hold for the future, and structural changes may therefore be missed.

2.1.2 Assumptions of the Model

- > The model assumes that b_x is invariant (remains constant) over time for all x.
- > It assumes that k_t is fixed over age-groups for all t.
- The practical use of the model assumes that the disturbances ε_{xt} are normally distributed i.e. $\varepsilon_{xt} \sim N(0, \sigma^2)$

2.2 Estimating the Parameters

The Lee-Carter model unlike most regression formulations is not a common model structure with observed variables on the right hand side and thus it cannot be fit with simple regression formulations. Hence, the estimation of \hat{b}_x and $\hat{\kappa}_t$ cannot be solved explicitly. The *singular value decomposition* (SVD) method was used to find a least squares solution in this study

The Singular Value Decomposition (SVD) is a technique used to decompose a matrix into several component matrices, exposing many of the useful and interesting properties of the original matrix. Ideally, the matrix is decomposed into a set of factors (often orthogonal or independent) that are optimal based on some criterion. In other words, any real mxn matrix A can be decomposed uniquely as:



$A = UDV^T$

U is *mxn* and orthogonal (its columns are eigenvectors of AA^{T}) *V* is *nxn* and orthogonal (its columns are eigenvectors of $A^{T}A$) *D* is *nxn* diagonal (non-negative real values called *singular* values) $D = diag (\rho_1, \rho_2, ..., \rho_n)$ ordered so that $\rho_1 \ge \rho_2 \ge \cdots \ge \rho_n$ (If ρ is a singular value of *A*, it's square is an eigenvalue of $A^{T}A$) The application of the SVD approach follows about 6 steps (see: *Chukwu & Oladipupo 2012*)

The purpose of using Singular value decomposition is to transfer the task of forecasting an age-specific vector $\ln m_{xt}$ into forecasting a scalar κ_t , with small error. As earlier stated, the model is given by: $\ln m_{x(t)} = \hat{a}_x + \hat{b}_x \hat{k}_{(t)} + \varepsilon_{x,(t)}$ and we need to estimate a_x , b_x and κ_t .

In order to achieve a unique solution the following restrictions are used;

$$\sum_{x\min}^{x=m} b_x^2 = 1$$
 (1)

$$\sum_{t\min}^{t=n} \kappa_t = 0 \tag{2}$$

Where x = 1.... m age groups and t = 1..... calendar years.

Where $\ln m_{xt}$ and Q are subject to the constraints in (3.1) and (3.2) above.

To find values that minimize Q we introduce Lagrange's multipliers; α and β

$$R = Q - \alpha \sum_{t} \kappa_{t} - \beta \sum_{x} b_{x}^{2} . \qquad (4)$$

We substitute equation (3) into (4) and then minimize;

$$R = \sum_{x,t} (a_x + b_x \kappa_t - q_{xt})^2 - \alpha \sum_t \kappa_t - \beta \sum_x b_x^2$$

Thus we first take the derivative of R in respect of a_x , b_x and κ_t respectively;

The derivatives are set equal to

Mathematical Theory and Modeling ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online) Vol.5, No.4, 2015

$$\frac{dR}{da_x} = 0$$
 , $\frac{dR}{db_x} = 0$ and $\frac{dR}{dk_t} = 0$

We solve equation (3.5);

$$2\sum_{t} (a_{x} + b_{x}\kappa_{t} - q_{xt}) = 0 \qquad \qquad =>$$

$$na_{x} + b_{x}\sum_{t}\kappa_{t} - \sum_{t}q_{xt} = 0 \qquad (\sum_{t \min}^{t=n}\kappa_{t} = 0) \qquad =>$$

$$na_{x} - \sum_{t}q_{xt} = 0 \qquad \qquad =>$$

$$\hat{a}_{x} = \frac{1}{n}\sum_{t}q_{xt}$$

$$\hat{a}_x = \frac{1}{n} \sum_{t=1}^n \ln m_{xt}$$

The estimate for a_x is thus computed as the average over time of the logarithm of the central death, which corresponds to the definition of a_x in section 3.0.

We now define $z_{xt} = q_{xt} - a_x \implies \sum_t z_{xt} = 0$ for every x.

