Mortality Projection

Japan

Mahadev Suresh

2024-06-01

Table of Contents

# Pre-processing of data

## Read Data

Exposure data & death has been downloaded from **Human Mortality Database** <https://mortality.org/>.

exposure <- fread(paste0(path, "Mortality Data - Japan/", "Exposures\_1x1.txt"))  
death <- fread(paste0(path, "Mortality Data - Japan/", "Deaths\_1x1.txt")) %>%   
 mutate(Female = as.integer(Female), Male = as.integer(Male)) %>%  
 as.data.table()  
  
kableExtra::kable(head(exposure))

| Year | Age | Female | Male | Total |
| --- | --- | --- | --- | --- |
| 1947 | 0 | 1122695.6 | 1170640.0 | 2293336 |
| 1947 | 1 | 777087.8 | 800958.6 | 1578046 |
| 1947 | 2 | 850243.1 | 869133.1 | 1719376 |
| 1947 | 3 | 985176.3 | 1006140.7 | 1991317 |
| 1947 | 4 | 953958.8 | 976684.0 | 1930643 |
| 1947 | 5 | 995604.1 | 1020042.5 | 2015647 |

kableExtra::kable(head(death))

| Year | Age | Female | Male | Total |
| --- | --- | --- | --- | --- |
| 1947 | 0 | 93923 | 111818 | 205741.71 |
| 1947 | 1 | 27719 | 29671 | 57391.07 |
| 1947 | 2 | 14696 | 14962 | 29658.70 |
| 1947 | 3 | 11068 | 11408 | 22477.48 |
| 1947 | 4 | 6704 | 7299 | 14003.90 |
| 1947 | 5 | 4812 | 5216 | 10028.53 |

## Reshaping Data

Converting Data in wide format to long format.

## Warning in melt.data.table(setDT(death), id.vars = c("Age", "Year"),  
## variable.name = "Gender", : 'measure.vars' [Female, Male, Total, ...] are not  
## all of the same type. By order of hierarchy, the molten data value column will  
## be of type 'double'. All measure variables not of type 'double' will be coerced  
## too. Check DETAILS in ?melt.data.table for more on coercion.

## Warning: There was 1 warning in `mutate()`.  
## ℹ In argument: `Age1 = if\_else(Age == "110+", 110, as.double(Age))`.  
## Caused by warning in `if\_else()`:  
## ! NAs introduced by coercion

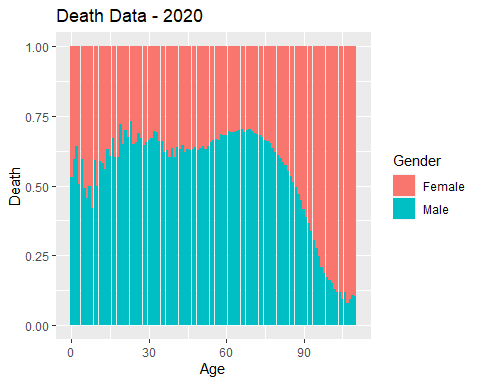
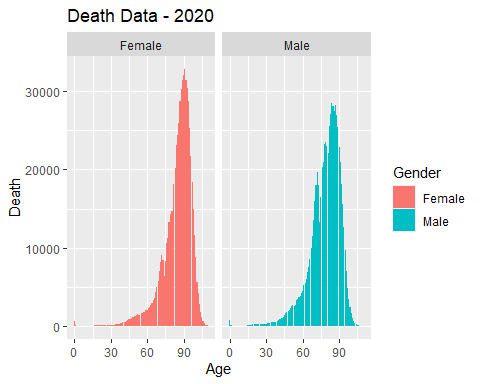
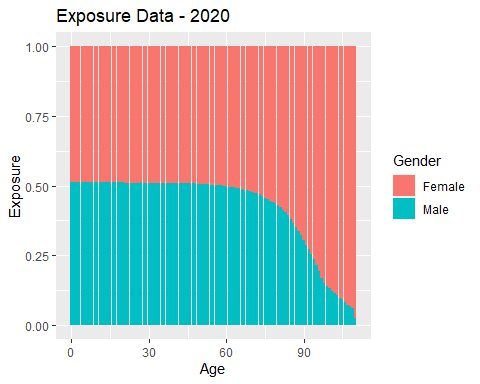
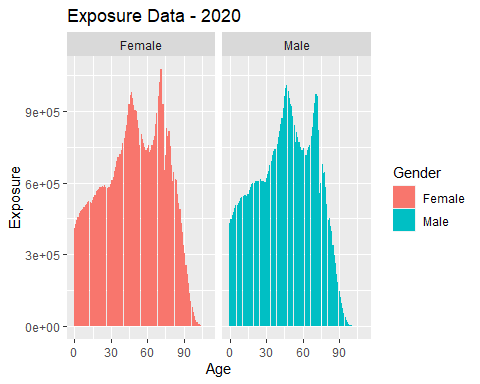
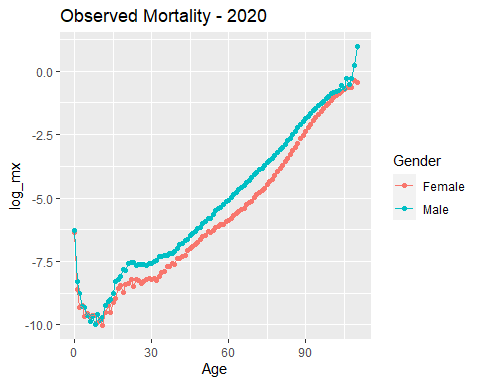
| Age | Year | Gender | Exposure |
| --- | --- | --- | --- |
| 0 | 1947 | Female | 1122695.6 |
| 1 | 1947 | Female | 777087.8 |
| 2 | 1947 | Female | 850243.1 |
| 3 | 1947 | Female | 985176.3 |
| 4 | 1947 | Female | 953958.8 |
| 5 | 1947 | Female | 995604.1 |

