Journal of Statistical and Econometric Methods, vol.6, no.2, 2017, 43-68
ISSN: 1792-6602 (print), 1792-6939 (online)
Scienpress Ltd, 2017

Projecting the long run relationship of multi-population life expectancy by race

A. Ntamjokouen[1], S. Haberman[2] and G. Consigli[3]

Abstract

Census data demonstrate that life expectancy has improved over the last 100 years. In this paper, we attempt to predict life expectancy in the USA by both race(black people, white people, and the general population) and gender. In doing this, we employ methods of cointegration analysis that have appeared recently in the actuarial and demography literature. We investigate the dependence(through cointegration analysis) between the six variables and it shows a better fit with better performance than others models such as VAR and ARIMA in predicting life expectancy at birth by race. We show that there are similar long-term trends in the average life expectancy of the members of all the main racial groups residing in the USA. Our study offers new insights to demographers with regard to predicting the average future life expectancy of members of different racial groups.

Mathematics Subject Classification: J11; J14; J15; C01; C52; C53

[1] Department of Management Science, Economics and Quantitative Methods, Università degli Studi di Bergamo, Italy. E-mail: achille.ntamjokouen@unibg.it

[2] Cass Business School, City University of London, UK. E-mail: s.haberman@city.ac.uk

[3] Università degli Studi di Bergamo, Italy. E-mail: giorgio.consigli@unibg.it

Article Info: *Received* : November 3, 2016. *Revised* : December 5, 2016.
Published online : April 1, 2017

Keywords: Life expectancy at birth; race; forecasting; VECM; ARIMA.

1 Introduction

Life expectancy and mortality rates are improving in developed countries, as several studies such as Tulgapurkar et. al(2007) and Oeppen(2002) have shown. The USA is not an exception: there has been improved life expectancy of the population in general and of the different racial groups living there. To illustrate this, in 1970 average life expectancy was 68 years for white males and 60 years for black males. By 2010, these figures had risen to 76.5 and 71.8 years, respectively. The statistics for females are similar: in 1970 average life expectancy was 75.6 years for white females and 68.3 years for black females; by 2010 these figures had risen to 81.3 and 78 years respectively. Historically, population censuses in the USA population mainly gathered data on two racial groups: white and black people. In recent years, however, many other groups racial groups such as Latin Americans and Asians - have immigrated to the USA and their presence is becoming visible there. Most of the literature on this topic - For example, Rives 1977; NCHS (1975), Manton(1980, 1982), Manton et. al(1979), Philipps and Burch(1960), Woodbury et. al(1981), Manton et. al(1979) and Carter(2010) - has focused on predicting the pair black-white death rates and life expectancy. In this study, we focus not only on the average life expectancy of the members of these two racial groups but also of common trends to predict the average life expectancy of each *racial/ethnic* group and of the general population. Our study also takes account of common trends to predict the average life expectancy from each group and also from all races in general.

In 1610, all the inhabitants of the USA were white according to official statistics we found; and 10 years later 0.9% all of newcomers were black (see Table 1 retrieved from U.S Bureau Census 1975). In 1710 86.5% of the population were white and the rest were black(see Humes(2009)). From 1850, data on the Hispanic population was included in the USA census statistics(In 1850 they accounted for 0.50% of the total population). By 1910, the official census data showed that Asian accounted for 0.2% of the total population and Indians 0.1%. In 1970, people of Hispanic origins made up 0.1% of the

American population. In 2010, there were four main(or ethnic groups) in the USA: Asian, Hispanic, white and black people. For four centuries the latter two racial groups have comprised the majority of the USA's population (as Table 1 and 2).

However, black people are no longer the second largest group: The number of black people has recently been exceeded by the number of hispanic (see Table 2). There is also a growing new population of people with Asian origins (with origins in China, Korea, India...etc). We analyze long-term relation between the life expectancy (by gender) of six different groups of people living in the USA including white people, black people and all races in general as these are all influenced by the same socio-economic factors. The main contribution of this paper is to use cointegration analysis that takes into account the dependencies between the different ethnic groups in order to forecast the average life expectancy of all races in general; we have rejected the usual method of demographers(which is to make predict the average life expectancy of the males and females of inidividual *racial/ethnic* groups independently, ARIMA models). Further, we produce forecasts of the average life expectancy of all races in general.

Life expectancy (Figure 1) shows that there was an upward trend in the average life expectancy of the members of all six groups during the period 1975 to 2010. The average life expectancy of black males was the lowest and that of white females the highest. Two gender's life expectancy including white females and females of all races in general show almost the same level particularly from 2000 onward. The general consensus is confirmed here as female's life expectancy is greater than for males in each type of race as it can be seen from Figure 1. We can see for illustraton that average life expectancy level for females in general is greater than for males. Further, life expectancy of white females is greater than that of white males.

The life expectancy data have been garnered from the National Vital Statistics Reports available in Arias(2012). These reports are compiled by the government of the USA. This database has been created in order to provide information about(including trends in) mortality rates, life expectancy, infant mortality rates in the USA(by age, sex, race, cause of death, and so forth) to researchers interested in the dynamic evolution of demography. It supplies annual data on the life expectancy of white people, black people and the members of all races or origin in general and covers the period from 1940 onwards. However, Arias did not provide information regarding Indians Americans. In this paper, all races groups refers to white, black, hispanic and asian people. We work with data of all races as they can give us an indication on the future life expectancy of other *races/origin*(asia and hispanic origin) and for both sexes, because a database for such groups such as hispanic and Asians began to be collected only recently. We begin the next; by reviewing the literature on the autoregressive integrated moving average(ARIMA) model before discussing the cointegration approach.

The cointegration methodology, that we adopted for our study, involved the following steps:

- the determination of the order of integration for each of the six average life expectancies using the Augmented Dickey Fuller test (ADF), Philips-Perron test (PP) and the Kwiatowski-Phillips-Schmidt-Shin test (KPSS);
- the computation of the optimal lag value for our vector autoregressive model (VAR) model:;
- performing the Johansen cointegration test in order to compute the cointegration rank and to identilfy the variables to be entered in the cointegrating equations and in a vector error correction model (VECM);
- estimation of the VECM and VAR models and the derivation of forecasts from the models.
- comparing the forecasting of an ARIMA model and the VAR model with that of the VECM in terms of goodness of fit(we discuss this in the third section of the paper). Cointegration analysis is described in detail by Juselius(2006) and Harris and Sollis(2003).

