Projecting the long run relationship of Multi-population life expectancy by races

A. Ntamjokouen[*]

Abstract

Over the last century, life expectancy of US population is being increased as past and actual census investigate. In this paper we attempt to predict future life expectancy by race for Black, white and all races in general in the USA. In doing this, we apply the cointegration analysis appeared recently in the actuarial literature. It takes into account dependence between between the six variables and shows a better fit with better performance in predicting life expectancy at birth by races. Also, it reveals that long-run relationship between life expectancy of all races do exist. These results are helping and giving some new insights to demographers on the performace of future life expectancy for each group of race living in the USA.

Keywords: Life expectancy by race, VAR, VECM

1 Introduction

Life expectancy as well as mortality rates are getting improved, across developped countries, as several studies such as Tulgapurkar et. al(2007) and Oeppen(2002) have shown. The United States of America are not an exception since they have highlighted also signs of improvements of life expectancy of the population in general and also the different races groups living in the

---
[*]corresponding author, email: achille.ntamjokouen@unibg.it, Department of Management science, Economics and Quantitative methods, University of Bergamo, Italy

country. To illustrate this, in 1970 life expectancy for male people was respectively of 68 for white race and 60 years old for black. In 2010, there were 76.5 and 71.8. Same results are observed for females as in 1970 white female life expectancy was 75.6 and black 68.3. Furthermore, in 2010 there were 81.3 and 78 for respectively white and black. The historical census of the USA population take on mainly of 2 group races: white and black people. However in recent years, there are many other group's races of people who have immigrated into the country such as Latinos and Asian and the incidence of their presence is becoming visible in the american society. Most literature existing on this topic have focused on predicting the pair black-white death rates such as Rives 1977; NCHS (1975), Manton(1980, 1982), Manton et. al(1979) Philipps and Burch(1960), Woodbury et. al(1981), Manton et. al(1979). We will not focus our study only on two races groups mainly described in time series database by black and white but we include also life expectancy from all races in general. In this, our study will also take into account of the common trends to predict life expectancy from each group and also from all races in general.

In 1610, 100% of inhabitants in the USA were white according to official statistics and 10 years later 0.9% of newcommers were black(see table 9 retrieved from U.S Bureau Census 1975). In 1780 while 79.9% of population were white, 20.7% the rest was black(see Humes(2001,2010)). From 1850, hispanic population was added into the statistics in the USA census as 0.5% of population. After 10 years, the official census investigates asian newcomers as 0.1% of population and indians as 0.1%. In 1970 people of hispanic origin were 4.4% of the american population. In 2010, there were mainly 4 ethnicities in the USA including asian, hispanic, white as well as blackasian. During 5 centuries the latter two racial group have been the two main habitants of the USA as we can see statistics here from table 1. The old pair is not anymore the major group as black people have been surpassed by Latinos recently(see table 2). There is also a growing new population mainly driven by people with asian origin(chinese, corean, indians..etc). Since, the population in the USA is not anymore a mixture of 2 groups, it arises issues on life expectancy of population by race with additional new comers. We aim at analysing crossover of 6 different groups of people living in the USA including white, black and all races in general with both females and males sexes since there are influenced by the same socio-economic factors. The main contribution of this paper is to use cointregration analysis that take

2

into account the dependencies between different ethhnic group to forecast life expectancy of all races in general turning down the fact that usually demographers do the predictions of each category independently through the model ARIMA. Further, we produce forecast of life expectancy for all races in general from the pair black and white races population.

Life expectancy (figure 1) shows an increasing trend for all the races during the period which spans from 1975 to 2010. We observe that Black male life expectancy show the lowest level and white females the highest. Three gender's life expectancy including black females, white males as well as males of other sexes show almost the same level particularly from 2000 onward. Furthermore, white females and females from other sexes present almost the same level from 1975 to 2010. Overall, life expectancy from different races in the USA show an increasing trend and also some convergence point between some of the component under study. 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.

Life expectancy data have been provided by the National Vital Statistics Reports available at www.cdc.gov. It is managed by the Government of the United States of America. This database has been created in order to provide information on mortality data, life expectancy, infant mortality and trends by characteristics such as age, sex, race, cause of death in USA to anyone interested in the dynamic evolution of demography. It supplies annual data on life expectancy which span from 1940 to 2010 for white, black and other races. In this, reference to all races group refer to white, black, hispanic and asian into the same pool. We will work with data from other races as they can give us an indication on future life expectancy from other races and for both sexes. To begin with, we present the literature on ARIMA model before looking at the cointegration approach model.
The latter methodology, which applies the cointegration, includes several steps as follows:
- The determination of order of integration for each of the 6 life expectancy using the Augmented Dickey Fuller, Philips-Perron as well as KPSS Test;
- The computation of the optimal value of lag of vector autoregressive model;
- the Johansen cointegration test which investigates the cointegration rank and specify which variable will enter in the cointegrated equations and in the Vector Error Correction model;

- furthermore the estimation of VECM and VAR and the forecasting of derived models.
- finally, third part will be dedicated to the comparison between the ARIMA, VAR model forecasting and the VECM in terms of goodness of fit. Steps of cointegration analysis are described by Juselius(2006) and Harris and Sollis(2003).