| Age | Year | Gender | Death |
| --- | --- | --- | --- |
| 0 | 1947 | Female | 93923 |
| 1 | 1947 | Female | 27719 |
| 2 | 1947 | Female | 14696 |
| 3 | 1947 | Female | 11068 |
| 4 | 1947 | Female | 6704 |
| 5 | 1947 | Female | 4812 |

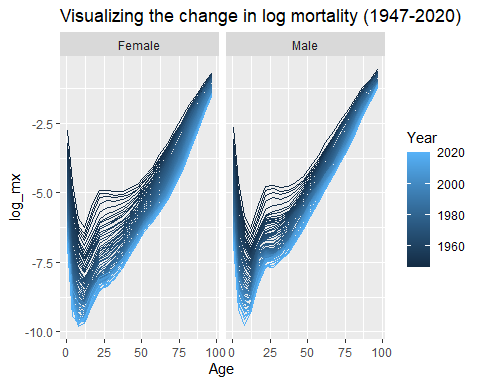
| Age | Year | Gender | Exposure | Death | mx |
| --- | --- | --- | --- | --- | --- |
| 0 | 1947 | Female | 1122695.6 | 93923 | 0.0836585 |
| 1 | 1947 | Female | 777087.8 | 27719 | 0.0356704 |
| 2 | 1947 | Female | 850243.1 | 14696 | 0.0172845 |
| 3 | 1947 | Female | 985176.3 | 11068 | 0.0112345 |
| 4 | 1947 | Female | 953958.8 | 6704 | 0.0070276 |
| 5 | 1947 | Female | 995604.1 | 4812 | 0.0048332 |

## `summarise()` has grouped output by 'AgeBand', 'Gender'. You can override using  
## the `.groups` argument.

## Visulaizing the pre-COVID mortality



## Mortality Curve over the years.



## Data Reshaping for Modelling

male\_mortality <- combined %>%   
 filter(Gender == "Male", Year >= 1980) %>%  
 pivot\_wider(names\_from = Year, values\_from = mx, id\_cols = Age) %>%   
 select(-Age) %>%   
 as.matrix()  
  
female\_mortality <- combined %>%   
 filter(Gender == "Female", Year >= 1980) %>%  
 pivot\_wider(names\_from = Year, values\_from = mx, id\_cols = Age) %>%   
 select(-Age) %>%   
 as.matrix()  
  
male\_exposure <- combined %>%   
 filter(Gender == "Male", Year >= 1980) %>%  
 pivot\_wider(names\_from = Year, values\_from = Exposure, id\_cols = Age) %>%   
 select(-Age) %>%   
 as.matrix()  
  
female\_exposure <- combined %>%   
 filter(Gender == "Female", Year >= 1980) %>%  
 pivot\_wider(names\_from = Year, values\_from = Exposure, id\_cols = Age) %>%   
 select(-Age) %>%   
 as.matrix()  
  
rows <- unique(combined$Age)  
cols <- as.numeric(colnames(male\_mortality))  
male\_demog <- demography::demogdata(male\_mortality, male\_exposure, years = cols, ages = rows, name = "male", "Japan - HMD", type = "mortality")

## Warning in .recacheSubclasses(def@className, def, env): undefined subclass  
## "ndiMatrix" of class "replValueSp"; definition not updated