2 Models and methodologies

In order to perform the analysis using the three models, we first visualize that life expectancy at birth for each race is non stationary(see Figure 1). We conduct ADF , PP and KPSS test unit root tests of the life expectancy data on the *racial/ethnic* groups under study. As for the unit root tests, we compute the results from the ADF tests. Under test statistics with trend criterion at 5% significance level, it shows that life expectancy is integrated of order 1. The results of using a drift criterion at 5% significant level confirm that the life expectancies of the various groups are integrated of order 1. These results are confirmed also by PP as well as KPSS tests(the results of these analysis are available upon request).

Life expectancy is modeled as a stochastic process. Rusolillo(2005), has shown that, for predicting life expectancy, the ARIMA model is superior to the Lee Carter model(1992). The ARIMA methodology consists of three phases: identification, estimation and diagnostics. The three steps are all described in Box and Jenkins(1976) and Hyndman(2013) which explain the process of choosing an appropriate ARIMA(p,d,q) for a variable. The goal is to identify the most suitable model for a time series. In prior studies, two options have been employed select the best model to use: selection of the ARIMA model by the user(Jenkins, 1976) or automatic ARIMA(see Hyndman, 2013). The general form of the ARIMA model is given by the equation (1):

$$L_t = A_0 + A_1 L_{t-1} + e_t \tag{1}$$

where A_0 is the drift term , A_1, L_{t-1} is the lagged time series and e_t is the error term.

The principal steps of the procedure are as follow:
Identification model: It consists to plot data and identify the pattern of the time series. Figure 1 shows, that life expectancies exhibit an upward trend with drift.The basics analysis also involves differencing data until they appear to be stationary. The(ADF), (PP) and (KPSS) unit root tests are used from the three unit roots corresponds to the value of parameter d.

Identification of the order of the model: After derivation of the order of stationarity, one should experiment various combinations of p, d and q with p the number of autoregressive parameters d drift, q the moving average parameters (q) to produce the best model. The decision is made consists of choosing the best model corresponding to the lowest Akaike information criterion(AIC). A Box-Cox transformation can be used to stabilize the variance if necessary.

In the third step of a standard, the researcher experiments with various combinations of p and q with the number of autoregressive parameter d(derived in the first phase), q the moving average parameter (q) to produce the best model. It is recommended that at this stage to examine the autocorrelation(ACF), the partial autocorrelation(PACF) and the diagnostic of residuals graphs be examined to help in choosing the appropriate model. Hyndman(2013) uses an automated algorithm that includes a constant. Jenkins advice to rely on the AIC and SIC(Schwarz criterion) to choose the best model.

The Fourth step is to check the diagnostics of the residuals of the choosen models by plotting the autocorrelations among them and conducting a portmanteau test. Next, the residuals are checked to determine whether they are white noise. The procedure is completed by computing the forecasts of each life expectancy from the best fitting model. Table 3 displays the best numerical results of the procedure described above. The Portmanteau test(see Table 4) indicates significant residual autocorrelation with 4, 10, 15 and 20 lags for each of the provinces. ARIMA model appears to well behave(meaning the residuals are white noise) to forecast future life expectancy as Portmanteau test suggests. All these steps are described in Russolilllo(2005).

2.1 Cointegration methodology and forecasting future life expectancy

A valid representation of a multivariate model can be done with VAR modeling. Lutkepohl(2005) describes the VAR model in general in the equation (2) as a set of dynamically dependent stationary time series:

$$L_t = \nu + \eta_1 L_{t-1} + \eta_2 L_{t-2} + ...\eta_p L_{t-p} + \mathbf{d}t + e_t \tag{2}$$

where $L_t = (L_{1,t}, L_{2,t}, ...L_{K,t})'$ is a K-dimensional time series,: η_i are matrices with the coefficient parameters $(K * K)$;, $\nu = (\nu_1, \nu_2, ...\nu_K)'$ is the intercept, e_t is the residuals part with white noise of K-dimensional with time invariant positive definite covariance matrix $E(u_t u_t^{'}) = \sum_u$, $t = 0, 1, ...T$ and p is the last lag order.

For estimating the VAR model with the objective of making predictions, we need to select the optimal lag p, which corresponds to the best predictive model with the least errors. This particular lag length of variables in the VAR model (and see below for the VECM) is derived by choosing the order p such that the value of the information criteria model such as Akaike(AIC), the HQ(Hannan-Quinn), the Schwarz Criterion(SC) - which is widely used in actuarial literature as Bayes Information criterion - and Final Prediction Error criteria(FPE)(see major details on Lutkepohl(2005)) as stated here below in the equations (3), (4), (5) and (6) are minimized:

$$AIC(p) = ln\ det(\sum(p)) + \frac{2}{T}pK^2 \tag{3}$$

$$HQ(p) = ln\ det(\sum(p)) + \frac{2ln(ln(T))}{T}pK^2 \tag{4}$$

$$BIC(p) = ln\ det\sum(p) + \frac{lnT}{T}pK^2 \tag{5}$$

$$FPE(p) = (\frac{T+p^*}{T-p^*})det\sum(p) \tag{6}$$

where T is the sample size, K is the dimension of the time series, $\sum(p)$ is estimated by $T^{-1}\sum_{i=1}^{n} e_t e_t^{'}$, p^* is the total number of parameters in each equation of the model with p^*=Kp+1, p is the lag order of the VAR model fitted to the data, and $\sum(p)$ is the estimated variance covariance matrix.

These various information criteria may identify different choices for the lag order p. However, Lutkepohl(2005) suggests that, in the case where information criteria identify different lags, the SC is the most parsimonious in choosing the best forecasting model because it produces the lowest forecast of error variance. The term p^* is the number of local parameters in each equation and p assigns the lag order VAR(p), which is said to be stationary if and only if : $det(I_k - A_1 z - .... - A_p z^p) \neq 0$ with $-1 \leq z \leq 1$. If $z = 1$, at least one of the variables inserted into VAR(p) is integrated of order 1 and therefore the variables are cointegrated. Suppose each variable of a VAR(p) process is I(d), then if each component is differenced individually this may distort the relationship between original variables. Thus, the VAR model may not be adequate. That is the why we introduce the VECM.

This conclusion introduces the Vector Error Correction Model(VECM). Let us suppose that the variables are collected in the vector $y_t = (y_{1t}, ...y_{kt})'$ and their long-run equilibrium is given by the following equation in (7):

$$\beta_1 y_{1,t} + ...\beta_K y_{K,t} = 0 \tag{7}$$

This relation may not always hold: we may sometimes have (8)

$$Z_t = \beta_{m,n} k_t \tag{8}$$

where Z_t is a stochastic variable representing the deviations from the equilibrium.