2   Models and methodologies

In order to perfom the analysis from the three models, we first visualize(see figure 1) that life expectancy at birth for each race are non stationary. We compute the unit root test from the life expectancy data of each race through the ADF test, PP test along with KPSS test. As for unit root tests we compute the results from ADF tests. Compared with the test statistics with trend criterion, it shows that life expectancy is integrated of order 1. The results under drift criterion also strenghtens that life expectancy from the various groups are integrated of order 1. These results are confirmed also by both results obtained by PP-tests as well as KPSS tests.

The life expectancy is modelled as stochastic processes. In Rusolillo(2005), it is shown that ARIMA lead to better performance over the Lee Carter model(1992) in predicting life expectancy. The ARIMA methodology consists of three phases including identification, estimation and diagnostics. The three steps are all described in Box and Jenkins(1976) or Hyndman(2013) which explain the process to choose an appropriate ARIMA(p,d,q) for a variable. The goal is to identify the correct model that fits well time series. There are two options used by existing literature to select the best model to use: selection of model by the user(Jenkins, 1976) or the automatic arima(see Hyndman, 2013). In general ARIMA is described as in (4):

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

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

The principal steps of the procedure are the following:

Identification model: It consists to plot data and identify the pattern of

the time series. As we can observe from figure 1, it emerges that life expectancy presents a positive trend with drift.The basics analysis consist also of differencing data until they appear to be stationary. The unit root tests Augmented Dickey Fuller(ADF), Philips-Perron(PP) as well as Kwiatkowski-Phillips-Schmidt-Shin(KPSS) tests are useful to determine the level of stationarity. The order of integration derived from the 3 unit roots corresponds to the value of parameter d. The successive path consists of choosing the best model corresponding to the lowest Akaike information criterion(AIC).

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. A Box-Cox transformation could be used to stabilize the variance if necessary.

As for the third step: the researcher, in a standard procedure, experiments 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 at this stage to examine the autocorrelation(ACF) and the partial autocorrelation(PACF) as well as the diagnostic of residuals graphs to help chosing of the appropriate model. Hyndman(2013) uses an automated algorithm which consists of the inclusion of a constant. Also, Jenkins advise to choose the best model relying on the AIC and SIC(Schwarz criterion).

Fourth phase is devoted to check the diagnostics of residuals from the choosen models by plotting the autocorrelation residuals and experimenting a Portmanteau test of the residuals. Next, the residuals are checked to see if they are white noise. The procedure is completed by computing the forecasts through the choice of the best fitting model. Table 11 displays the best numerical results of the procedure described above. The Portmanteau test(see table 10) indicates significant residual autocorrelation with 4, 10, 15, 20 lags for each of the provinces. These results suggest that ARIMA appears to well behave with white noise disturbances to forecast future life expectancy.

2.1 Cointegration methodology and forecasting future life expectancy

A valid representation of a multivariate model can be done with Vector Autoregressive as Lutkepohl(2005) describes in general in the equation (4) as a

5

set of dynamically dependent stationary time series described by:

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

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

For estimating the VAR with the objective of making prediction, we need to select the optimal lag p which corresponds to the best predictive model with least error. This particular lag length of variables in the VAR model (and the VECM see later) is derived by choosing the order p which minimizes 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 - as well as Final Prediction Error criteria(FPE)(see major details on Lutkepohl(2005)) as stated here below in the equations (5), (6), (7) and (8):

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

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

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

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

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 fitted with the data, $\sum(p)$ is the estimated variance covariance marix, .
The different information criteria may identify different choices for the lag order p. However, Lutkepohl(2005) suggests that, in the case where criteria choose different lags, the SC is the most parsimonious in choosing the best forecasting model because it produces the lowest forecast of error variance.

p^* is the local number of parameters in each equation and p assigns the lag order VAR(p) which is said to be stationary if and only if : $det(I_k - A_1z - .... - A_pz^p) \neq 0$ with $-1 \leq z \leq 1$. If $z = 1$, at least one of these the variables inserted into VAR(p) are integrated of order 1 and therefore cointegration does exist between the variables. Suppose each variable of a VAR(p) process is I(d), then each component is differenced individually it may distort the relationship between original variables. Therefore the VAR model may be not adequate. That is the reason we introduce the cointegrated models.

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

$$\beta_1y_{1t} + ...\beta_Ky_{Kt} = 0 \tag{7}$$

This relation may not hold at any time and we may sometimes have

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

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

In an extended form we have:

$$Z_{t-1} = \beta_1k_{1,t-1} + \beta_2k_{2,t-1} + \beta_3k_{n,t-1} \tag{9}$$

and in an extended form we have:

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

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

3 Estimation procedure of the two models

This equilibrium between variables means that there are moving together and Z_t is stable. However this does not exclude that variables wander as a group and may be driven by common stochastic trend or better there may

7

be cointegrated. The variables of interest in the vector y_t are called 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 $d - b$. So the vector β is called cointegrating vector. Multiplying the vector β by a non zero constant we obtain other vectors. This indicate 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 $\alpha \beta'$ are of dimension $K * r$ amd of rank r. The matrix β is called a cointagrating matirx and α is a loading matrix. The decision for the cointegration rank is based on the trace test and the maximum eigenvalue test of Johansen(1988 and 1991). For further details on Johansen methodology, see Juselius(2006).

