



Munich Personal RePEc Archive

Multi-Scenarios Population Projection for Algeria using R

Flici, Farid

Research Center in Applied Economics for Development (CREAD)

6 May 2020

Online at <https://mpra.ub.uni-muenchen.de/119600/>
MPRA Paper No. 119600, posted 06 Jan 2024 20:57 UTC

Muti-Scenarios Population Projection for Algeria using R

Farid FLICI *

* Centre for Research in Applied Economics for Development (CREAD).

BP 197, Street Djamel eddine El-Afghani, Rostomia, Bouzareah, 16011, Algiers, Algeria

e-mail: farid.flici@cread.dz

webpage: farid-flici.github.io

May 6, 2020

Contents

1	Notions	3
2	Data preparation	4
2.1	Baseline population	4
2.2	Mortality Forecasts	4
2.3	Fertility Forecasts	6
2.4	Number of males for 1 female among newborn	7
3	Multi-scenarios projection	7
3.1	Defining the scenarios	7
3.2	Results	8
4	Plotting dynamic multi-scenarios population pyramids	9
5	References	12

- To cite this document:

FLICI, F (2020). Multi-scenarios Population Projection for Algeria Using R. A textbook published with Gitbook, Available at: [<https://farid-flici.gitbook.io/multi-scenarios-population-projection-for-algeria-/>], Version of 06-05-2020.

In this manual, we present a methodology of doing a multi-scenarios population projections for Algeria using R. The methodology used consists of the Cohort-Component Method. A detailed methodology of doing mono-scenario population projection in R was already presented in Flici (2020). Here, mortality Scenarios are defined using the predictive intervals of a stochastic forecast, while the Fertility scenarios are expert-based. Here, 5 scenarios are simulated, the Age-Specific fertility rates are then defined using the Lee-Carter Model for fertility (Lee, 1993). Later on, we shows how to make dynamic visualizations of the multi-scenario projections.

1 Notions

The method used is the multi component method. This method considers a population during a year 0 to be composed from different *cohorts* in regard to their year of birth. Hence, the population P_t is composed from different sub-populations $P_{x,t}$, with x going from age 0 to age w . Then, the time evolution of each cohort is deduced based of the projected mortality surface.

For x from 0 to w , $P_{x+1,t+1} = P_{x,t} * (1 - qx,t)$ with qx,t is the Age-Specific-Mortality Rate (ASMR).

Then, the population aged 0 in the beginning of the year $t + 1$ is deduced from the population of females P^f during the year t (mid-year population) at the procreation ages, i.e., 15 – 49, and the corresponding Age Specific Fertility Rates (ASFRs), noted $f_{x,t}$. We can write:

$$P_{0,t+1} = \sum_{x=15}^{49} P_{x,t} * \frac{P^f_{x,t} + P^f_{x,t+1}}{2} * f_{x,t}$$

If immigration data is available and has a significant effect on population dynamics, the population evolution is adjusted accordingly. If not, we assume simply that immigration sold is null.

A detailed overview of the cohort component method can be found in Smith et al. (2013).

There is also an *R*package *popReconstruct* allowing populations projections using R. [<https://cran.r-project.org/web/packages/popReconstruct/popReconstruct.pdf>]

In order to perform the projection, a set of information need to be made available:

- The horizon of the projection;
- A baseline population;
- Projected Age Specific Mortality Rates, for males and females, extended from the reference year up to the horizon of the projection;
- Projected Age Specific Fertility Rates, for females at reproduction ages;
- The Sex ratio at birth, or the number of males corresponding to 1 female among newborns;

Using a multi-scenarios approach aims at considering the uncertainty in presenting the projection results. The scenarios of the evolution of each component need to be defined; based on it, the population evolution scenarios are defined.

Here, Following the central scenario, Age Specific mortality rates (ASMRs) are projected using the coherent model of Hyndman et al. (2013). A detailed overview can be found in Flici (2016a). The scenarios "High" and "Low" are defined based on the 95% prediction interval.

