The User Guide to Pension Simulation Project[1]

(funded by the Steven H. Sandell Grant Program, U.S. Social Security Administration via Boston College Center for Retirement Research)

The purpose of the project is to simulate all possible normal cost and actuarial accrued liability based upon sixteen different parameters and actuarial assumptions. The final products of the projects include the R codes, seven comma-separated values (CSV) files and this user guide, which can be accessed through the following link: https://github.com/sunypension/pensionsimulation

The R codes are created to calculate possible normal cost and actuarial accrued liability at the individual level based upon all possible actuarial assumptions within defined parameters. There are sixteen parameters that can be adjusted in the code. (see the following Table 1)

Table 1. 16 Parameters and Current Values for Setting the Data Set

Parameters	Descriptions	Current Values
1. Entry Age	The starting age of plan participant	25, 30 and 35
2. Retirement Age	The last year of participant included in payroll	55, 60 and 65
3. Current Age	The current year of plan participant	25, 26, … ,84 and 85
4. Minimum Age Limit of the Model	The minimum possible age included in the plan	25
5. Maximum Age Limit of the Model	The maximum possible age included in the plan	85
6. Inflation Rate	Assumed annual inflation rate	0.02
7. Past Salary Growth Rate	The actual annual salary growth rate	0.04, 0.05,0.06 and 0.07
8. Future Salary Growth Rate	The expected annual salary growth rate	0.04, 0.05,0.06 and 0.07
9. Discount Rate	The assumed annual investment return rate of pension fund	0.04, 0.05,0.06,0.07,0.08 and 0.09
10. Benefit Factor	The final factor used in pension benefits calculation is a benefit multiplier	0.01, 0.02, 0.03 and 0.04
11. COLA (cost of living adjustment)	The adjustment made to salary to counteract the effects of inflation.	0, 0.01, 0.02, 0.03 and 0.04
12. AFC (average final compensation)	The numbers of last years' salary to be included for calculation	5
13. Vesting Period	The minimum required years to be qualified for the calculation	5
14. Actuarial Cost Methods	Cost Methods to be entry age normal (EAN) and projected unit calculation(PUC)]	EAN' or 'PUC'
15. Mortality Table Choices	Mortality Table choices to be 2(RP2014), 4(RP2000) and 6(RP 2010)	2, 4 and 6
16. Starting Salary	The starting salary of the first year	$1

1 This pension simulation project is co-led by Dr. David S. T. Matkin and Dr. Gang Chen with the State University of New York at Albany. Check the paper for more detailed findings: Chen, G., & Matkin, D. S. (2017). Actuarial Inputs and the Valuation of Public Pension Liabilities and Contribution Requirements: A Simulation Approach. *Public Budgeting & Finance*, *37*(1), 68-87.

The created R codes include three major files: *Run_Simulation.R*, *Functions.R* and *Generatecsv.R*. (see the following Table 2)

Table 2. The Created Three Codes

File Name	File Description
1. Run_Simulation.R	The main file to create data frame and is currently set to the default value of actuarial assumptions for creating the final data set of this project. The values of sixteen parameters within line 50 to 100 can be adjusted to different requirements;
2. Functions.R	The file includes all related calculation that Run_Simulation.R calls from;
3. Generatecsv.R	The code to create the final CSV files using the data frame created by *Run_Simulation.R*. The current results include six separate data set, based upon the combination of two calculation choices [entry age normal (EAN) and projected unit calculation(PUC)] and three mortality table choices (mortality table 2000, 2010 and 2014), and a combined data set which consists of the six data sets.

Seven CSV files are created using the current values of 16 parameters through the three above R codes. (see the following Table 3)

Table 3. The Created Seven CSV Files (https://tinyurl.com/m7kb3yf)

File Name	Observation Number	File Size (MB)
EANrp2000full	967,680	86.7
EANrp2010full	967,680	86.7
EANrp2014full	967,680	85.7
PUCrp2000full	967,680	86.7
PUCrp2010full	967,680	86.7
PUCrp2014full	967,680	86.7
Full	5,806,080	519

To fully reproduce the result, all files in the folder named "R_Code" are required to be downloaded and unzipped into one newly created folder in addition to the most updated version of R [2]and RStudio [3](preferably coupling R studio with R version 3.3.3. released on Monday 2017-03-06).

In the folder "R_Code", there are three R codes files, one text file and another folder "Data" which includes two R data sets, *Mortality.RData and Termination.RData*.

The R studio is the preferred operating environment and the R codes can be directly used by R studio. The following steps are written for running the simulation in the RStudio environment. The downloaded codes can also be applied in the original R environment through coping all the contents from *R_Code.txt* and pasting to R console after adjusting the default saving route. The step of adjusting saving route can be done by substituting "XXX" part in the first line with the downloaded folder location and then type "enter."

- The Eight Steps to Replicate the Pension Simulation Data Set:

1. Setting up the R simulating environment with appropriate packages: download and install R and R studio with appropriate packages[4] together with the folder from: https://github.com/sunypension/pensionsimulation
   The eight following R Packages in Table 4 are required to be installed:

Table 4. The Eight Packages and Descriptions

R Package	Description
1. BayesBridge	The package to create different kinds of distributions of population;
2. dplyr	The package to manipulate data tasks within the model;
3. foreach	The package to enable for writing loops in R coding;
4. FinCal	The package to do financial calculation, time series analysis and computational finance etc.;
5. magrittr	The package to improve readability and maintainability of code;
6. rJava	The package to allow creation of objects and calling methods;
7. Scales	The package to enable number formatting and visualize data set used in the model;
8. XLConnect	The package to import mortality tables and termination rates as well as export the final csv files.