The VAR model helps to estimate the parameters on the equation (9). According to Pfaff(2008) and Engle(1987), the VAR (p) in the equation (4) can be converted into VECM. The latter form of VECM, which will be used in our case studies is defined, in (9) as follows:

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

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: Engle and Granger methodology which perform OLS regression of one variable on another and Johansen methodology which will be used here because it is suitable for multiple time series. The forecasts of mortality indices are yielded through the Johansen Maximum Likelihood methodology. It seeks the linear combination that is most stationary whereas Engle and Granger looks for stationary relationship that has minimum variance. The decisions on whether there are cointegrated equations between variables are taken according to the following hypotheses.
If $r = K$, the number of cointegrated variables, r, which is stationary equals the rank(K) of Π and then the model will be estimated by 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 above.
If $0 < r < K$ there exists 2 matrices, α and β, such that $\Pi = \alpha \beta'$, there will

be r cointegrating relationships or $n - r$ common trends.α represent loading matrix measures the impact of cointegrating relations have on the variables in the cointegrated system while each column of β represents one long run relationship. The test of cointegration is reduced to the two following hypotheses:

The rank test, which is specified by the hypotheses in (10) as follows:

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

and the likelihood ratio value of such tests are measured in (11):

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

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

In order to start model's estimations from different life expectancy, we performed the Augmented Dickey Fuller(see Dickey and Fuller(1979)), the Philips Perron (see Philipps and Perron(1988)) as well as the KPSS(Kwiatkowski et.al(1992)) tests for the variables under study. These tests allow us to determinate whether variables are stationary or not. Further, the first difference tests under the three tests are also perfomed. The PP and KPSS tests confirm the same results as the ADF. The critical statistics indicate that 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. Consequently life expectancy from each group is non stationary.

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

We compute the optimal lag length of the VAR model. The Information criteria(see table 5) shows the following results: AIC choose 3 optimal lags, HQ and FPE the value of 2 while SC only 1. Since they differ each from others, Lutkepohl(2005) shows that the preference will be given to SC. Consequently, the lag lenth is 1.

[table 5]

9

Denoting by A= all races males, B=All races Females, C= White Males, D=White Females, E= Black Males, F=Black Females, The VAR model applied to the life expectancy by race in the USA is described empirically as follows:

Diagram of fit and residuals for a
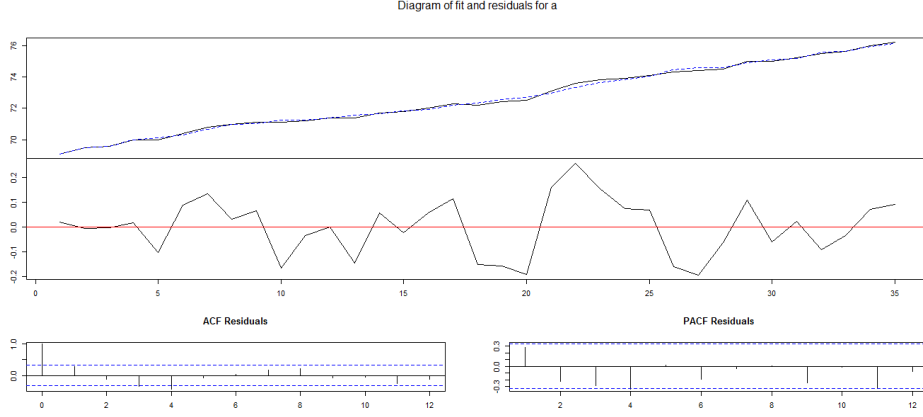
Diagnostics of residuals with reference to All sexes

$$L_{a,t} = -0.34L_{a,t-1} - 1.04L_{b,t-1} + 1.26L_{c,t-1} - 0.03L_{d,t-1} - 0.041L_{e,t-1} + 0.54L_{f,t-1} + 0.06\lambda + 51.42$$

$$L_{b,t} = -0.33L_{a,t-1} - 0.53L_{b,t-1} + 0.33L_{c,t-1} + 0.17L_{d,t-1} - 0.17L_{e,t-1} + 0.70L_{f,t-1} + 0.07\lambda + 65.38$$

$$L_{c,t} = -0.27L_{a,t-1} - 0.84L_{b,t-1} + 1.03L_{c,t-1} + 0.14d, t - 1 - 0.079L_{e,t-1} + 0.34L_{f,t-1} + 0.085\lambda + 50.42$$

$$L_{d,t} = -0.43L_{a,t-1} - 0.39L_{b,t-1} + 0.32L_{c,t-1} + 0.31L_{d,t-1} - 0.14L_{e,t-1} + 0.49Lf, t - 1 + 0.084\lambda + 63.65$$

$$L_{e,t-1} = -0.60L_{a,t-1} - 3.11L_{b,t-1} + 1.18L_{c,t-1} + 0.67L_{d,t-1} + 0.59e, t - 1 + 1.28f, t - 1 + 0.06\lambda + 80.067$$

$$L_{f,t-1} = 0.18L_{a,t-1} - 1.32L_{b,t-1} - 0.51L_{c,t-1} + 0.46L_{d,t-1} - 0.064L_{e,t-1} + 1.36L_{f,t-1} + 0.12\lambda + 67.35$$

The diagnostic test on residuals(table 6) show evidence of normality (with p-value=0.77) as well as non autocorrelation of residuals (with p-value=0.91). Since normality as well as non autocorrelation of residuals are held, we conclude that this model, which shows white noise on residuals, well behaves to forecast future life expectancy from different groups in the USA.