The scenario of fertility were defined to give levels of total fertility rate (TFR) corresponding to "1.5", "2.1", "2.5", "3", and "3.5". Then the corresponding $f_{x,t}$ surfaces are calculated using the model of Lee (1993). This method, as well as an application on the Algerian data, is explained in Flici (2016b).

The final number of scenarios is related to the number of scenario used for each component. Here, in our case, Three scenarios for mortality and five for fertility are adopted. It makes 15 scenarios in total.

2 Data preparation

In order to perform a multi-scenarios projections, the different components need to be projected according to the corresponding scenarios. A baseline population is also needed.

2.1 Baseline population

We use the population of Algeria, 2015, by single age as a reference population. Then, the year-to-year population growth is conducted according the the mortality projected rates. First, we upload the baseline population, while separating males and females.

```
pop_2015 <- read.table('https://firebasestorage.googleapis.com/v0/b/gitbook-28427.
.appspot.com/o/assets%2F-MaP249sVLyWPT_DFN0y%2F-MaP41tA697v7RX1Pxx5%2F-MaP545y8GWTM5Kjff
B8%2FBaseline_pop_dz_2015.txt?alt=media&token=f73ca16f-e067-46ff-b121-e92690043472',
header=T)
PopM15<-as.matrix(pop_2015$males)
rownames(PopM15)<-c(0:99)
PopF15<-as.matrix(pop_2015$females)
rownames(PopF15)<-c(0:99)
```

2.2 Mortality Forecasts

Next, the projected Age-Specific Mortality Rates (ASMRs) need to be uploaded for males and females, age detailed from 0 to 120. In concern of the mortality evolution assumptions, 3 scenarios are considered based on the results of the stockastic forecasts and the underlying confidence intervals at a confidence level of 95%.

```
MortMM<- read.table("https://firebasestorage.googleapis.com/v0/b/gitbook-28427.appspot
.com/o/assets%2F-MaP249sVLyWPT_DFN0y%2F-MaP41tA697v7RX1Pxx5%2F-MaP5y6C8HRCWkVrf65X%2FM
```

```

M_m.txt?alt=media&token=d5106285-fdc3-4e92-97b5-ba8c7a2305bb", header=T)
MortMM<-as.matrix(MortMM)
MortMM<-MortMM[,2:56]
rownames(MortMM)<-c(0:120)
MortMF <- read.table("https://firebasestorage.googleapis.com/v0/b/gitbook-28427.appspot.com/o/assets%2F-MaP249sVLyWPT_DFN0y%2F-MaP41tA697v7RX1Pkx5%2F-MaP5u0Nnt0rqq2ErAQ8%2FM_f.txt?alt=media&token=9cfc9d43-2b6e-412e-b79b-3a84c66f1dd2", header=T)
MortMF<-as.matrix(MortMF)
MortMF<-MortMF[,2:56]
rownames(MortMF)<-c(0:120)
MorthM <- read.table("https://firebasestorage.googleapis.com/v0/b/gitbook-28427.appspot.com/o/assets%2F-MaP249sVLyWPT_DFN0y%2F-MaP41tA697v7RX1Pkx5%2F-MaP51dt93uhYbt2WY2A%2FH_m.txt?alt=media&token=f18b35d3-b7cf-4543-8143-7d13d2baa059", header=T)
MorthM<-as.matrix(MorthM)
MorthM<-MorthM[,2:56]
rownames(MorthM)<-c(0:120)
MorthF <- read.table("https://firebasestorage.googleapis.com/v0/b/gitbook-28427.appspot.com/o/assets%2F-MaP249sVLyWPT_DFN0y%2F-MaP41tA697v7RX1Pkx5%2F-MaP5jmL4b_kQWP5IfGc%2MF_h_f.txt?alt=media&token=c68edb7e-c22e-430c-97ea-3f40cd5689b0", header=T)
MorthF<-as.matrix(MorthF)
MorthF<-MorthF[,2:56]
rownames(MorthF)<-c(0:120)
MortLM <- read.table("https://firebasestorage.googleapis.com/v0/b/gitbook-28427.appspot.com/o/assets%2F-MaP249sVLyWPT_DFN0y%2F-MaP41tA697v7RX1Pkx5%2F-MaP5scHLsDS1bHofgw9%2FM_L_m.txt?alt=media&token=f2686b91-a52c-4810-91a0-07b89a8e19b7", header=T)
MortLM<-as.matrix(MortLM)
MortLM<-MortLM[,2:56]
rownames(MortLM)<-c(0:120)
MortLF <- read.table("https://firebasestorage.googleapis.com/v0/b/gitbook-28427.appspot.com/o/assets%2F-MaP249sVLyWPT_DFN0y%2F-MaP41tA697v7RX1Pkx5%2F-MaP5r2AT9xJjOPeZmA3%2MF_L_f.txt?alt=media&token=89273d33-3934-4744-a7fd-c6e370425559", header=T)
MortLF<-as.matrix(MortLF)
MortLF<-MortLF[,2:56]
rownames(MortLF)<-c(0:120)