## Registered S3 methods overwritten by 'demography':  
## method from   
## print.lca e1071  
## summary.lca e1071

female\_demog <- demography::demogdata(female\_mortality, female\_exposure, years = cols, ages = rows, name = "female", "Japan - HMD", type = "mortality")  
male\_demog\_stomomo <- StMoMoData(male\_demog)  
female\_demog\_stomomo <- StMoMoData(female\_demog)

# Lee Carter Model

## Model Fitting

LC\_log <- lc(link = "log")  
   
LCfit\_male <- fit(LC\_log, data = male\_demog\_stomomo)

## Warning in fit.StMoMo(LC\_log, data = male\_demog\_stomomo): StMoMo: 11 data points have   
## non-positive exposures and have been zero weighted

## Warning in fit.StMoMo(LC\_log, data = male\_demog\_stomomo): StMoMo: 11 missing values which have been zero weighted

## StMoMo: Start fitting with gnm  
## Initialising  
## Running start-up iterations..  
## Running main iterations......  
## Done  
## StMoMo: Finish fitting with gnm

LCfit\_female <- fit(LC\_log, data = female\_demog\_stomomo)

## Warning in fit.StMoMo(LC\_log, data = female\_demog\_stomomo): StMoMo: 2 data points have   
## non-positive exposures and have been zero weighted

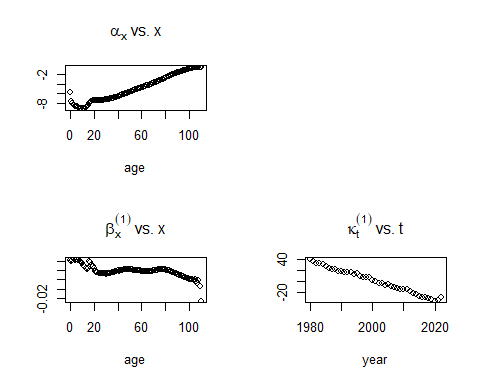
## Warning in fit.StMoMo(LC\_log, data = female\_demog\_stomomo): StMoMo: 2 missing values which have been zero weighted

## StMoMo: The following cohorts have been zero weigthed: 1870   
## StMoMo: Start fitting with gnm  
## Initialising  
## Running start-up iterations..  
## Running main iterations......  
## Done  
## StMoMo: Finish fitting with gnm

#ax is similar to the mortality rates observed in the past.   
#bx is smooth for most part except for 10-14, the improvement for 20-40 is as low as people in 80s. Interpolation maybe required.   
#kt seems alright, but I suspect the forecast will overstate the improvements.   
plot(LCfit\_male, type = "point")

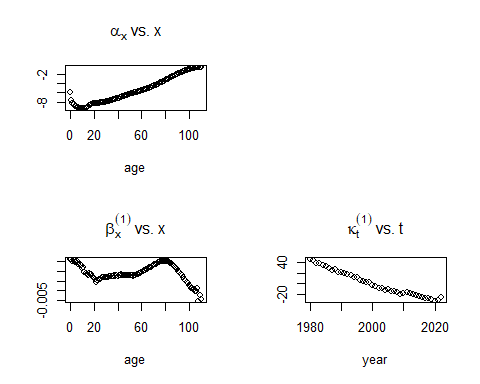
## Warning in plot.xy(xy, type, ...): plot type 'point' will be truncated to first  
## character

## Warning in plot.xy(xy, type, ...): plot type 'point' will be truncated to first  
## character  
  
## Warning in plot.xy(xy, type, ...): plot type 'point' will be truncated to first  
## character



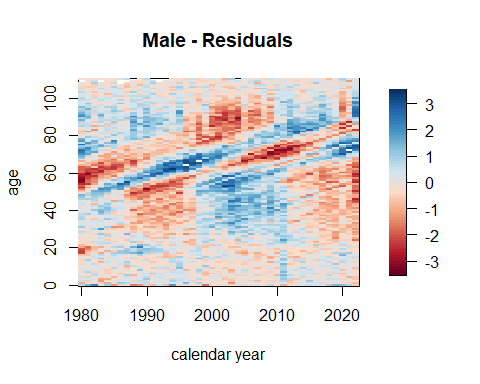
plot(LCfit\_female, type = "point")