Equation (3.7) can now be rewritten as:

$$2\sum_{x} b_x (a_x + b_x \kappa_t - q_{xt}) - \alpha = 2\sum_{x} b_x (b_x \kappa_t - z_{xt}) - \alpha$$

If we set the new expression of the derivative with regard to K_t equal to zero we have;

$$2\sum_{x} b_{x}(b_{x}\kappa_{t} - z_{xt}) - \alpha = 0 \qquad \Longrightarrow$$

$$\sum_{x} b_{x}(b_{x}\kappa_{t} - z_{xt}) = \frac{\alpha}{2} \qquad \Longrightarrow$$

$$\kappa_{t} \sum_{x} b_{x}^{2} - \sum_{x} b_{x} z_{xt} = \frac{\alpha}{2} \qquad (\sum_{x\min}^{x=m} b_{x}^{2} = 1)$$

$$\kappa_{t} - \sum_{x} b_{x} z_{xt} = \frac{\alpha}{2} \qquad ((\sum_{x\min}^{x=m} b_{x}^{2} = 1))$$

$$((\sum_{x\min}^{x=m} b_{x}^{2} = 1)) \qquad ((\sum_{x\min}^{x=m} b_{x}^{2} = 1))$$

$$((\sum_{x\min}^{x=m} b_{x}^{2} = 1)) \qquad ((\sum_{x\min}^{x=m} b_{x}^{2} = 1))$$

$$((\sum_{x\min}^{x=m} b_{x}^{2} = 1)) \qquad ((\sum_{x\min}^{x=m} b_{x}^{2} = 1)) \qquad ((\sum_{x\min}^{x=m} b_{x}^{2} = 1)) \qquad ((\sum_{x\min}^{x=m} b_{x}^{2} = 1))$$

$$((\sum_{x\min}^{x=m} b_{x}^{2} = 1) \qquad ((\sum_{x\min}^{x=m} b_{x}^{2} = 1)) \qquad ((\sum_{x\min}^{x=m} b_{x}^{2} = 1) \qquad ((\sum_{x\min}^{x=m} b_{x}^{2} = 1)) \qquad ((\sum_{x\min}^{x=m} b_{x}^{2} = 1)) \qquad ((\sum_{x\min}^{x=m} b_{x}^{2} = 1) \qquad ((\sum_{x\min}^{x=m} b_{x}^{2} = 1)) \qquad ((\sum_{x\min}^{x=m} b_{x}^{2} = 1) \qquad ((\sum_{x\min}^{x=m} b_{x}^{2} = 1)) \qquad ((\sum_{x\min}^{x=m} b_{x}^{2} = 1) \qquad ((\sum_{x\min}^{x=m} b_{x}^{2} = 1)) \qquad ((\sum_{x\min}^{x=m} b_{x}^{2} = 1) \qquad ((\sum_{x\max}^{x=m} b_{x}^{2} = 1) \qquad ((\sum_{x\max}^{x=m} b_{x}^{2} = 1) \qquad ((\sum_{x\max}^{x=m}$$

Taking the sum over t in the equation (8) above we get,

$$\sum_{t} (\kappa_{t} - \sum_{x} b_{x} z_{xt}) = \sum_{t} \frac{\alpha}{2} = \sum_{t} \kappa_{t} - \sum_{t} \sum_{x} b_{x} z_{xt} = \sum_{t} \frac{\alpha}{2} = \sum_{t} \frac{\alpha}{2} = \sum_{t} \kappa_{t} = 0 = \sum_{t} \sum_{x} b_{x} z_{xt} = \sum_{t} \frac{\alpha}{2} = \sum_{t} \sum_{x} b_{x} z_{xt} = \sum_{t} \sum_{x} b_{x} z_{$$

$$m\sum_{x} b_{x} \sum_{t} z_{xt} = \frac{m\alpha}{2} \qquad (\sum_{t} z_{xt} = 0) \implies \alpha = 0$$

We now have an expression for κ_t by putting $\alpha = 0$ in equation (3.8).

Hence,
$$\kappa_t = \sum_x b_x z_{xt}$$
 for every t(9)

The constraint for K_t is now fulfilled.