[table 6]

3.1 Vector Error Correction model

After obtaining VAR, we can derive easily Vector Error Correction Model. It is fitted on the 6 life expectancy which span from the period 1975 to 2010. We performed the trace test and the Eigenvalue maximum for the cointegrated equations. We test the Johansen cointegration on the life expectancy from different groups in the USA that are all integrated of order 1. In this way, long run relationships are estimated between the 6 variables under study. They are estimated using the maximum likelihood estimators and the results are presented here below for the trace test values:

The Johansen cointegration tests results are shown here in table 7. The trace test statistic indicates that there is one cointegrating relationship at the 5% significance level. Therefore we conclude that there is one long-run relationship among the 6 various groups in both cases of trace and eigenvalue(the results from eigenvalue test are available upon request) tests driven by 5 common trends.

[table 7]

The estimated cointegrated relation is:

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

Here $Z_{i;t}$ represents the stationary variable which quantify the deviation from the equilibrium of the various life expectancies analysed.
This equation shows the changing environment on the historical life expectancy between races in the USA. To illustrate, according to the equation, an increment in the variation of life expectancy of white females will produce a decrease in either white males and black females and black males

11

but an increase in all races life expectancy both for males and females. The cointegrating equations shows the dependence of each race life expectancy group with others and improve the understanding of multi-population life expectancy in the long run. In these equations, we see that there is an inter-dependence between various races in the United States of America regarding life expectancy. The life expectancy improvement in one group race is not only influenced by the lagged life expectancy of the same race but also influenced by improvement from another group as it is described by the following equations.

$$
\begin{bmatrix} \Delta a \\ \Delta b \\ \Delta c \\ \Delta d \\ \Delta e \\ \Delta f \end{bmatrix} = \begin{bmatrix} 0.32 & -1.01 & -0.007 & 0.69 & -0.026 & 0.17 \\ -0.48 & -0.83 & 0.95 & 0.07 & -0.22 & 0.67 \\ 0.062 & -0.99 & -0.003 & 0.53 & -0.03 & -0.42 \\ -0.42 & -0.70 & 0.58 & 0.27 & -0.21 & 0.63 \\ 1.82 & -1.35 & -1.11 & -0.34 & -0.35 & 1.50 \\ 0.72 & -0.44 & -0.38 & -0.44 & -0.45 & 0.94 \end{bmatrix} \begin{bmatrix} \Delta A(-1) \\ \Delta BC(-1) \\ \Delta NB(-1) \\ \Delta NS(-1) \\ \Delta ON(-1) \\ \Delta Q(-1) \end{bmatrix} +
$$

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

Analysis on diagnostic of residual test(see table 8) shows also that the residuals are normal (with p-value equals to 0.50) and are non autocorre-lated(p= 0.98). We observe that for the normality test, we have a test based on the skewness(with p-value =0.42), another based on kurtosis(p value=0.50) and finally both based on kurtosis and skewness(p-value=0.50). Since these two tests show that residuals are non autocorellated and normal, we can concluded that VECM is appropriate to forecast future life expectancy.

[table 8]

3.2 Forecasting procedure and Backtesting of the various models

In the phase of backtesting, we compute the MAPE each model from 6 different sample period 2000-2010, 2001-2010, 2003-2010, 2004-2010, 2005-2010 as well as 2006-2010, 2007-2010, 2008-2010. We oberve first that the VAR and VECM are reliable in fitting well data. The error are close to 0.50% for each sample in table 9. Second, ARIMA model present definetly results with higher error performance mean which is around 7%. The added components on the firsts two models have improved the performance of life expectancy prediction for each model. The VECM realized better performance than VAR.

[table 9]

These findings are strengthened by results showed in table 10 which show the confidence interval. In the 6 races as you can see in figures 5 to 9, the confidence interval produced by VECM presents better performance over VAR. However, the VECM model present similar confidence interval for some races with ARIMA even though for Black race it is better in explaining uncertainty. It allows to account for more risk than other models. The error term is really higher for ARIMA. Consequently the VECM presents better results in terms of backtesting out-of-sample and of confidence interval. Accordingly, life expectancy by races must be explained by taking into consideration the long run relationships between races.

[table 10]

We observe that the future life expectancy improvements will be taking into account more risk as table 11 shows the largest confidence interval corresponding to the VECM of each group race. Given that, it is the model with the lowest value error predictions we can compute their life expectancy reachable by each group as in table 11. We remark that life expectancy for white males will be increasing by 8 years and females by 4 years while Black males and females respectively by 11 years and 8 years within 50 years of predictions. All races life expectancy in general will gain 9 years for males and 5 years for females. Overall we observe that life expectancy will be increasing in the future for all groups of races in the USA.