## Warning in plot.xy(xy, type, ...): plot type 'point' will be truncated to first  
## character  
  
## Warning in plot.xy(xy, type, ...): plot type 'point' will be truncated to first  
## character  
  
## Warning in plot.xy(xy, type, ...): plot type 'point' will be truncated to first  
## character

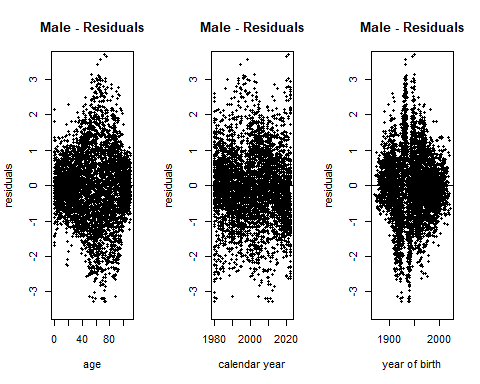


## Goodness of fit

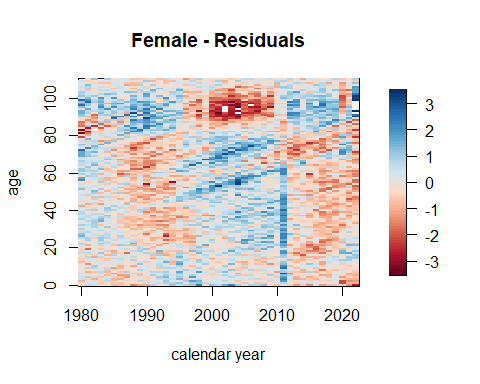
LCres\_male <- residuals(LCfit\_male)  
LCres\_female <- residuals(LCfit\_female)  
plot(LCres\_male, type = "colourmap", reslim = c(-3.5, 3.5), main = "Male - Residuals")



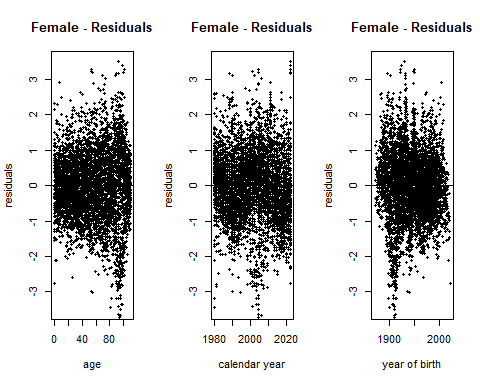
# The fit overestimates mortality for ages around 60.   
plot(LCres\_male, type = "scatter", reslim = c(-3.5, 3.5), main = "Male - Residuals")



# Deviation observed for ages between 60-70.   
  
plot(LCres\_female, type = "colourmap", reslim = c(-3.5, 3.5), main = "Female - Residuals")



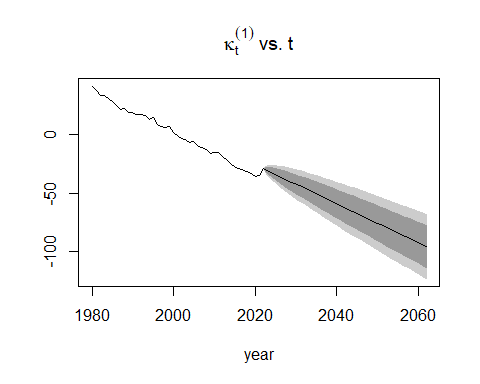
plot(LCres\_female, type = "scatter", reslim = c(-3.5, 3.5), main = "Female - Residuals")



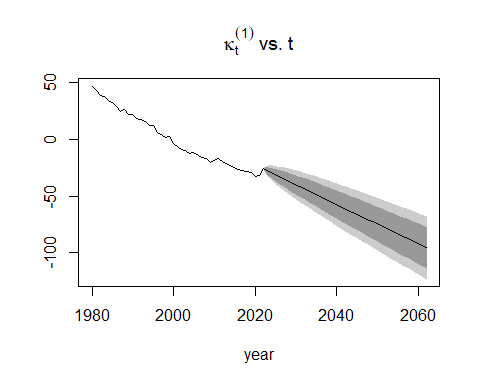
# Female residuals are better (closer to zero, lower deviations) on average compared to the males. LC model fit better for females.   
# Cohort effect is visible for both, but for males its seems to be higher.