If we substitute the expression $z_{xt} = q_{xt} - a_x$ into equation (3.6), we obtain;

$$2\sum_{t}\kappa(a_{x}+b_{x}\kappa_{t}-q_{xt})-2\beta b_{x}=2\sum_{t}\kappa_{t}(\kappa_{t}b_{x}-z_{xt})-2\beta b_{x}$$

Setting this equal to zero we have,

We take the square of both sides and summarise over x and get

We are able to get an expression for b_x by inserting the equation for β in (11) into equation (10):

$$b_{x}\left(\sum_{t}\kappa_{t}^{2}-\beta\right)=\sum_{t}\kappa_{t}z_{xt}$$

$$(\beta=\sum_{t}\kappa_{t}^{2}-\sqrt{\sum_{x}\left(\sum_{t}\kappa_{t}z_{xt}\right)^{2}}) \qquad =>$$

$$b_{x}\left(\sum_{t}\kappa_{t}^{2}-\sum_{t}\kappa_{t}^{2}+\sqrt{\sum_{x}\left(\sum_{t}\kappa_{t}z_{xt}\right)^{2}}\right)=\sum_{t}\kappa_{t}z_{xt} \qquad =>$$

$$\hat{b}_{x} = \frac{\sum_{t} \kappa_{t} z_{xt}}{\sqrt{\sum_{x} (\sum_{t} \kappa_{t} z_{xt})^{2}}} \quad \forall x.$$
Hence, $\hat{a}_{x} = \frac{1}{n} \sum_{t=1}^{n} \ln m_{xt}$, $\hat{b}_{x} = \frac{\sum_{t} \kappa_{t} z_{xt}}{\sqrt{\sum_{x} (\sum_{t} \kappa_{t} z_{xt})^{2}}} \quad \hat{\kappa}_{t} = \sum_{x} b_{x} z_{xt}$

2.3 Forecasting $\hat{k}_{u(t)}$ Using Data at Unequal Intervals

After fitting the mortality data to the model and the values of the vectors \hat{a}_x , \hat{b}_x and $\hat{k}_{u(t)}$ are found, only the mortality index $\hat{k}_{u(t)}$ needs to be predicted. Suppose mortality data is collected at times u(0), u(1)... u(T). The random walk with drift equation becomes

$$\hat{k}_{u(t)} = \hat{k}_{u(t-1)} + \hat{\theta}[u(t) - u(t-1)] + [(\varepsilon_{u(t-1)} + 1) + ... + \varepsilon_{u(t)}]$$

So $\hat{k}_{u(t)} - \hat{k}_{u(t-1)} = \hat{\theta}[u(t) - u(t-1)] + [(\varepsilon_{u(t-1)} + 1) + ... + \varepsilon_{u(t)}]$(12)

Because the means of the second term in the right-hand side of equation (12) are zero, the unbiased estimate of $\hat{\theta}$ is obtained as:

$$\hat{\theta} = \frac{\sum_{t=1}^{T} \hat{k}_{u(t)} - \hat{k}_{u(t-1)}}{\sum_{t=1}^{T} [u(t) - u(t-1)]} = \frac{\hat{k}_{u(T)} - \hat{k}_{u(0)}}{u(T) - u(0)}$$

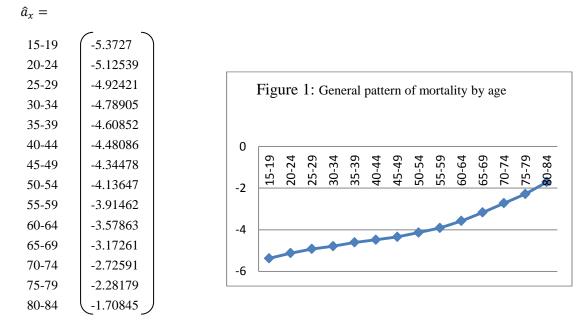
$$\operatorname{Var}\left(\varepsilon_{u(t)}\right) = \frac{\sum_{t=1}^{T} \left[\hat{k}_{u(t)} - \hat{k}_{u(t-1)} - \hat{\theta}[u(t) - u(t-1)]\right]^{2}}{[u(T) - u(0) - \frac{\sum_{t=1}^{T} [u(t) - u(t-1)]^{2}}{u(T) - u(0)}]}$$

$$\operatorname{Var}\left(\hat{\theta}\right) = \frac{var\{\sum_{t=1}^{T} [(\varepsilon_{u(t-1)} + 1) + \dots + \varepsilon_{u(t)}]]}{[u(T) - u(0)]^{2}}$$

$$= \frac{\sigma^{2}}{u(T) - u(0)} \approx \frac{\operatorname{Var}\left(\varepsilon_{u(t)}\right)}{u(T) - u(0)}$$

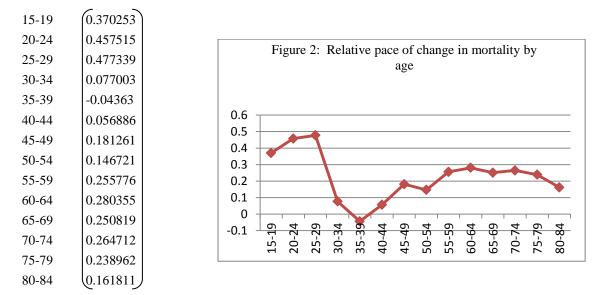
3. Summary of Results

We present below a summary of our findings.