[table 11]

4   Conclusion

Since the last century, life expectancy trend is increasing in developed countries. In this article We have explored new developements on forecasting life expectancy by race in the USA. We used the cointegration analysis through Vector Error Correction model to estimate long run relationship between differents groups in the USA subdivided by race. The cointegration analysis are applied on life expectancy by races through the period from 1975 to 2010. Vector Autoregressive and Vector Error Correction Models have shown better performance than ARIMA model in predicting the 6 variables. We have seen that the life expectancy from differents groups are dependent from each other in the long run. Furthermore, the results show also that life expectancy will be improving for all the group races in the future as well as their future trends. These results are helping and giving some insights to demographers on the performace of future life expectancy for each group of race(Black and white and females) also the life expectancy from all races in the USA.

References

[1] Biffis, E., Blake D., *Mortality-Linked Securities and Derivatives.* Discussion Paper PI-0901-http://www.pensions-institute.org/

[2] Carter, L.R.,*Long-run relationship in differential US mortality forecasts by race and sex: tests for co-integration, Ageing in advanced industrial states, Springer, ISBN 978904811135523, 47-75* 2010.

[3] Gibson, C. Jung K.,*Historical Census Statistics on Population Totals By Race, 1790 to 1990, and By Hispanic Origin, 1970 to 1990, For the United States, Regions, Divisions, and States". Population Division, U.S. Census Bureau, Working Paper Series 56*2002.

[4] Grieco E., Cassidy R., *Overview of Race and Hispanic Origin.*2001

[5] Harris , R. Sollis,R., *Applied Time Series Modelling and Forecasting, Wiley.*2003

[6] Heligman L. and J. H. Pollard. *The age pattern of mortality Journal of the Institute of Actuaries 107:49-80.*1980.

[7] Humes, K., Nicholas J., and Roberto R., *Overview of Race and Hispanic Origin: Census Briefs. U.S. Census Bureau*2001

[8] Humes, K., Nicholas J., and Roberto R., *Overview of Race and Hispanic Origin: Census Briefs. U.S. Census Bureau*2010.

[9] Hyndman R. *A state space framework for automatic forecasting using exponential smoothing methods, International Journal of Forecasting, 18(3),439-454.*2002

[10] Hyndman, R.J. and Athanasopoulos. *Forecasting: principles and practice, http:*
otexts.com/fpp/ Accessed on 05 september 2013.

[11] Jenkins G, *Time series analysis: Forecasting and control, 575 pp.*1976.

[12] Joliffe, I.T. *Principal Component Analysis, 2nd ed. 2002, XXIX, 489 p,Springer.*2002.

[13] Johansen, S. *Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. Econometrica, JSTOR 2938278 , 59(6):1551-1580.*1991

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

[15] Juselius, K.. *The Integrated VAR Model: Methodology and Applications. Advanced Texts in Econometrics, New York: Oxford University Press.*2006.

[16] Lee, R.D. and L.R. Carter. *Modeling and Forecasting U. S. Mortality. Journal of the American Statistical Association, 1992. 87: 659-675..*1992

[17] Manton, K. G. *ex and race specific mortality differentials in multiple causes of death data.The Gerontologist, 20, 480493 .*1980

[18] Manton, K. G. *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, 170179. .*1982

[19] Manton, K. G., Poss, S. S., & Wing, S. *The Black/white mortality crossover: Investigation from the perspective of the components of aging. The Gerontologist, 19, 291299. .*1979.

[20] 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

[21] Njenga, C., Sherris, M. *Longevity Risk and the Econometric Analysis of Mortality Trends and Volatility. Asia-Pacific Journal of Risk and Insurance, Vol. 5,P. 1:52..*2011.

[22] Pfaff, B. *VAR, SVAR and SVEC Models: Implementation with the R Package vars. Journal of Statistical Software, 27(4).*2008.

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

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

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

[26] Tuljapurkar, S., Puleston, C.O., Gurven, M.D. *Why Men Matter: Mating Patterns Drive Evolution of Human Lifespan. PLoS ONE 2(8): e785. doi:10.1371. .*2007.

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

[28] Woodbury, M. A., Manton, K. G., Stallard, E. *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, 355366 .*1981.

[29] Yang S, Yue Jack, Huang H., *Modeling Longevity Risks using a Principal Component Approach: A Comparision with Existing Stochastic Mortality Models Working Paper.*2008
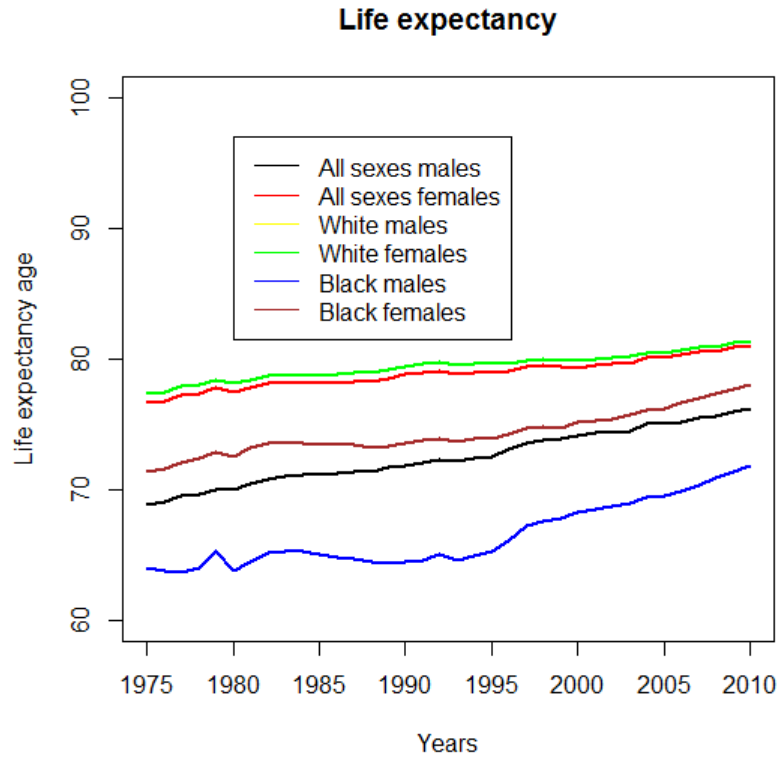
Figure 1: Life expectancy by races from all groups in the USA