## Forecast mortality

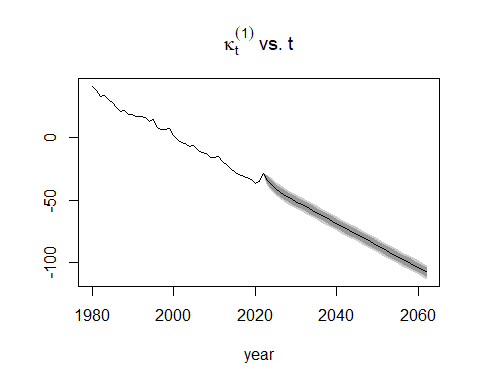
# Plot the forecast kt.   
LCfor\_male <- forecast(LCfit\_male, h = 40)  
plot(LCfor\_male, only.kt = TRUE)



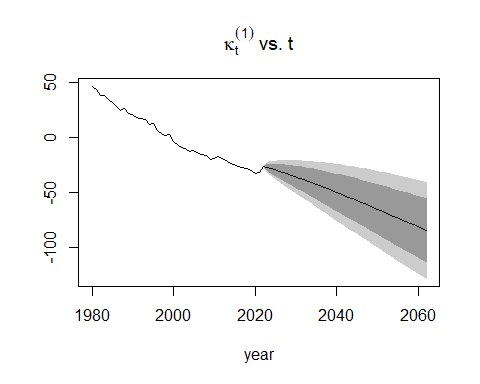
LCfor\_female <- forecast(LCfit\_female, h = 40)  
plot(LCfor\_female, only.kt = TRUE)



LCforArima\_male <- forecast(LCfit\_male, h = 40, kt.method = "iarima", kt.order = c(1, 1, 1))  
plot(LCforArima\_male, only.kt = TRUE)

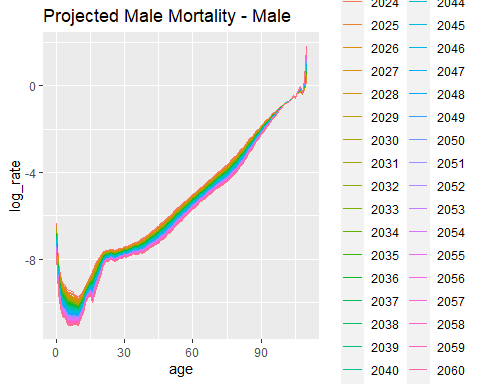


LCforArima\_female <- forecast(LCfit\_female, h = 40, kt.method = "iarima", kt.order = c(1, 1, 1))  
plot(LCforArima\_female, only.kt = TRUE)

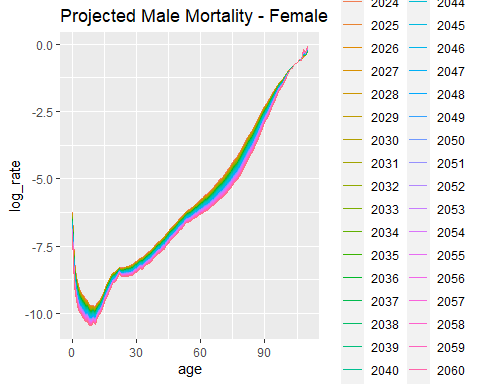


## Visualizing the projections - Lee Carter

forcast\_male <- LCfor\_male$rates %>%   
 as.data.table() %>%   
 mutate(age = rownames(LCfor\_male$rates), gender = "Male") %>%   
 pivot\_longer(cols = colnames(LCfor\_male$rates), values\_to = "rates", names\_to = "year")  
  
forcast\_female <- LCfor\_female$rates %>%   
 as.data.table() %>%   
 mutate(age = rownames(LCfor\_male$rates), gender = "Female") %>%   
 pivot\_longer(cols = colnames(LCfor\_male$rates), values\_to = "rates", names\_to = "year")  
  
forcast\_arima\_male <- LCforArima\_male$rates %>%   
 as.data.table() %>%   
 mutate(age = as.integer(rownames(LCforArima\_male$rates)), gender = "Male") %>%   
 pivot\_longer(cols = colnames(LCforArima\_male$rates), values\_to = "rates", names\_to = "year") %>%   
 mutate(log\_rate = log(rates)) %>%   
 arrange(year, age)  
  
forcast\_arima\_female <- LCforArima\_female$rates %>%   
 as.data.table() %>%   
 mutate(age = as.integer(rownames(LCforArima\_female$rates)), gender = "Female") %>%   
 pivot\_longer(cols = colnames(LCforArima\_female$rates), values\_to = "rates", names\_to = "year") %>%  
 mutate(log\_rate = log(rates)) %>%   
 arrange(year, age)   
  
ggplot(forcast\_arima\_male, aes(age, log\_rate, color = year, group = year)) +   
 geom\_line() +  
 ggtitle("Projected Male Mortality - Male")