In extended form, we have (9):

$$Z_{t-1} = \beta_1 k_{1,t-1} + \beta_2 k_{2,t-1} + \beta_3 k_{n,t-1} \tag{9}$$

and (10):

$$\beta_{m,n} = \begin{pmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,n} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m,1} & b_{m,2} & \cdots & b_{m,n} \end{pmatrix} \tag{10}$$

where $k_{t-1} = (k_{1t}k_{2t}\cdots k_{nt})$.

3 Estimation procedure for the two models

This equilibrium between the variables means that they are moving together and Z_t is stable. However, this does not exclude the possibility that the variables wander as a group and driven by a common stochastic trend or better there may be cointegrated. The variables of interest in the vector y_t are said to be cointegrated of order (d, b) $y_t \succ CI(d, b)$ of all components of y_t are $I(d)$ and there exists a linear combination $Z_t = \beta^{'} y_t$ with $\beta = (\beta_1....\beta_K)' \neq 0$ such that Z_t is integrated of order $d - b$. the vector β is therefore called a cointegrating vector. Multiplying the vector β by a non zero constant we obtain other vectors. This indicates that vector may not be unique. In addition, there may be various independent cointegrating vectors involving all the variables or a subset of the variables in the vector. The matrix product $\alpha\beta^{'}$ is of dimension $K * r$ and of rank r. The matrix β is called a cointegrating matrix and α is a loading matrix. The cointegration rank is determined by the trace test and the Johansen's maximum eigenvalue test of (1988 and 1991). For further details on Johansen methodology, see Juselius(2006).

The VAR model helps to estimate the parameters of the VECM. According to Pfaff(2008) and Engle(1987), the VAR(p) in Equation (2) can be converted into VECM. The latter form of the VECM, the one used in our study is defined, in (11) as follows:

$$\Delta k_t = \Gamma_1 \Delta k_{t-1} + \Gamma_2 \Delta k_{t-2} + ... + \Gamma_{p-1} \Delta k_{t-p+1} + \Pi k_{t-p} + \nu + \varepsilon_t \tag{11}$$

where $\Gamma_i = -(I - \eta_1 + ... - \eta_i)$, for $i = 1, ...p-1$ and $\Pi = -(I - \eta_1 - ...... - \eta_p)$.

The methodologies used to estimate and test cointegrated systems are: the Engle and Granger methodology(which involves performing OLS regression of one variable on another) and the Johansen methodology, which is the one used here because it is suitable for multiple time series. The forecasts of life expectancy are yielded through the Johansen maximum likelihood procedure, which is used to seek the linear combination that is the most stationary (the Engle and Granger approach is to look for the stationary relationship that has minimum variance). Deciding whether there are cointegrating equations between variables depends on the following hypotheses.

If $r = K$, the number of cointegrated variables, r, that are stationary equals the rank(K) of Π then the model can be estimated using the standard statistical model.

If $r = 0$, this means that there are no cointegrated relationships between the variables. The variables are stationary if we take the differences of variables under study.

If $0 < r < K$ there exists two matrices, α and β, such that $\Pi = \alpha\beta'$, and there are r cointegrating relationships or $n - r$ common trends.α represents a loading matrix that measures the impact of cointegrated relationships have on the variables in the cointegrated system and each column of β represents one long run relationship. The test of cointegration is reduced to the: the rank test and the likelihood ratio test. The rank test, which is specified by the the hypotheses in (12) as follows:

$$H_0 : rank(\Pi) = r, H_1 : rank(\Pi) > r \tag{12}$$

The likelihood ratio value of such tests are measured in (13):

$$LR(r) = -(T - p) \sum (1 - \lambda_i) \tag{13}$$

where T is the length of the sample, r equals the number of cointegrated relationships, $\lambda_1 > \lambda_2 > .... > \lambda_i$ is ith the eigenvalue associated with the linear relationship.

In order to start model's estimations from different life expectancy, we performed the ADF(see Dickey and Fuller(1979)), the PP (see Philipps and Perron(1988)) and the KPSS(Kwiatkowski et.al(1992)) tests of the variables under study. These tests allow us to determine whether the variables are stationary or not. Further, the first difference tests are also performed. The PP and KPSS tests confirm the results of the ADF tesr. Life expectancy data from each variable in this study are integrated of order 1 for the three tests performed above since the first difference from each historic data is stationary with a test at 5% significant level. Consequently, the life expectancy data for each group is non-stationary.

The first step to derive the cointegration analysis consists of computing the optimal lag that drives the dynamics of life expectancy.

We compute the optimal lag length for the VAR model. The information criteria (see Table 5) yields the following results: AIC choose 3 optimal lags, HQ and FPE the value of 2 while SC only 1. Because there are these differences, we follow the recommendation of Lutkepohl(2005) and preference is given to the SC. Consequently, the optimal lag length adopted is 1.

With A= males(all races), B=females(all races), C= males(white), D=females(white), E= males(black), F= females(black), the VAR model applied of life expectancy by race is described empirically as follows:

$L_{A,t} = -0.34L_{A,t-1} - 1.04L_{B,t-1} + 1.26L_{C,t-1} - 0.03L_{D,t-1} - 0.041L_{E,t-1} + 0.54L_{F,t-1} + 0.06\lambda + 51.42$

$L_{B,t} = -0.33L_{A,t-1} - 0.53L_{B,t-1} + 0.33L_{C,t-1} + 0.17L_{D,t-1} - 0.17L_{E,t-1} + 0.70L_{F,t-1} + 0.07\lambda + 65.38$

$L_{C,t} = -0.27L_{A,t-1} - 0.84L_{B,t-1} + 1.03L_{C,t-1} + 0.14D, t - 1 - 0.079L_{E,t-1} + 0.34L_{F,t-1} + 0.085\lambda + 50.42$

$L_{D,t} = -0.43L_{A,t-1} - 0.39L_{B,t-1} + 0.32L_{C,t-1} + 0.31L_{D,t-1} - 0.14L_{E,t-1} + 0.49LF, t - 1 + 0.084\lambda + 63.65$

$L_{E,t-1} = -0.60L_{A,t-1} - 3.11L_{B,t-1} + 1.18L_{C,t-1} + 0.67L_{D,t-1} + 0.59E, t - 1 + 1.28F, t - 1 + 0.06\lambda + 80.067$

$L_{F,t-1} = 0.18L_{A,t-1} - 1.32L_{B,t-1} - 0.51L_{C,t-1} + 0.46L_{D,t-1} - 0.064L_{E,t-1} + 1.36L_{F,t-1} + 0.12\lambda + 67.35$

The diagnostic tests of the residuals (Table 6) show evidence of normality (with p-value=0.77) and no autocorrelation (with p-value=0.91). We therefore conclude that the residuals of this model are white noise on residuals according to (Table 6), and the model can be used to reasonably accurately forecast the life expectancies of different *racial/ethnic* groups in the USA.