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/Indian	-	-	-	0.5%	0.3%
Asian	-	-	-	-	0.2%
Hispanic	-	-	-	0.50%	0.7%

Table 1: Statisitcs census of American population

Races	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 2: Statistics census of American population

Models	All Sex Ma	All Sex Fe	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 3: The best ARIMA models resulted from the analysis of life expectancy

lags	All Sex Ma	All Sex Fe	White Ma	White Fe	Black Ma	Black Fe
4 ags	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 4: P-values of Portmanteau test resulted from ARIMA models over the period 1921-2009

Information criteria	Life expectancy by race
AIC	3
HQ	2
SC	1
FPE	2

Table 5: Optimal lag length for both females and males

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

Table 6: The diagnostics tests of residuals of VAR

Cointegrating relationship	critical values	5pct	1pct
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 7: The cointegration relations under Trace test

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

Table 8: The diagnostics tests of residuals of VECM

Out-Of-Sample	VECM	VAR	ARIMA
h=2000-2010	0.5%	2.31%	5.1%
h=2001-2010	0.55%	2.3%	5.8%
h=2002-2010	0.41%	0.62%	6.2%
h= 2003-2010	1.02%	0.77%	6.41%
h=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 9: The average MAPE for models, ARIMA VAR and VECM for the 6 provinces

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 Confidence interval of models VAR, VECM and ARIMA for the 6 provinces derived from predictions 50 years ahead

Year	All males	All races females	White Males	White females	Black Males	Black Fem
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

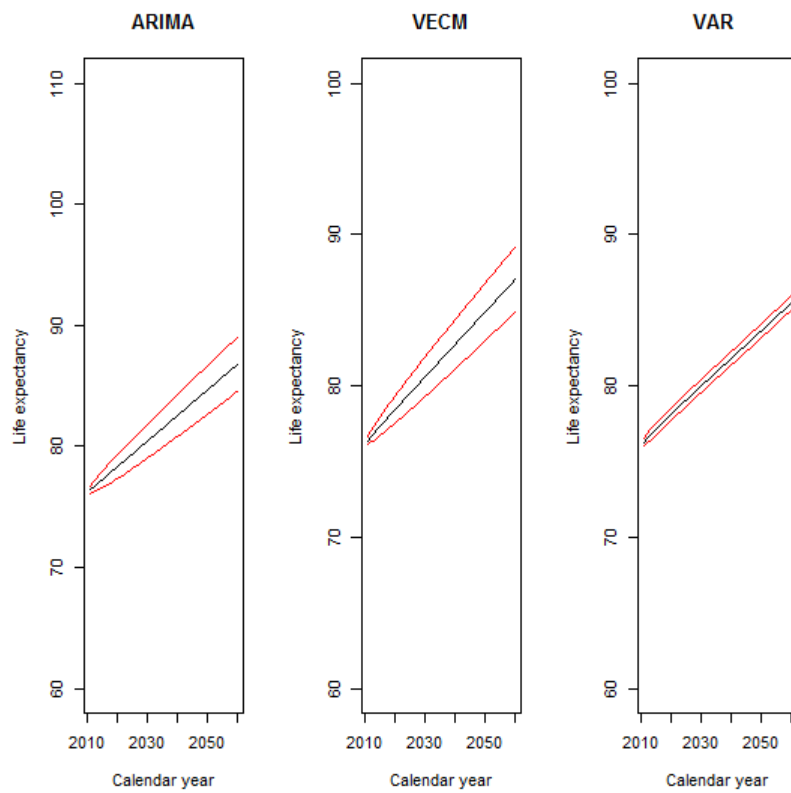Table 11: Future forecast of life expectancy with model VECM for the 6 provinces