```

In order to simplify the data structure and the combination of the scenarios when projecting the population, the different ASMRs by scenario are arranged in lists of matrices by sex, namely, *MortM* and *MortF*

for males and females respectively.

```
# Males
MortM<-vector("list",3)
MortM[[1]] <- MortHM
MortM[[2]] <- MortMM
MortM[[3]] <- MortLM
#Females
MortF<-vector("list",3)
MortF[[1]] <- MortHF
MortF[[2]] <- MortMF
MortF[[3]] <- MortLF
```

2.3 Fertility Forecasts

In concern of Fertility, 5 scenarios are considered : Very low, low, medium, high, very high, which correspond to Total Fertility Rates (TFR) of 1.5, 2.1, 2.5, 3 and 3.5 respectively by 2070.

```
FerM<- read.table("https://firebasestorage.googleapis.com/v0/b/gitbook-28427.appspot.com/o/assets%2F-MaP249sVLyWPT_DFN0y%2F-MaP41tA697v7RX1Pxx5%2F-MaP5iJR38pDNbd18tKV%2FFM.txt?alt=media&token=c287a6a4-1df0-4666-bb8d-362f2a795b44", header=T)
FerM<-FerM[1:35,2:57]
rownames(FerM)<-c(15:49)
```

```
FerL<- read.table("https://firebasestorage.googleapis.com/v0/b/gitbook-28427.appspot.com/o/assets%2F-MaP249sVLyWPT_DFN0y%2F-MaP41tA697v7RX1Pxx5%2F-MaP5UXAhG0aoxI13RA1%2FFL.txt?alt=media&token=a1e6d1b4-9b39-44a0-91ce-1398cbec3963", header=T)
FerL<-FerL[1:35,2:57]
rownames(FerL)<-c(15:49)
```

```
FerH<- read.table("https://firebasestorage.googleapis.com/v0/b/gitbook-28427.appspot.com/o/assets%2F-MaP249sVLyWPT_DFN0y%2F-MaP41tA697v7RX1Pxx5%2F-MaP5IaOTLKhfdRFoiQh%2FFH.txt?alt=media&token=ff398a34-4f02-445d-b7a5-6d23a0e6bbd9", header=T)
FerH<-FerH[1:35,2:57]
rownames(FerH)<-c(15:49)
```

```
FerLv<- read.table("https://firebasestorage.googleapis.com/v0/b/gitbook-28427.appspot.com/o/assets%2F-MaP249sVLyWPT_DFN0y%2F-MaP41tA697v7RX1Pxx5%2F-MaP5VqfgaxeC8ceKCAy%2FFL%2B.txt?alt=media&token=a1edacaf-bf82-4b77-ae2c-edec34aea329a", header=T)
FerLv<-FerLv[1:35,2:57]
rownames(FerLv)<-c(15:49)
```

```
FerHv<- read.table("https://firebasestorage.googleapis.com/v0/b/gitbook-28427.appspot.com/o/assets%2F-MaP249sVLyWPT_DFN0y%2F-MaP41tA697v7RX1Pxx5%2F-MaP5VqfgaxeC8ceKCAy%2FFH%2B.txt?alt=media&token=ff398a34-4f02-445d-b7a5-6d23a0e6bbd9", header=T)
```

```
t.com/o/assets%2F-MaP249sVLyPT_DFN0y%2F-MaP41tA697v7RX1Pxx5%2F-MaP5Jhw1ZBLhPhIuCRb%2FFH%
2B.txt?alt=media&token=4d13c49b-6e77-4fb2-8afd-6353777525e0", header=T)
FerHv<-FerHv[1:35,2:57]
rownames(FerHv)<-c(15:49)
```