Having obtained this VAR model, we can derive easily our VECM. It is fitted on the 6 life expectancy which span the period 1975 to 2010. We perform the trace test and compute the maximum eigenvalue for the cointegrated equations. We perform the Johansen cointegration test on the life expectancies of the different racial groups in the USA, which are all integrated of order 1. In this way, the long-run relationships between the six variables under study are estimated using maximum likelihood estimators; the results of the trace test and of the Johansen cointegration tests are shown in Table 7. The trace test statistic indicates that there is one cointegrating relationship at the 5% and 1% significance level. We therefore conclude that there is one long-run relationship among the six various groups in both cases of trace and eigenvalue(the results from eigenvalue test are available upon request) tests driven by five common trends.

The estimated cointegrated relationship is:

$$Z_{1,t-1} = A_{1,t-1} + 0.72B_{2,t-1} - 1.02C_{3,t-1} - 0.31D_{4,t-1} - 0.0655E_{5,t-1} - 0.24F_{6,t-1}$$

Here $Z_{i;t}$ represents the stationary variable that quantifies the deviation from the equilibrium of the various life expectancies analyzed.

This equation shows the dynamic evolution of the historical life expectancy between races in the USA. To illustrate, according to the equation, an increment in the variation of the average life expectancy of white females is driven by a decrease in either white males and black females and black males but an increase in the average life of both males and females of all races. The cointegrating equations shows the dependence of each race racial group's life expectancy with that of others racial groups and improve the understanding of multi-population life expectancy in the long run. These equations reveal that is a dependence between various races regarding life expectancy in the USA. The variations in the average life expectancy of one racial group is influenced not only by the lagged average life expectancy of the same group but also by improvement in the average life expectancy of other groups, as described by the following equations.

$$\begin{bmatrix} \Delta B \\ \Delta B \\ \Delta C \\ \Delta D \\ \Delta E \\ \Delta F \end{bmatrix} = \begin{bmatrix} 0.32 & -1.01 & -0.007 & 0.69 & -0.026 & 0.17 \\ -0.48 & -0.83 & 0.95 & 0.07 & -0.22 & 0.67 \\ 0.062 & -0.99 & -0.003 & 0.53 & -0.03 & -0.42 \\ -0.42 & -0.70 & 0.58 & 0.27 & -0.21 & 0.63 \\ 1.82 & -1.35 & -1.11 & -0.34 & -0.35 & 1.50 \\ 0.72 & -0.44 & -0.38 & -0.44 & -0.45 & 0.94 \end{bmatrix} \begin{bmatrix} \Delta A(-1) \\ \Delta B(-1) \\ \Delta C(-1) \\ \Delta D(-1) \\ \Delta E(-1) \\ \Delta F(-1) \end{bmatrix} +$$

$$\begin{bmatrix} -1.62 & -0.23 & 1.94 & -0.88 & -0.07 & 0.52 \\ 0.43 & -0.54 & -0.16 & -0.06 & -0.15 & 0.29 \\ -0.14 & 0.06 & 0.364 & -0.58 & -0.12 & 0.21 \\ 0.04 & 0.45 & 0.22 & -1.01 & -0.13 & 0.17 \\ -0.87 & -0.27 & 1.59 & -1.27 & -0.39 & 0.76 \\ -0.31 & -0.08 & 0.41 & -0.06 & 0.05 & -0.03 \end{bmatrix} \begin{bmatrix} A(-1) \\ B(-1) \\ C(-1) \\ D(-1) \\ E(-1) \\ F(-1) \end{bmatrix} + \begin{bmatrix} 31.31 \\ 17.48 \\ 17.65 \\ 20.82 \\ 39.08 \\ 4.68 \end{bmatrix}$$

Analysis of the diagnostic tests of the residuals(see Table 8) shows that the residuals are normal (with p-value equals to 0.50) and are non autocorrelated (p= 0.98). We conduct normality tests, based on measurement of skewness (with p-value =0.42), kurtosis (p value=0.50) and finally both kurtosis and skewness simultaneously (p-value=0.50). As these tests show that the residuals are not autocorrelated and are normal, we may conclude that the VECM is an appropriate one for forecasting future life expectancy.

In the backtesting phase, we compute the mean absolute percentage error(MAPE) values for each model using data fo the periods 2000-2010, 2001-2010, 2003-2010, 2004-2010, 2005-2010, 2006-2010, 2007-2010 and 2008-2010. First, we observe that the VAR model and the VECM fit the data well. The errors are close to 0.50% for each sample in Table 10. Second, the ARIMA model present clearly results in greater errors(with a mean of around 7%. The added component(lagged variation of each variable) on the first two models improved the performance of life expectancy prediction for each model. The VECM is found to be superior to the VAR model.

These findings are strengthened by the results shown in Table 9 which gives the various confidence intervals. Figures 2 to 8 show that the confidence interval yielded by the VECM are better than those yielded by the VAR model. However, for some of the racial groups the VECM yields confidence intervals similar to those of the ARIMA model, although for black Americans it is better in explaining uncertainty. It allows to account for more risk than other models. The error term is substantially higher with the ARIMA model. Thus, the VECM provides better results in terms of out-of-sample backtesting and confidence intervals Accordingly, life expectancy by races might be explained by taking into consideration the long run relationships between races.

We observe that the future life expectancy variations will be taking into account more risk as Table 9 shows the largest confidence interval corresponding to the VECM of each racial group. Given that, it is the model with the lowest value error predictions we can compute their life expectancy reachable by each group as in 11. We remark that the model predicts that within the next 50 years in the USA the average life expectancy of white males will increae by 8 years, that of black females by 4 years, that of black males by 11 years, and that of black females by 8 years. All races life expectancies in general will gain 9 years for males and 5 years for females. In sum we observe that model predicts tat in the USA life expectancy will continue to increase for all racial groups.