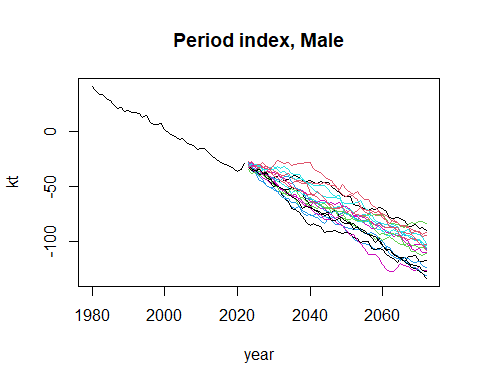


ggplot(forcast\_arima\_female, aes(age, log\_rate, color = year, group = year)) +   
 geom\_line() +   
 ggtitle("Projected Male Mortality - Female")

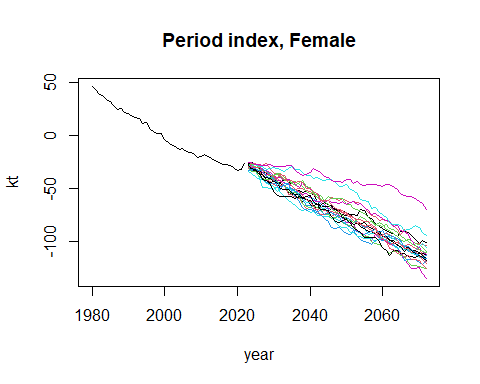


## Simulation

LCsim\_male <- simulate(LCfit\_male, nsim = 500, h = 50)  
  
plot(LCfit\_male$years, LCfit\_male$kt[1, ], xlim = range(LCfit\_male$years, LCsim\_male$years),   
 ylim = range(LCfit\_male$kt, LCsim\_male$kt.s$sim[1, , 1:20]),   
 type = "l", xlab = "year", ylab = "kt", main = "Period index, Male")  
  
matlines(LCsim\_male$kt.s$years, LCsim\_male$kt.s$sim[1, , 1:20], type = "l", lty = 1)



LCsim\_female <- simulate(LCfit\_female, nsim = 500, h = 50)  
  
plot(LCfit\_female$years, LCfit\_female$kt[1, ], xlim = range(LCfit\_female$years, LCsim\_female$years),   
 ylim = range(LCfit\_female$kt, LCsim\_female$kt.s$sim[1, , 1:20]),   
 type = "l", xlab = "year", ylab = "kt", main = "Period index, Female")  
  
matlines(LCsim\_female$kt.s$years, LCsim\_female$kt.s$sim[1, , 1:20], type = "l", lty = 1)



# Age-Period-Cohort

## Fitting APC Model

### Key difference is no b\_x instead we have an additive cohort factor.

wxt\_male <- genWeightMat(male\_demog\_stomomo$ages, male\_demog\_stomomo$years,   
 clip = 3)  
wxt\_female <- genWeightMat(female\_demog\_stomomo$ages, female\_demog\_stomomo$years,  
 clip = 3)  
  
male\_demog\_stomomo\_init <- central2initial(male\_demog\_stomomo)  
  
APCfit\_male <- fit(apc(link = "log"), data = male\_demog\_stomomo, wxt = wxt\_male)

## Warning in fit.StMoMo(apc(link = "log"), data = male\_demog\_stomomo, wxt = wxt\_male): StMoMo: 11 data points have   
## non-positive exposures and have been zero weighted

## Warning in fit.StMoMo(apc(link = "log"), data = male\_demog\_stomomo, wxt = wxt\_male): StMoMo: 11 missing values which have been zero weighted

## StMoMo: The following cohorts have been zero weigthed: 1870 1871 1872 2020 2021 2022   
## StMoMo: Start fitting with gnm  
## StMoMo: Finish fitting with gnm

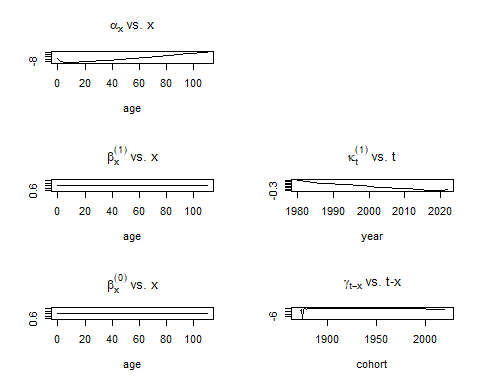
APCfit\_female <- fit(apc(link = "log"), data = female\_demog\_stomomo, wxt = wxt\_female)

## Warning in fit.StMoMo(apc(link = "log"), data = female\_demog\_stomomo, wxt = wxt\_female): StMoMo: 2 data points have   
## non-positive exposures and have been zero weighted