The projected Age-Specific Fertility Rates (ASFRs) provided by the different scenarios are arranged into a list of matrices from the lowest to the highest fertility levels.

```
Fer<-vector("list",5)
Fer[[1]] <- FerLv
Fer[[2]] <- FerL
Fer[[3]] <- FerM
Fer[[4]] <- FerH
Fer[[5]] <- FerHv
```

2.4 Number of males for 1 female among newborn

According to historical data, the number of males corresponding to 1 females is equal to 1.045. In other words, among 2,045 new-borns, 1,000 are girls and 1,045 are boys.

```
a=1.045
```

3 Multi-scenarios projection

3.1 Defining the scenarios

If we define the three scenarios of mortality evolution as (1,2,3) corresponding to (high, average, low) and the five scenarios of fertility evolution as (1,2,3,4,5) to correspond to (very low, low, average, high, very high). The combination of the possible scenarios can be obtained by the combination of the three scenarios about mortality with the five scenarios of fertility must lead to a total of 15 scenarios. This combination must be managed through a transition matrix.

```
Mort <-c(1,2,3)
Fert <-c(1,2,3,4,5)
# The combination of the different scenarios
SC_pop <- data.matrix(expand.grid(Mort, Fert))
colnames(SC_pop) <-c("Mort", "Fert")
show(SC_pop)
```

The matrix of scenarios has the following form:

```
Mort Fert
[1,] 1 1
[2,] 2 1
```



```
[3,] 3 1
[4,] 1 2
[5,] 2 2
[6,] 3 2
[7,] 1 3
[8,] 2 3
[9,] 3 3
[10,] 1 4
[11,] 2 4
[12,] 3 4
[13,] 1 5
[14,] 2 5
[15,] 3 5
```

It displays the combination of the 3 scenarios about futur evolution of mortality and the 5 scenarios regarding fertility into 15 scenarios.

3.2 Results

First, we create listes of matrices to be filled with the projection results. For each sex, a list of 15 matrices (1 matrice per scenario) is created. Each matrix has ages in rows (from 0 to 120) and years in columns (from 2015 to 2070). The first column is to be filled by the baseline population.

```
PopM <- vector("list",15)
PopF <- vector("list",15)
for (k in 1:15) {
PopM[[k]]<-matrix(0,nrow=121,ncol= 56)
rownames(PopM[[k]])<-c(0:120)
colnames(PopM[[k]])<-c(2015:2070)

PopF[[k]]<-matrix(0,nrow=121,ncol=56)
rownames(PopF[[k]])<-c(0:120)
colnames(PopF[[k]])<-c(2015:2070)
PopM[[k]][1:121,1]<-rbind(PopM15,as.matrix(rep(0,21),ncol=1,nrow=21))
PopF[[k]][1:121,1]<-rbind(PopF15,as.matrix(rep(0,21),ncol=1,nrow=21))
}
```

Then, the year-to-year population are deduced using the corresponding survival probabilities. If we set Px, t to be the population at age x during the year t , and qx, t to be the probability of dying at age x during the year t , then we can write:

$$Px + 1, t + 1 = Px, t * (1 - q_{x,t})$$

Then, the new-borns during the year t , noted B_t are estimated using the formula:

$$B_t = \sum_{x=15}^{49} \frac{P^f_{x-1,t-1} + P^f_{x,t}}{2} * f_{x-1,t-1}$$

with $P^f_{x,t}$ to be the population of females aged x at the beginning of the year t and $f_{x,t}$ the Age-Specific Fertility Rate (ASFR) at age x and year t .