4 Conclusion

Since the last century, there has been an upward trend in average life expectancy in the developed countries. In this article, we have explored new developements in the forecasting of life expectancy in the USA by race. We have used the cointegration analysis with a VECM to estimate long run relationship between the life expectancies of different racial groups in the USA. We have conducted various cointegration analyses of average life expectancy by racial group for the period 1975 to 2010. VAR models and VECMs perform better than ARIMA models in predict have shown better performance in thehe six variables. We have found that the life expectancies from of the different groups are dependent from each other in the long run. Furthermore, we have found that over the long term life expectancy will improve for all the group *racial/ethnic* groups in the future as well as their future trends. Our study offers new insights to demographers with regard to predicting the average life expectancy of members of different racial orethnic groups, particularly those found in the USA.

References

[1] B. Pfaff, VAR, SVAR and SVEC Models: Implementation within the R package vars, *Journal of Statistical Software*, **27**(4), (2008),1-32

[2] C. Njenga-, and M., Sherris, Longevity risk and the econometric analysis of mortality trends and volatility, *Asia-Pacific Journal of Risk and Insurance*, **5**(2), (2011), 1-52.

[3] D.A. Dickey and W.A. Fuller, Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association*, **74**(366), (1979), 427-431.

[4] D. Kwiatkowski, P.C.B. Phillips, P. Schmidt and Y. Shin, Testing the null hypothesis of stationarity against the alternative of a unit root, *Journal of Econometrics*, **54**(1-3), (1992), 159-178.

[5] E. Arias, M. Heron, and J. Xu, *United States life tables, 2012. National Vital Statistics Reports*, **65**(8), (2012), Hyattsville, MD: National Center for Health Statistics.

[6] G. Jenkins, *Time series analysis: Forecasting and control*, Wiley, 1976.

[7] H. Ltkepohl, *New introduction to multiple time series analysis*, Springer 2005, XXI, (2005).

[8] J.H., Phillips, and G.E. Burch, A review of cardiovascular diseases in the white and negro races. *Medicine*, **39**(2), (1960), 241288.

[9] J. Oeppen, and J. W. Vaupel, Broken limits to life expectancy, *Science's Compass*, **296**(5570), (2002), 1029-1031.

[10] K. G. Manton, Sex and race specific mortality differentials in multiple causes of death data, *The Gerontologist*, **20**(4), (1980), 480493.

[11] K.G. Manton, S.S. Poss, and S. Wing, The black/white mortality crossover: Investigation from the perspective of the components of aging, *The Gerontologist*, **19**(3), (1979), 291299.

[12] K. G. Manton, Temporal and age variation of United States black/white cause-specific mortality differentials: A study of the recent changes in the relative health status of the United States black population, *The Gerontologist*, **22**(2), (1982), 170179.

[13] K. Juselius, *The co-integrated VAR model: Methodology and applications, Advanced Texts in Econometrics*, Oxford University Press, New York, 2006.

[14] K. Humes and H. Howard, Measurement of race and ethnicity in a changing, multicultural America, *race and Social Problems*, **1**(111), (2009), 1029-1031.

[15] L.R. Carter, Long-run relationship in differential US mortality forecasts by race and sex: Tests for co-integration, *Ageing in Advanced Industrial States, Springer*, **57**(2), (2010), 47-75.

[16] M. A. Woodbury, K.G. Manton and E. Stallard, A dynamic analysis of chronic disease development: A study of sex specific changes in coronary heart disease incidence and risk factors in Framingham, *International Journal of Epidemiology*, **10**(4), (1981), 355366.

[17] M. Russolillo and S. Haberman, Lee-Carter mortality forecasting: application to the Italian population. Faculty of Actuarial Science and Statistics, Cass Business School, London *Working paper*, (2005).

[18] National Center for Health Statistics, United States life tables: 1969-1971, Department of Health, Education, and Welfare Pub. No. (HRA) 75-1150. U.S. decennial life tables for 1969-71, 1(1). Rockville, MD: Public Health Service, (1975), Humes (2001, 2010).

[19] N. W., Rives, The effect of census errors on life table estimates of black mortality, *American Journal of Public Health*, **67**(9), (1977), 867868.

[20] P.C.B. Phillips, and P. Perron, Testing for a unit root in time series regression, *Biometrika*, **75**(2), (1988), 335-346.

[21] R.D. Lee and L.R. Carter, Modeling and forecasting U. S. mortality, *Journal of the American Statistical Association*, **87**, (1992), 659-675.

[22] R.F. Engle and C.W.J. Granger, Cointegration and error-correction: Representation, estimation and testing, *Econometrica*, **55**(2), (1992), 251-276.

[23] R. Harris and R. Sollis, *Applied tiime series modelling and forecasting*, Wiley, 2003.

[24] R. Hyndman, and G. Athanasopoulos, State space framework for automatic forecasting using exponential smoothing methods, *International Journal of Forecasting*, **18**(3), (2002), 439-454.

[25] R. Hyndman, and G. Athanasopoulos, *Forecasting: Principles and practice*, http:
otexts.com/fpp/ OTexts: Melbourne, Australia. http://otexts.org/fpp/. Accessed on 05 september 2013.

[26] S. Johansen, Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. Econometrica, *International Journal of Forecasting*, **59**(6), (2002), 1551-1580.

[27] S. Johansen, Statistical analysis of cointegration vectors,*Journal of Economic Dynamics and Control*, **12**(2-3), (1988), 231-254.

[28] S. Tulgapurkar, C.O. Puleston and M.D. Gurven, Why men matter: Mating patterns drive evolution of human lifespan. PLoS ONE, **2**(8), (2007), e785.

[29] U.S. Census Bureau, *Historical statistics of the United States, colonial times to 1970*, Bicenternnial Edition, Part 2, Chapter Z: Colonial and Pre-Federal Statistics, Washington, D.C. Series Z, **1-19**, (1975).