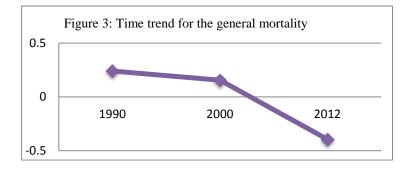
Parameter \hat{a}_x represents the general pattern of mortality by age. Our results show that \hat{a}_x values are increasing with age. This implies an upward trend in mortality with respect to the age-groups. As shown in figure 1 above, the younger ages have a lower mortality rate than the older ages. The reason for this disparity is probably due to differences in exposure to risk such as family responsibilities, stress, health issues so on.

 $\hat{b}_x =$

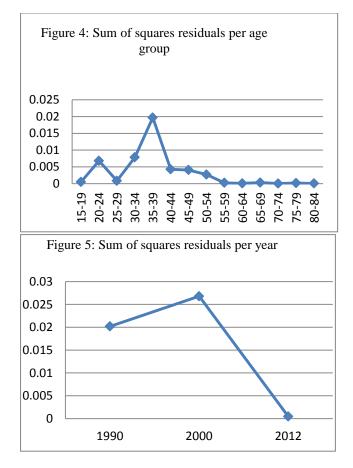


Parameter \hat{b}_x describes the tendency of mortality at age x to change as the general level of mortality changes. The larger the value of \hat{b}_x at a particular age-group, the more fluctuant the mortality rate at that age-group as compared to the general level of mortality change. The results obtained show that persons aged 20-29 have a more fluctuant mortality pattern than other age groups with age group 25-29 having the highest fluctuant mortality pattern. The relative pace of change is lowest in age group 35-39 as illustrated in figure 2 above.

1990	2000	2012
[0.23988	0.15597	-0.39586]



Parameter $\hat{k}_{u(t)}$ is the mortality index and it captures the main time trend on the logarithmic scale in death rates at all ages. When $\hat{k}_{u(t)}$ values decrease overtime, it signifies a decline in mortality trend as the years increase. Our findings show a downward trend from 1990 to 2012 as shown in figure 3. However, it is important to note that the straight line obtained in figure 3 is as a result of the 3-time point nature of our data. It does not necessarily imply that the mortality pattern between 1990-2000 and 2000-2009 was not a fluctuating one. Furthermore, the significant decline in mortality in between the years can be attributed to social-economic factors such as improved Government policies on health, for instance the National health insurance scheme (NHIS).



Sum of squares residuals per age group & per year: The residuals indicate the age specific influences not captured by the model. As shown in figure 6 above, residuals were lowest at 2012.

4. Conclusion

The goodness of fit as shown in figure 7 of the appendix shows that the Lee-Carter model follows the mortality pattern very well for most of the ages despite the heterogenous nature of the data. On the overall, an R^2 of 0.8341 was obtained as the goodness of fit of the model. The forecast of the mortality time trend show a gradual decline in mortality from 2013-2025 as expressed in figure 8 while the general pattern of mortality shows an upward trend from age 15-84 years all things being equal. We can therefore say that the lee-carter model can be used to model the Nigeria mortality data even when the data available is heterogenous in nature. However, the level of data heterogeneity for which the lee-carter model will give successful result is another issue for discussion.

References

B âte, D. & Scherp, H. (2007). Applying the Lee-Carter model to countries in Eastern Europe and the former Soviet Union. *Dissertation for Actuarial Diploma, Swedish Society of Actuaries*.

Brouhns, N., Denuit, M. and Vermunt, J. (2002). A Poisson log-bilinear regression approach to the construction of projected life tables. *Insurance: Mathematics and Economics*, 31, 373–393.