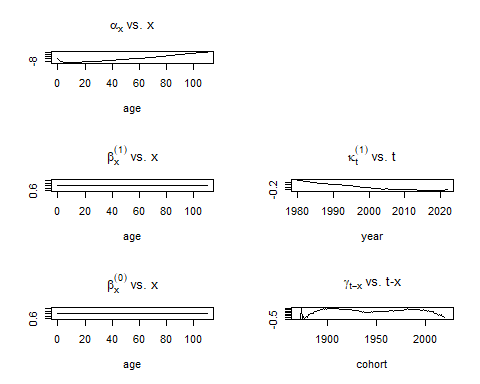
## Warning in fit.StMoMo(apc(link = "log"), data = female\_demog\_stomomo, wxt = wxt\_female): StMoMo: 2 missing values which have been zero weighted

## StMoMo: The following cohorts have been zero weigthed: 1870 1871 1872 2020 2021 2022   
## StMoMo: Start fitting with gnm  
## StMoMo: Finish fitting with gnm

plot(APCfit\_male)

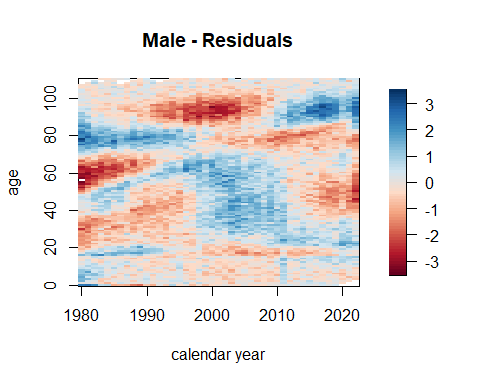


plot(APCfit\_female)

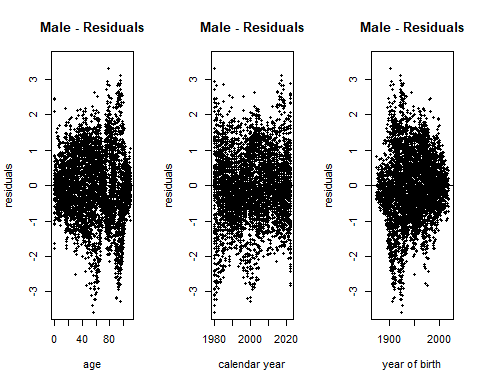


## Analysis of Goodness of Fit

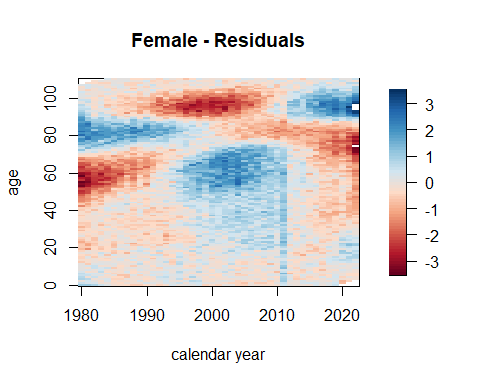
APCres\_male <- residuals(APCfit\_male)  
APCres\_female <- residuals(APCfit\_female)  
  
plot(APCres\_male, type = "colourmap", reslim = c(-3.5, 3.5), main = "Male - Residuals")



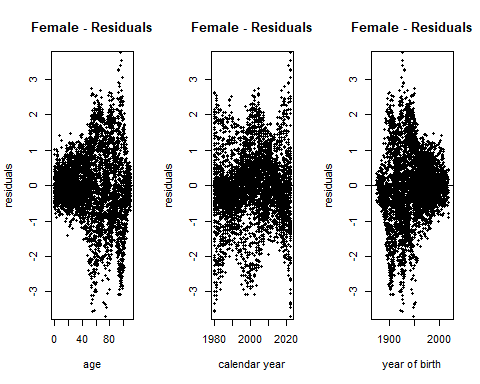
# The fit overestimates mortality for ages around 60.   
plot(APCres\_male, type = "scatter", reslim = c(-3.5, 3.5), main = "Male - Residuals")



plot(APCres\_female, type = "colourmap", reslim = c(-3.5, 3.5), main = "Female - Residuals")



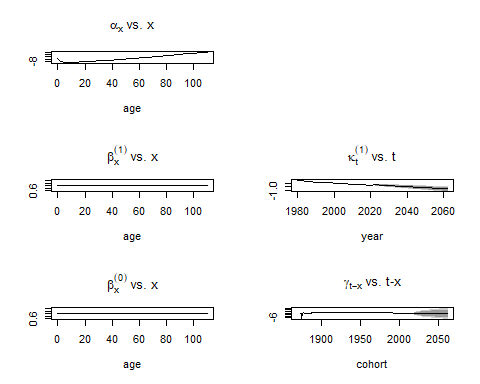
# The fit overestimates mortality for ages around 60.   
plot(APCres\_female, type = "scatter", reslim = c(-3.5, 3.5), main = "Female - Residuals")