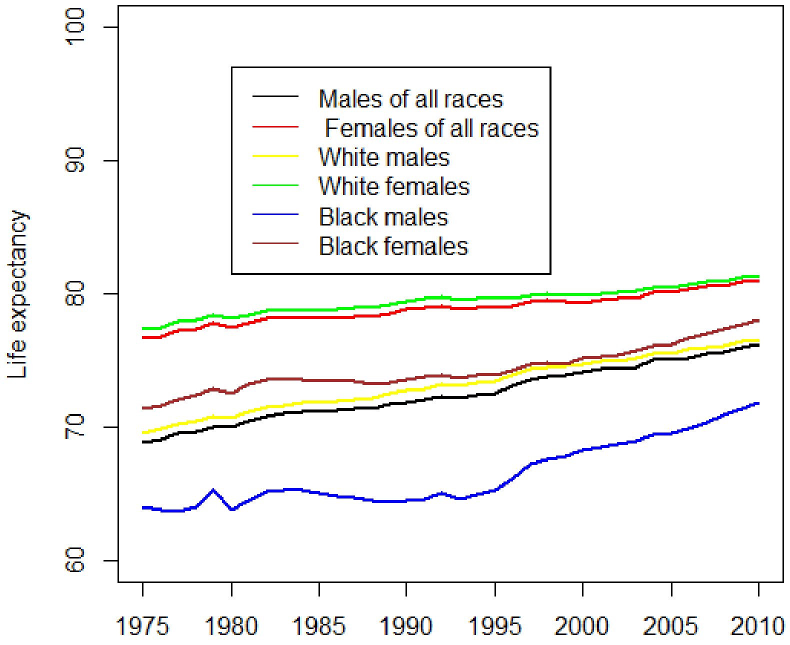

Figure 1: Average life expectancy in the USA by race and genders for the period 1975-2010

Table 1: Statisitcs census of American population 1610 to 1900

Races	1610	1710	1800	1850	1900
White	100%	86.5%	81.1%	84.3%	87.9%
Black	0%	13.5%	18.9%	15.7%	11.6%
American Indians	-	-	-	0.5%	0.3%
Asian	-	-	-	-	0.2%
Hispanic	-	-	-	0.50%	0.7%