2. R can be downloaded at https://www.r-project.org/
3. R studio can be downloaded at https://www.rstudio.com/products/rstudio/
4. More details instruction with examples about the installation process can be found in the following: https://www.r-bloggers.com/installing-r-packages/

2. Open the *Run_Simulation.R* using RStduio by double clicking the file;
3. Select all the codes in the Source panel through right licking mouse within the source panel and click “Select All” in the popped up Common Tasks Bar or simply typing “Ctrl+A”;
   The code within line 50 to line 100 are the sixteen parameters are set to the default values and can be adjusted to your needs (see the further notes below);
4. Click “Run” button at the top right corner of the Source panel or simply type “Ctrl+R”;
5. The code will start to run (the speed of this process depends upon the chosen parameters and computing capacity of your computer) to generate the originally uploaded data set and it will generate an R data frame named *simulation_result.RData* in the same folder “R_Code”;
6. Once the step 5 is finished (refer to the console result and it will show the time when the computing is finished), open the Generatecsv.*R* in the folder using RStduio by double clicking the file;
7. Repeat steps 3 and 4 to run the whole coding within Generatecsv.*R*;
8. The codes will import the newly created R data frame *simulation_result.RData* and generate the final seven CSV files in the folder (see the further notes below). The speed of this process also depends on the computing capacity of your computer.

- The Four Steps to Produce Pension Data with Different Parameters:

  There are four major different steps involved to produce pension data with different parameters:

1. Repeat the previous steps 1 and 2 to set up the R simulation environment and files;
2. Follow the previous step 3, adjust the16 parameters within the lines from 50 to 100 of *Run_Simulation.*R. See he following Table 5 for more details;
3. Repeat the previous steps 4 to 6 to produce an R data frame *simulation_result.RData* with adjusted 16 parameters;
4. Follow the previous step 7 and 8, you can adjust the lines from 5 to 19 based upon your previous choices cost methods and mortality table within *Generatecsv.R.* For example, If PUC is deleted, then line 8, 9, 10 should be deleted together with lines 15, 16, 17. Meanwhile, if mortality table 2010 is deleted, then line 6 and line 9 together with line 13 and line 16 should also be removed.

Table 5. The Codes of 16 Parameters and Explanations

Location	Code	Explanation
1. Line 51	e_a<-seq(25,35,5)	this defines the range of entry age to be 25 to 35 with 5 incremental increase within the chosen range, which is 25, 30, 35. You can adjust the range and the rate of incremental change;
2. Line 54	r_a<-seq(55,65,5)	this defines the range of retirement age to be 55 to 65 with 5 incremental increase within the chosen range, which is 55, 60, 65. You can adjust the range and the rate of incremental change;
3. Line 57	ca<-seq(25,85,1)	this defines the range of current age to be 25 to 85 with 1 incremental increase within the chosen range, which is 25, 26, …,84,85. You can adjust the range and the rate of incremental change;
4. Line 60	min_age<-25	this defines the minimum age for the group (including active employees and retirees) to be 25 in this model and you can adjust the value to your needs;
5. Line 63	max_age<-85	this defines the maximum age for the group (including active employees and retirees) to be 85 in this model and you can adjust the value to your needs;
6. Line 66	inflation<-0.02	this defines annual inflation rate to be 0.02 or 2 percent in this model and you can adjust the value to your needs;
7. Line 69	past_sgr<-seq(0.02,0.05,0.01)+inflation	this defines the range of salary growth rate before current age to be 0.02 to 0.05 with 0.01 incremental increase within the chosen range and the inflation rate 0.02, which makes the value to be 0.04, 0.05,0.06 and 0.07. You can adjust the range and the rate of incremental change;
8. Line 72	future_sgr<-seq(0.02,0.05,0.01)+inflation	this defines the future expected salary growth rate to be 0.02 to 0.05 with 0.01 incremental increase within the chosen range and the inflation rate 0.02, which makes the value to be 0.04, 0.05,0.06 and 0.07. You can adjust the range and the rate of incremental change;
9. Line 75	i_r<-seq(0.02,0.07,0.01)+inflation	this defines the discount rate to be 0.02 to 0.07 with 0.01 incremental increase within the chosen range and the inflation rate 0.02, which makes the value to be 0.04, 0.05,0.06,0.07,0.08 and 0.09. You can adjust the range and the rate of incremental change;
10. Line 78	bf_p<-seq(0.01,0.04,0.01)	this defines the benefit factor to be 0.01 to 0.04 with 0.01 incremental increase within the chosen range, which makes the value to be 0.01, 0.02, 0.03 and 0.04. You can adjust the range and the rate of incremental change;
11. Line 81	cola_r<-seq(0,0.04,0.01)	this defines the COLA to be 0 to 0.04 with 0.01 incremental increase within the chosen range, which makes the value to be 0, 0.01, 0.02, 0.03 and 0.04. You can adjust the range and the rate of incremental change;
12. Line 84	afc_p<-5	this defines the AFC (average final compensation, which is the number of years to calculate pension benefit) to be 5 years. You can adjust the value to your needs;
13. Line 87	vesting<-5	this defines the Vesting Period to be 5 years. You can adjust the value to your needs;
14. Line 90	c_m<-c('EAN','PUC')	this defines the Actuarial Cost Methods to be entry age normal (EAN) and projected unit calculation(PUC)]. You can keep both or only one of it;
15. Line 93	mort_num<-c(2,4,6)	this defines the Mortality Table choices to be 2(RP2014), 4(RP2000) and 6(RP 2010) You can keep any one or two or all of them;
16. Line 96	sal<-1	this defines the starting salary to be $1 and You can adjust it your own numbers.