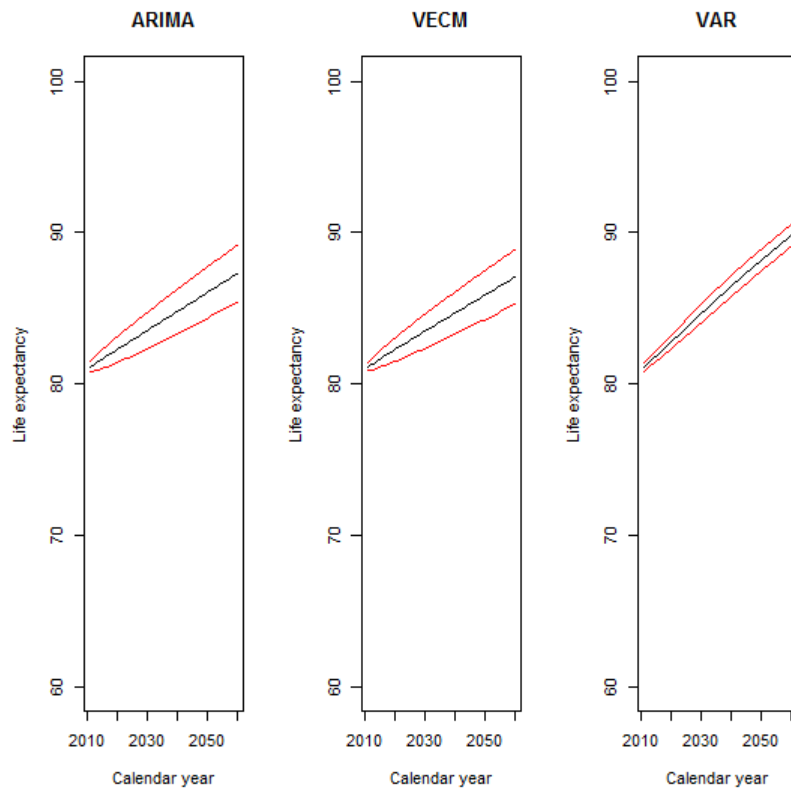Figure 2: Projections males life expectancy from all other races in the USA

Figure 3: Projections females life expectancy from all other races in the USA
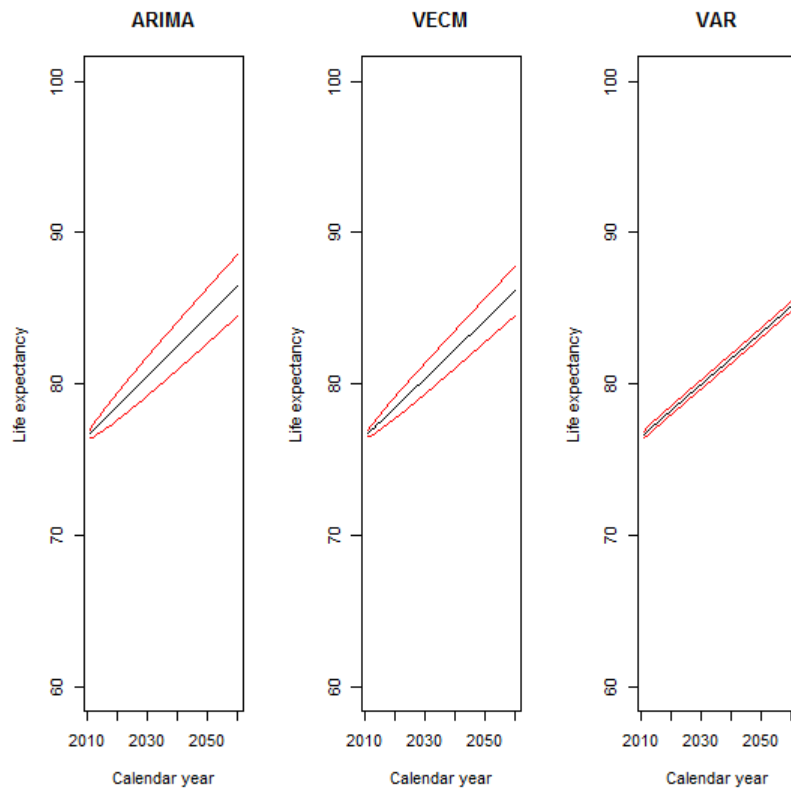
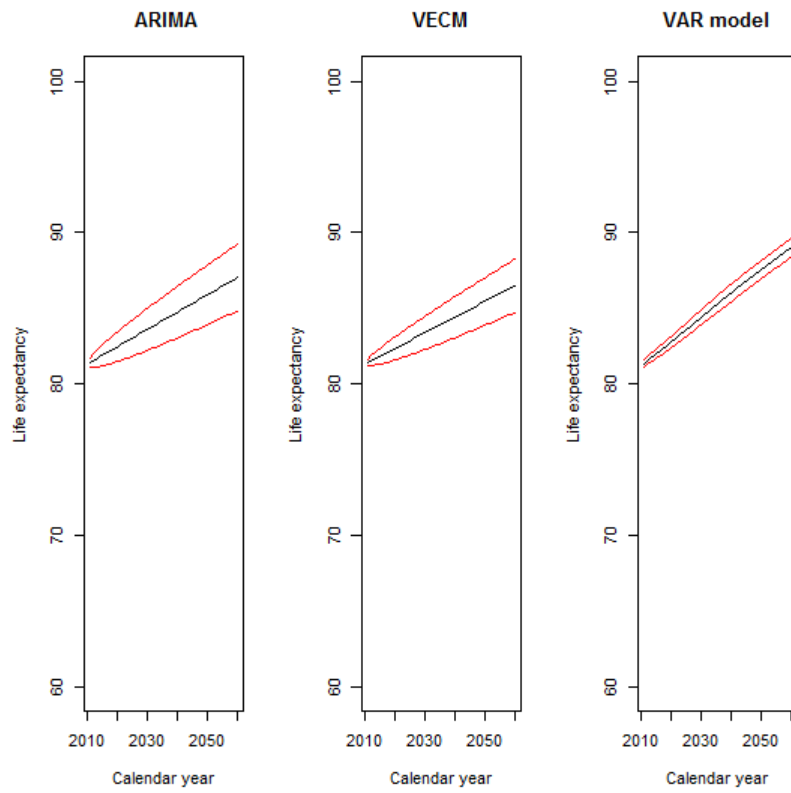Figure 4: Projections life expectancy from white males in the USA