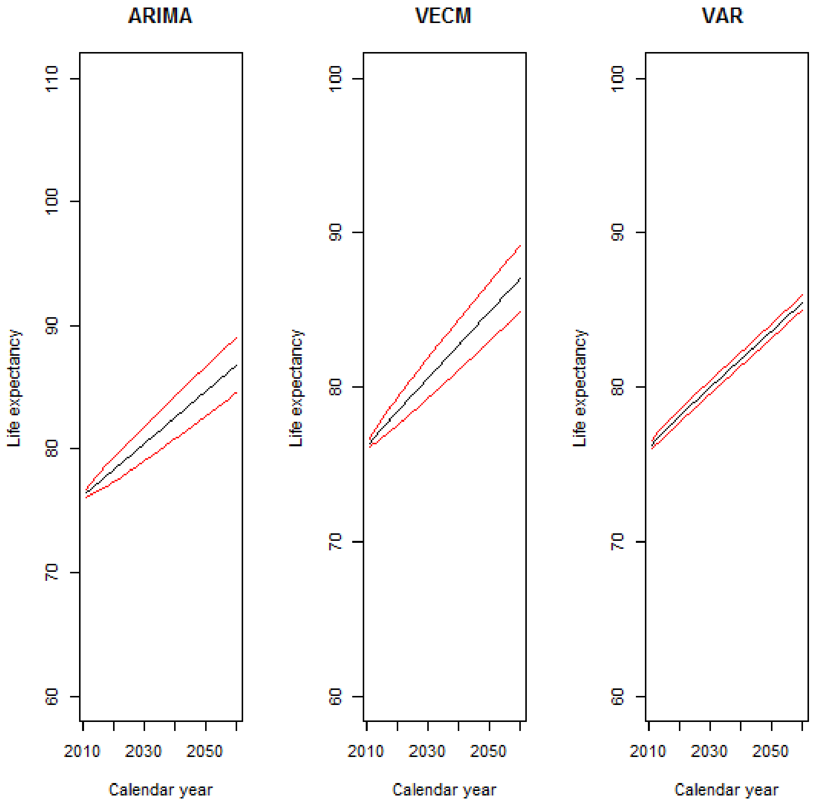

Figure 2: Projections of life expectancy in the USA: all Races

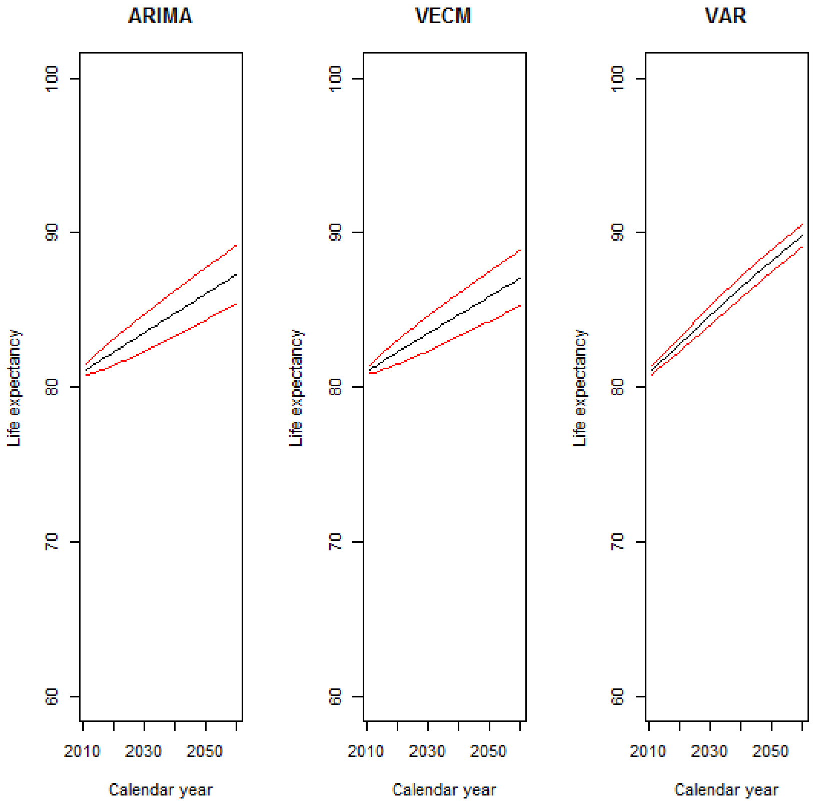

Figure 3: Projections of life expectancy in the USA: All races in general

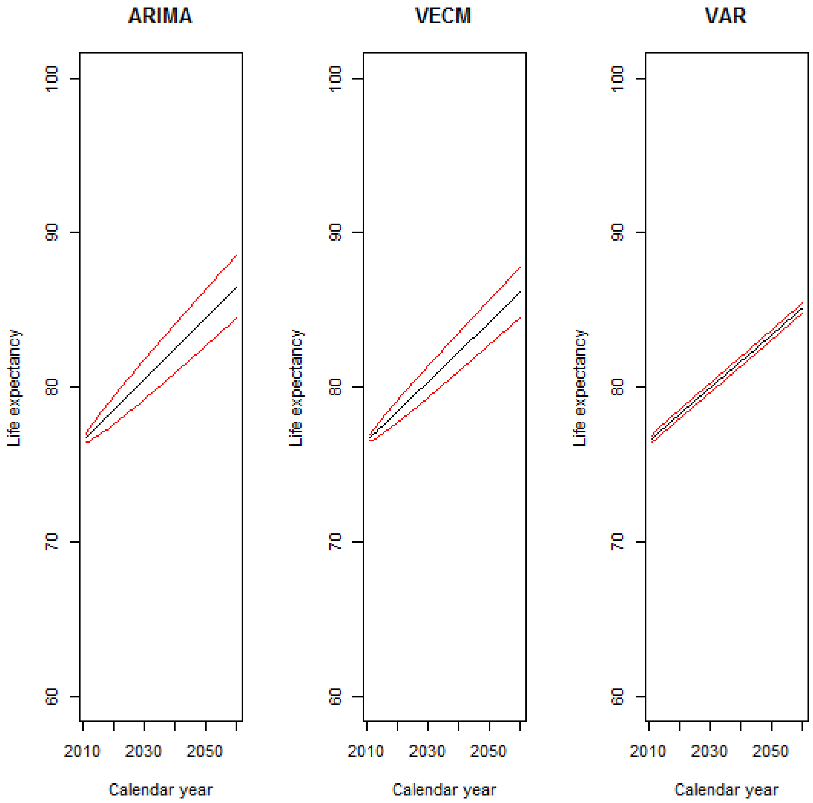

Figure 4: Projections of life expectancy in the USA: white males

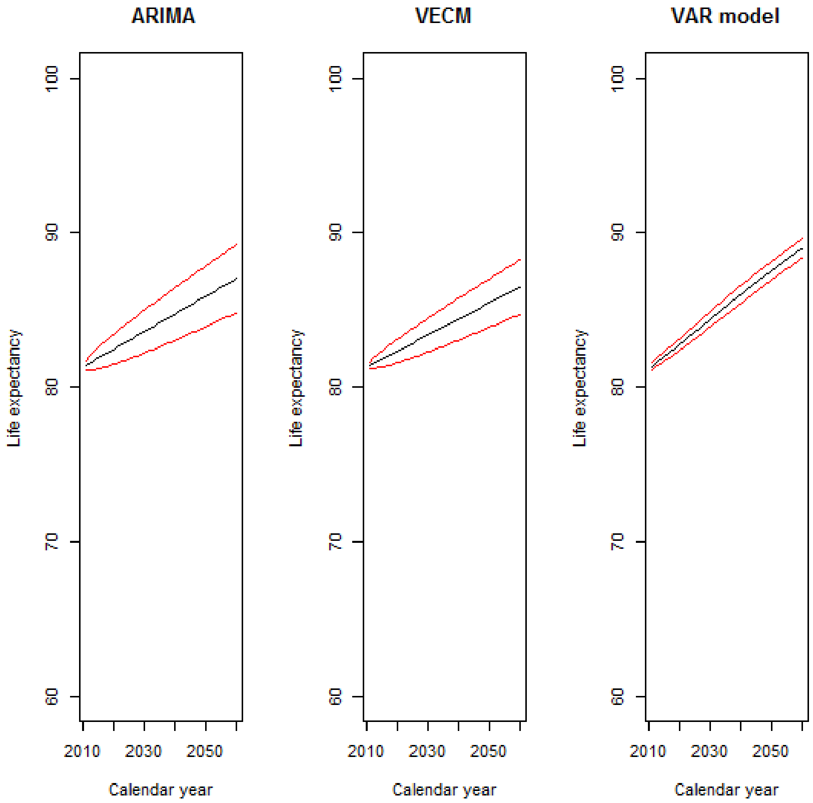

Figure 5: Projections of life expectancy in the USA: white females

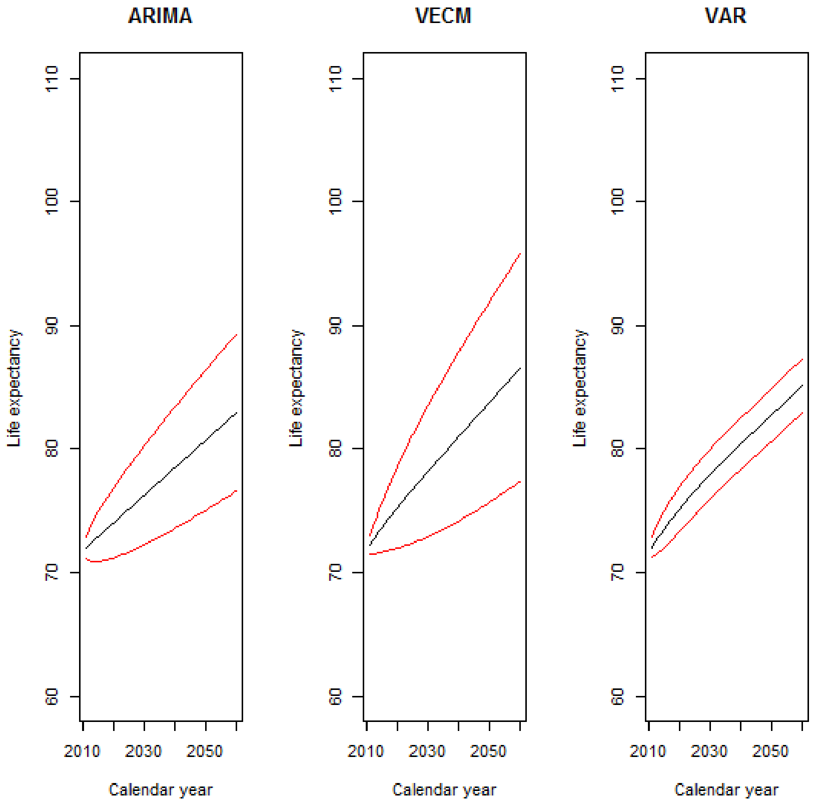

Figure 6: fig:7: Projections of life expectancy in the USA: black males

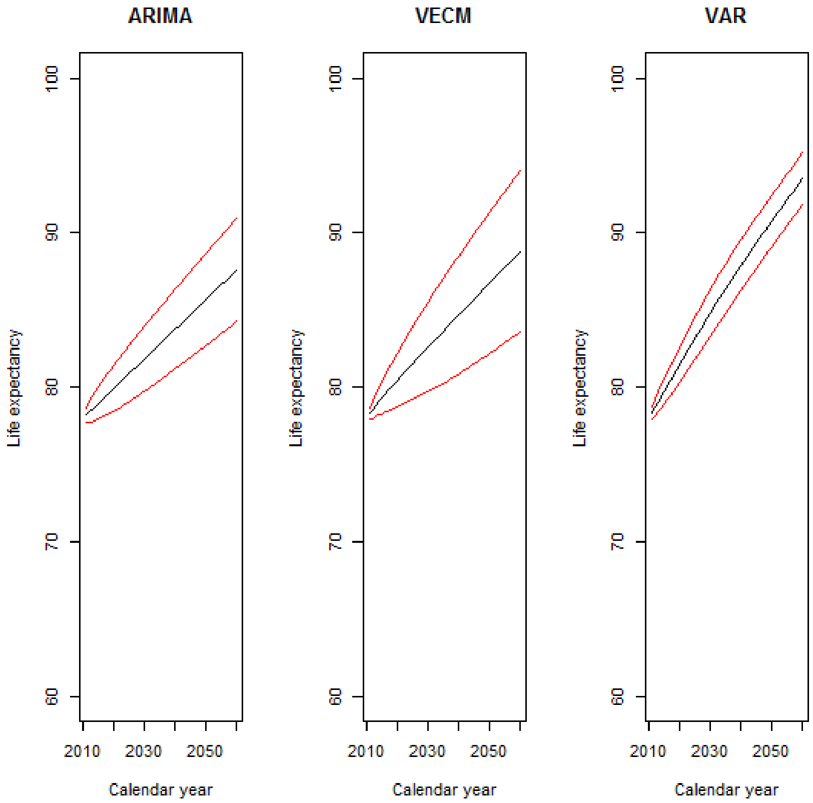

Figure 7: Projections of life expectancy in the USA: black females

Table 2: Census statistics American population 1910 to 2010

Race	1910	1950	1970	2000	2010
White	88.9%	89.5%	87.7%	75.1%	72.4%
Black	10.7%	10%	11.1%	12.3%	12.6%
American/Indian	0.3%	0.2%	0.8%	3.8%	4.9%
Asian	0.2%	0.2%	0.8%	3.8%	4.9%
Hispanic	0.9%	0.8%	0.1%	12.5%	16.3%

Table 3: The ARIMA models resulted for the analysis of life expectancy

Models	All(Males)	All(Females)	White Ma	White Fe	Black Ma	Black Fe
ARIMA(p,d,q)	(0,1,0)	(1,1,0)	(0,1,0)	(0,1,0)	(0,1,0)	(0,1,0)
ar1		-0.32	-			-
(se)		(0.15)	-	(0.10)	-	-
ma1					-	-
(se)		-	-	-	-	
ma2		-	-	-	-	-
(se)	-	-	-	-	-	-
drift	0.21	0.12	0.20	0.20	0.22	0.19
(se)	(0.02)	(0.02)	(0.025)	(0.02)	(0.077)	(0.04)

Table 4: P-values of Portmanteau test resulted from ARIMA models over the period 1921-2009

lags	A	B	C	D	E	F
4 lags	0.63	0.77	0.09	0.53	0.63	0.57
10 lags	0.66	0.87	0.24	0.91	0.66	0.94
15 lags	0.10	0.66	0.08	0.45	0.10	0.93
20 lags	0.11	0.59	0.13	0.11	0.11	0.75

Table 5: Optimal lag length for both the females and males

Information criteria	Optimal lag length
AIC	3
HQ	2
SC	1
FPE	2

Table 6: The diagnostics tests of the VAR residuals