Chukwu, A.U. & Oladipupo, E.O. (2012). Modelling Adult Mortality in Nigeria: An Analysis Based on the Lee-Carter Model. *Studies in Mathematical Sciences*, 5 (2), 1-11. Available at http://www.cscanada.net/index php/sms /article/view/j.sms.1923845220120502.1986 DOI: 10.3968/j.sms.1923845220120502.1986

Koissi, M., Shapiro, A. & Högnäs, (2004). Fitting and forecasting mortality rates for Nordic countries using the Lee-Carter model, *Proceeding of the 39th Actuarial Research Conference, ARCH*

Lee, R. & Carter, L. (1992). Modeling and Forecasting the Time Series of U.S. Mortality. *Journal of the American Statistical Association*, 87: 65-71.

Lee, R. & Nault F., (1993). Modelling and forecasting provincial mortality in Canada. paper presented at the *World Congress of the International Union for Scientific Study of Population*, Montreal, 1993.

Lee, R. & Rofman, R. (1994) "Modelacion Proyeccion de la Mortalidad en Chile. NOTAS 22 (59), 182-213.

Li, N., Lee, R. & Tuljapurkar, S. (2002). Using the Lee-Carter method to forecast mortality in populations with limited data. *International Statistical Review*, 72 (1), 19-36.

Li, S. & Chan, W. (2007). The Lee-Carter Model for forecasting mortality, revisited. *North American actuarial journal*, 11(1), 68-87. DOI:10.1080/10920277.2007.10597438

Lundström, H. & Qvist, J. (2004). Mortality Forecasting and Trend Shifts: an Application of the Lee-Carter Model to Swedish Mortality Data. *International Statistical Review*, (72)1, 37-50.

Preston, S. & Wang, H. (2009). Forecasting United States mortality using cohort smoking histories. *PNAS*, 106 (2), 393-398. Available at www.pnas.org/cgi/doi/10.1073/pnas.0811809106

Renshaw, A. & Haberman, S. (2003). Lee-Carter mortality forecasting: a parallel generalized linear modelling approach for England and Wales mortality projections. *Journal of the Royal Statistical Society (Applied Statistics)*, 52 (1), 119-137.

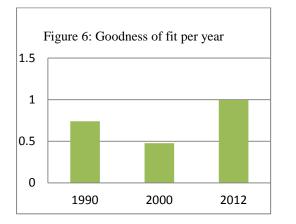
Wang, C.W. & Liu, Y.L. (2010). Comparisons of mortality modelling and forecasting- empirical evidence from Taiwan. *International Research Journal of Finance and Economics*, ISSN 1450-2887 issue 37. Available at http://www.eurojournals.com/finance.htmWang, J.Z. (2007). Fitting and forecasting mortality for Sweden: applying the Lee-Carter model. Retrieved from http://www2.math.su.se/matstat/reports/serieb/2007/rep1/report.pdf

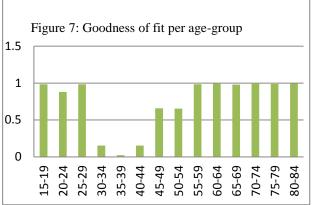
Wilmoth, J. R., (1998). Is the pace of Japanese mortality decline converging toward international trends? *Population and Development Review* 24(3), 593-600. Retrieved from http://demog.berkeley.edu/~jrw/Eprints/wilmoth.1998.MortFore1a.pd

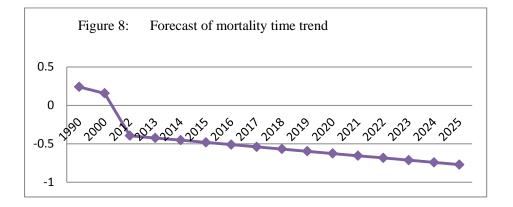
World Health Statistics Report (2014).Retrieved from

http://apps.who.int/iris/bitstream/10665/112738/1/9789240692671_eng.pdf?ua=1.

Appendix







The IISTE is a pioneer in the Open-Access hosting service and academic event management. The aim of the firm is Accelerating Global Knowledge Sharing.

More information about the firm can be found on the homepage: <u>http://www.iiste.org</u>

CALL FOR JOURNAL PAPERS

There are more than 30 peer-reviewed academic journals hosted under the hosting platform.

Prospective authors of journals can find the submission instruction on the following page: <u>http://www.iiste.org/journals/</u> All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Paper version of the journals is also available upon request of readers and authors.

MORE RESOURCES

Book publication information: http://www.iiste.org/book/

Academic conference: http://www.iiste.org/conference/upcoming-conferences-call-for-paper/

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digtial Library, NewJour, Google Scholar