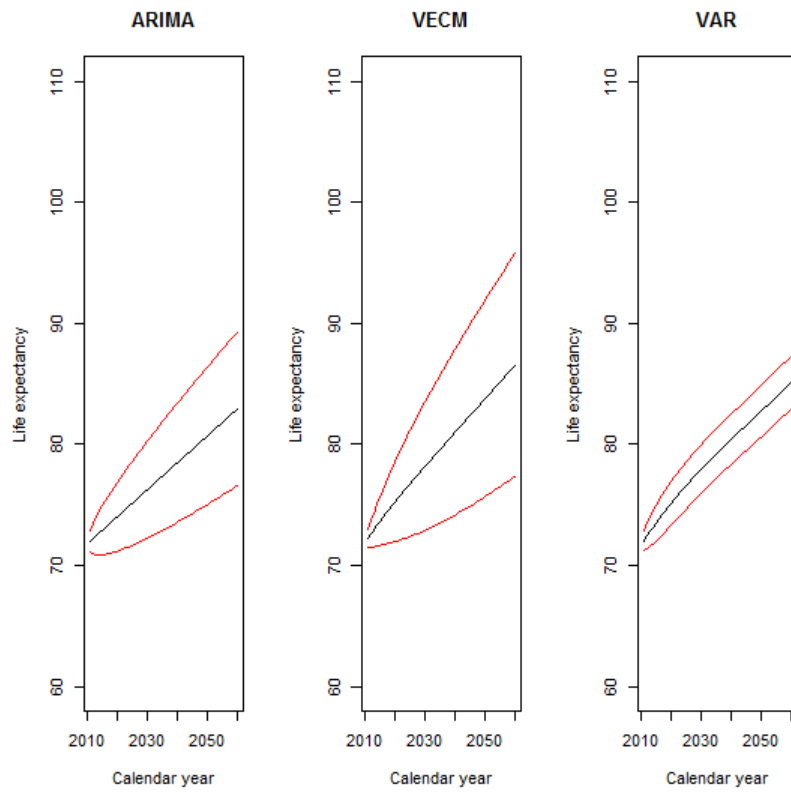Figure 5: Projections life expectancy by races from white females in the USA
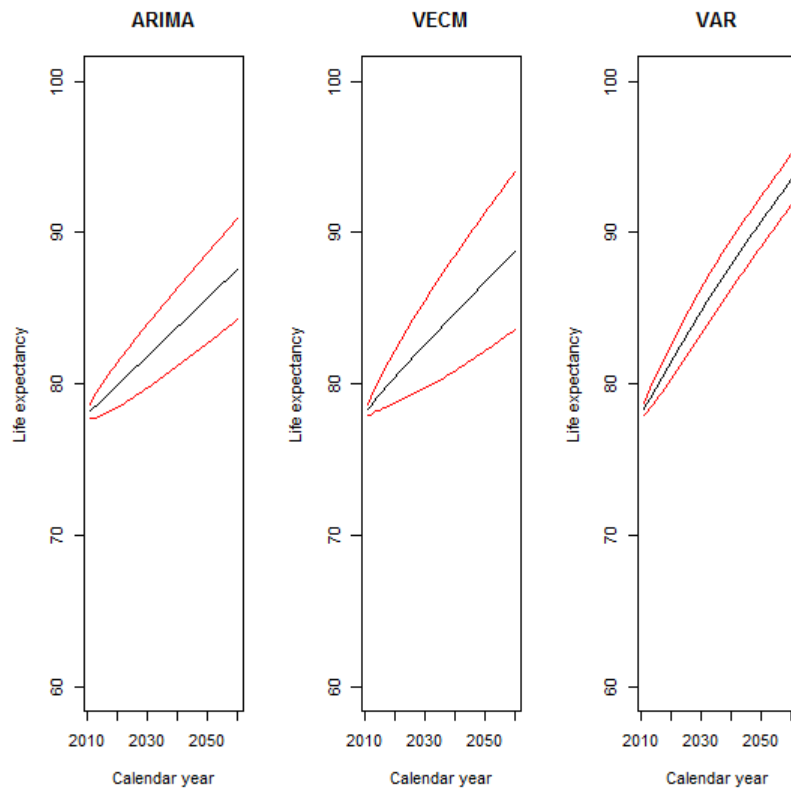
Figure 6: Projections life expectancy from black males in the USA

Figure 7: Projections life expectancy from black females the USA