The number of new-borns B_t are split into boys and girls and are introduced as Populations at age 0 during the next year. We can write: $P^m_{0,t} = B_t - 1 * \frac{1.045}{1+1.045}$ for males and $P^f_{0,t} = B_t - 1 * \frac{1}{1+1.045}$ for females.

```

for (k in 1:15){
for (i in 2 : 56) {
for (j in 2: 121) {
  PopM[[k]][j,i]<-PopM[[k]][j-1,i-1]*(1-MortM[[ SC_pop[k,1] ]][j-1,i-1])
  PopF[[k]][j,i]<-PopF[[k]][j-1,i-1]*(1-MortF[[ SC_pop[k,1] ]][j-1,i-1])
}
  PopM[[k]][1,i]<-as.matrix(t(PopF[[k]][16:50,i-1]+PopF[[k]][16:50,i])/2)
  %% as.numeric(as.matrix(Fer[[ SC_pop[k,2] ]][,i-1]))*(a/(1+a))
  PopF[[k]][1,i]<-as.matrix(t(PopF[[k]][16:50,i-1]+PopF[[k]][16:50,i])/2)
  %% as.numeric(as.matrix(Fer[[ SC_pop[k,2] ]][,i-1]))*(1/(1+a))
}
}

```

It results in:

```

PopM[[4]][1:3,1:5]
  2015 2016 2017 2018 2019
0 525931 511352.1 514473.8 516157.9 516325.4
1 505803 511707.4 497080.7 499867.6 501359.5
2 494789 505140.0 511015.0 496396.3 499172.5

```

4 Plotting dynamic multi-scenarios population pyramids

Trying to make different *ggplot* objects in a *list* then trying to plot then using a *loop* will results in duplicating the last plot on all the subplots (for all scenario). The solution consists in gathering all the datasets in one data frame. It means passing from a matricial form where data is initially structured by age (in rows) and years (in columns) and population number in cells; separatly by sex and scenario into a linear form where each line

gives a combination of "Sex", "Age", "year", "Scenario", and "population number". For each scenario, the corresponding scenario of mortality and fecundity are added. This can be done using *melt* function which is available under *reshape2* package.

```
library(reshape2)
Pop <-cbind(rbind( melt(PopM), melt(PopF)), c(rep(1, nrow(melt(PopM) ) ),
rep(2, nrow(melt(PopF) ) )))
add two columns for fertility (1.5, 2.1, 2.5, 3) and LE (low, Medium, High)
Pop <- cbind(Pop, matrix(rep(NA), ncol=2, nrow=nrow(Pop)))
for (i in 1:15) {
Pop[ Pop[,4]== i, 6] <- SC_pop[i,1] Pop[ Pop[,4]== i, 7] <- SC_pop[i,2]
}
colnames(Pop)<-c("age", "year", "population", "scenario", "sex",
"Life_Expectancy", "Fertility")
```

Population numbers for males are set with a negative sign

```
Pop[Pop[,5]==1, 3] <- - Pop[Pop[,5]==1, 3]
```

This results in a new data structure:

```
Pop[1:4,]
2015 2016 2017 2018 2019
0 525931 511352.1 514473.8 516157.9 516325.4
1 505803 511707.4 497080.7 499867.6 501359.5
2 494789 505140.0 511015.0 496396.3 499172.5
```

We call the packages needed in plotting a dynamic multi-plot.

```
library(ggplot2)
library(animation)
library(dplyr)
library(ggthemes)
library(gridExtra)
library(grid)
library(reshape2)
```