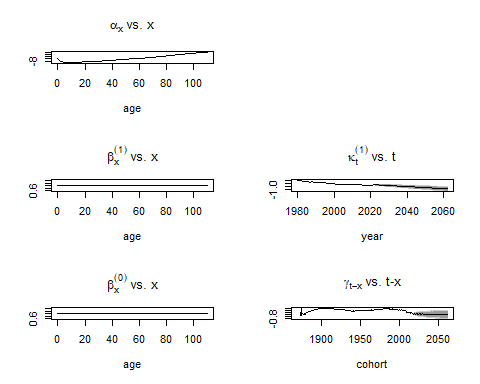
# Much poorer fit than Lee-Carter for Males & Females

## Forecast mortality - APC

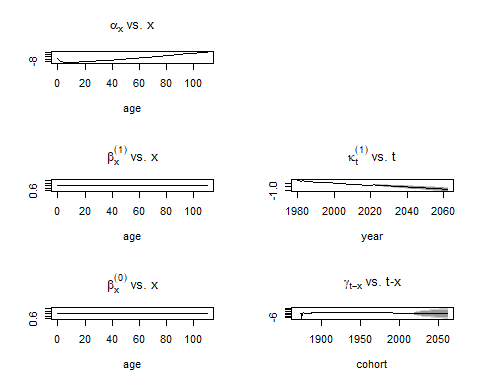
# Plot the forecast kt.   
APCfor\_male <- forecast(APCfit\_male, h = 40)  
plot(APCfor\_male)



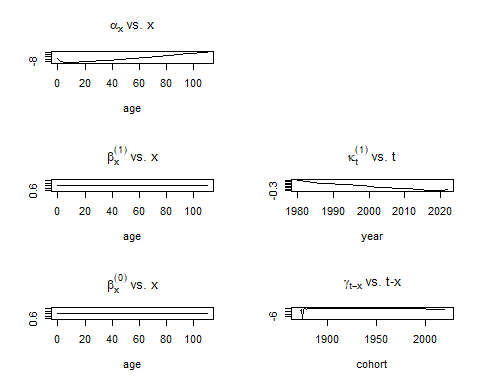
APCfor\_female <- forecast(APCfit\_female, h = 40)  
plot(APCfor\_female)



APCforArima\_male <- forecast(APCfit\_male, h = 40, kt.method = "iarima", kt.order = c(1, 1, 1))  
plot(APCforArima\_male)

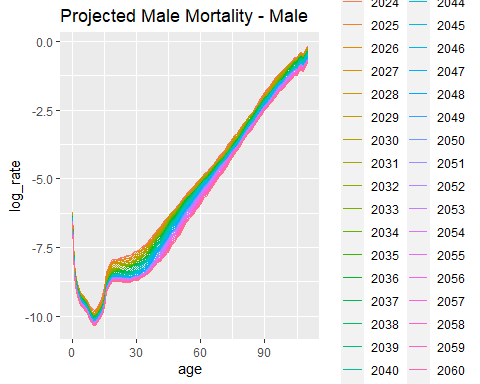


APCforArima\_female <- forecast(APCfit\_female, h = 40, kt.method = "iarima", kt.order = c(1, 1, 1))  
plot(APCfit\_male)



## Visualizing the projections - APC

forcast\_male <- APCfor\_male$rates %>%   
 as.data.table() %>%   
 mutate(age = as.integer(rownames(APCfor\_male$rates)), gender = "Male") %>%   
 pivot\_longer(cols = colnames(APCfor\_male$rates), values\_to = "rates", names\_to = "year") %>%   
 mutate(log\_rate = log(rates)) %>%   
 arrange(year, age)   
  
forcast\_female <- APCfor\_female$rates %>%   
 as.data.table() %>%   
 mutate(age = as.integer(rownames(APCfor\_female$rates)), gender = "Female") %>%   
 pivot\_longer(cols = colnames(APCfor\_female$rates), values\_to = "rates", names\_to = "year") %>%   
 mutate(log\_rate = log(rates)) %>%   
 arrange(year, age)   
  
ggplot(forcast\_male, aes(age, log\_rate, color = year, group = year)) +   
 geom\_line() +  
 ggtitle("Projected Male Mortality - Male")



ggplot(forcast\_female, aes(age, log\_rate, color = year, group = year)) +   
 geom\_line() +   
 ggtitle("Projected Male Mortality - Female")