Type of test	Autocorrelation	p-values
Autocorrelation	Portmanteau(4 lags)	0.91
Normality	Both	0.77
	Kurtosis	0.55
	Skewness	0.42

Table 7: The cointegration relationships indicated by the Trace test

Cointegrating relationship	critical values	5%	1%
5	0.64	8.18	11.65
4	8.02	14.90	19.19
3	13.19	21.07	25.75
2	19.65	27.14	32.14
1	23.58	33.32	38.78
0	57.79	39.43	46.82

Table 8: The diagnostics tests of the VECM residuals

Type of test	Autocorrelation	p-values
Autocorrelation	Portmanteau(4 lags)	0.98
Normality	Both	0.5076
	Kurtosis	0.5078
	Skewness	0.42

Table 9: The Confidence interval of models VAR, VECM and ARIMA for the 6 groups derived from predictions 50 years ahead

Races	VECM	VAR	ARIMA
All sexes Males	(0.23-2.13)	(0.24-0.46)	(0.31-2.24)
All sexes Females	(0.23-1.82)	(0.26-0.72)	(0.35-1.89)
White females	(0.21-9.21)	(0.23-0.31)	(0.28-2.04)
White Males	(0.35-5.21)	(0.23-0.62)	(0.31-3.12)
Black Females	(0.35-7.66)	(0.80-2.17)	(0.9-6.35)
Black Males	(1.08-6.33)	(0.40-1.68)	(0.47-4.72)

Table 10: The average MAPE for ARIMA, VAR and VECM models of the six groups

Out-of-sample period	VECM	VAR	ARIMA
2000-2010	0.5%	2.31%	5.1%
2001-2010	0.55%	2.3%	5.8%
2002-2010	0.41%	0.62%	6.2%
2003-2010	1.02%	0.77%	6.41%
2004-2010	1.1%	0.60%	6.69%
h=2005-2010	1.39%	0.48%	7.37%
h=2006-2010	0.280%	0.62%	7.34%
h=2007-2010	0.29%	0.32%	7.9%
h=2008-2010	0.19%	0.42%	8.39%

Table 11: VECM forecasts of average life expectancy(in years) for the six groups

Year	All males	All races females	White males	White females	Black males	Black females
10	78.43	82.25	78.42	82.32	75.27	80.46
20	80.59	83.46	80.34	83.35	78.18	82.59
30	82.73	84.65	82.27	84.39	80.39	84.67
40	84.87	85.85	84.20	85.43	83.77	86.73
50	87.01	87.05	86.112	86.47	86.56	88.79