To change the label of the variables, we use:

```
LE.labs =c("Life Expectancy: Low", "Life Expectancy: Medium",
"Life Expectancy: High")
names(LE.labs)=c("1", "2", "3")
FER.labs=c("Fertility: 1.5", "Fertility: 2.1", "Fertility: 2.5",
"Fertility: 3", "Fertility: 3.5")
names(FER.labs)=c("1", "2", "3", "4", "5")
```

Finally, we create a Gif of the evolution of the different scenarios.

```

saveGIF({
for (t in 2015:2070){
E <- ggplot( Pop[ Pop[,2]==t ,] ,aes(x=Pop[ Pop[,2]==t ,1],
fill=as.factor( Pop[ Pop[,2]==t, 5] )))
  E1 <- E + geom_bar(aes(y=Pop[ Pop[,2]==2015 ,3]), stat="identity", width=0.75,alpha=1)
+ coord_flip()
  E2 <- E1+ geom_bar(aes(y=Pop[ Pop[,2]==t ,3]), stat="identity", width=1,alpha=0.5)
  E3 <- E2 + facet_grid( Life_Expectancy ~ Fertility, labeller=labeller( Life_Expectancy=
LE.labs, Fertility=FER.labs))
  E4 <- E3 + scale_y_continuous(breaks = seq(-1000000, 1000000, 500000), labels =paste(
as.character(c(seq(1, 0, -0.5), seq(0.5,1,0.5))), "m"), limits= c( -1050000,1050000))
  E5 <- E4 + scale_x_continuous(limits=c(0,110), breaks = seq(0, 100, by=20),
labels = seq(0, 100,by=20))
  E6 <- E5 + theme(strip.text.x = element_text(size = 14, face="bold"), strip.text.y =
element_text( size = 14, face="bold") , plot.title = element_text(hjust = 0.5,
size=18,face="bold"),panel.border = element_rect(fill=NA, size=1,colour="black"),
axis.text=element_text(size=14, face="bold"))
  E7 <- E6 + ylab("Population Number")+ xlab("Age")
+ ggtitle(paste(" Population Projection, Algeria :", t, "Vs. 2015"))
  E8 <- E7 + scale_fill_manual(values=c("blue", "red"), labels=c("M", "F")) +
guides(fill=guide_legend(title="Sex", size=14, face="bold"))
  plot(E8)
} }, movie.name = 'pyrampopdz.gif', interval = 0.3, ani.width = 1400, ani.height = 720)

```

It results in the following visualization:

Figure 1: Dynamic multi-scenarios population pyramid

5 References

- Flici, F. (2016a). Coherent mortality forecasting for the Algerian population. Presented at Samos Conference in Actuarial Sciences and Finance, Samos, Greece (May).
- Flici, F. (2016b). Projection des taux de fécondité de la population algérienne à l horizon 2050. MPRA Paper No. 99077, posted 12 Mar 2020. [https://mpra.ub.uni-muenchen.de/99077/1/MPRA_paper_99077.pdf]
- Flici, F. (2017). Longevity and pension plan sustainability in Algeria: Taking the re- tirees mortality experience into account. Doctoral dissertation, Higher National School of Statistics and Applied Economics (ENSSEA), Kolea, Algeria.
- Flici, F. (2020). Populations Projections Using R - Including Dynamic graphical Visualisations. A textbook published with Gitbook, Available at: [<https://farid-flici.gitbook.io/pop-proj-dz/>], Version of 2020-04-30.
- Hyndman, R. J., Booth, H., & Yasmeen, F. (2013). Coherent mortality forecasting: the product-ratio method with functional time series models. *Demography*, 50 (1), 261-283.

- Lee, R. D. (1993). Modeling and forecasting the time series of US fertility: Age distribution, range, and ultimate level. *International Journal of Forecasting*, 9(2), 187-202.
- Smith, S. K., Tayman, J., & Swanson, D. A. (2013). *Overview of the Cohort-Component Method*. In *A Practitioner's Guide to State and Local Population Projections* (pp. 45-50). Springer, Dordrecht. DOI: 10.1007/978-94-007-7551-0_3