Scotland Mortality Projections

Jon Minton

# Introduction

This document will present life expectancy and age specific mortality observed in Scotland since 2012 in comparison with projections based on earlier trends.

Two periods will be considered when informing the pre-2012 trends:

* 1950-2011
* 1990-2011

## Background literature: why ARIMA(0,1,0) is OK

[Lee (2016?)](https://link.springer.com/chapter/10.1007%2F978-3-030-05075-7_14), who co-created the [Lee-Carter model](https://www.jstor.org/stable/2290201?seq=1#page_scan_tab_contents) for mortality projection in the early 1990s, accepted in the mid/late 2000s, as a result of analyses by [White (2002)](https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1728-4457.2002.00059.x) and [Oeppen & Vaupel (2002)](http://truemedmd.com/wp-content/uploads/2013/07/Life_expectancy_scienceMay2002.pdf), that a much simpler modelling approach is often adequate for life expectancy forecasting. This approach involves modelling life expectancy at birth () directly, rather than individual age specific mortality rates () and quantities derived from them (a ‘drift parameter’ ) as in Lee-Carter’s original model specification.

Although [Lee (2002)](https://link.springer.com/chapter/10.1007%2F978-3-030-05075-7_14) presents a number of refinements of a basic linear life expectancy forecase model specification, including models which allow for negative autocorrelation between change in years, the article states that the basic strategy for forecasting should be “to use the appropriate or preferred equation for [change in life expectancy] to estimate one year later, and then continue recursively.” This is the approach this projection exercise will follow

## Modelling strategy

We have , the life expectancy at birth for sex and time , extracted from the Human Mortality Database (HMD) from 1855 to 2016, and from the ONS for 2017 and 2018 (by males and females only, not total population). We define as change in life expectancy for sex from year to year .

We first plot both against by , and against by , to identify appropriate pre-slowdown time periods, i.e. the range of years prior to 2012 (our assumed breakpoint in mortality trends) to use to calibrate projections of life expectancy trends from 2012 to 2018 onwards. Based on these plots and background knowledge, we selected two pre-slowdown periods:

* : 1950-2011 inclusive
* : 1990-2011 inclusive

For both of these pre-slowdown periods, we calculated the mean annual change observed by sex (i.e. ) and the variance in mean annual changes over the same period (i.e. ).

To project forward estimates of life expectancy trends based on pre-slowdown observations in a way that takes into account observed variation in annual changes, we sample and accumulate draws from Normal distributions calibrated on the sample means and variances from the pre-slowdown period ( or ), where refers to the number of annual periods we want to project forward from 2011. As we want to project forward to 2018 we therefore select . This exercise is repeated times to allow credible intervals in forecasts to be calculated using a Monte-Carlo approach. Life expectancy projections for each year 2012-2018 inclusive are produced by summing up the draws from the first projected period onwards; the life expectancy from 2011 (the last observation in the pre-slowdown period) is added to these summed values to produce predicted life expectancies rather than cumulative gains or losses in life expectancy since 2011. Algebraically this can be expressed as follows:

Where refers to 2011, the discrete number of periods to forecast after , and the series of annual changes over the pre-slowdown period .

As for each sex and number of discrete projection periods (from 1 to 7 periods), replicates have been produced, the observed life expectancy can be compared against the Monte-Carlo distribution of corresponding projections for the same period . This comparison is shown both graphically, using shaded regions to indicate the 90% and 95% credible intervals from the projected distributions, and by counting up the proportion of the projected values which exceed the observed values for each post 2011 period . For both pre-trend periods (1950-2011) and (1990-2011) the same random number seed was used. All data preparation and analysis were performed using the [R programming language](https://cran.r-project.org/).

Load packages required

pacman::p\_load(HMDHFDplus, tidyverse, broom, readxl)

# Data

The data used are described below. The code will not be run/evaluated as the data have already been downloaded.

# Single year life tables for Scotland available here:  
  
# https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/singleyearlifetablesuk1980to2018  
  
# Downloaded to data folder (see singleyearlifetablesscland.xls)  
  
# load e0 for scotland  
# dta\_scot <- HMDHFDplus::readHMDweb("GBR\_SCO", item = "E0per", username = userInput(), password = userInput())  
# dta\_scot  
  
# load mx for scotland  
# dta\_scot\_mx <- HMDHFDplus::readHMDweb("GBR\_SCO", item = "Mx\_1x1", username = userInput(), password = userInput())  
# HMDHFDplus::getHMDitemavail("GBR\_SCO", username = userInput(), password = userInput())  
  
# bring to tidy format  
# dta\_scot\_tidy <-   
# dta\_scot %>%   
# as\_tibble() %>%   
# rename\_all(tolower) %>%   
# gather(-year, key = "sex", value = "e0")  
  
# write e0 to directory  
# write\_csv(dta\_scot\_tidy, "data/e0\_scot.csv")

The data will now be loaded locally, and augmented with the last couple of years’ e0

dta\_scot\_tidy <- read\_csv("data/e0\_scot.csv")

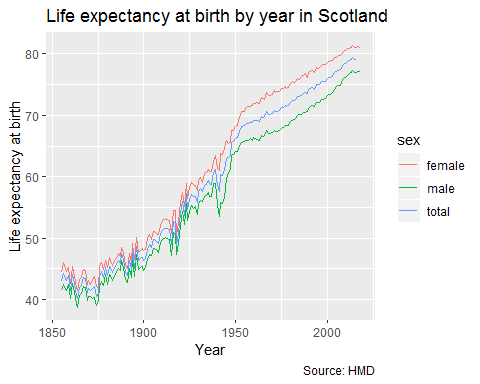
## Parsed with column specification:  
## cols(  
## year = col\_double(),  
## sex = col\_character(),  
## e0 = col\_double()  
## )

# manually adding last two years from table listed above  
dta\_scot\_tidy <-   
 dta\_scot\_tidy %>%   
 bind\_rows(  
 tribble(  
 ~year, ~sex, ~e0,  
 2017, "male", 77.13,  
 2017, "female", 81.16,  
 2018, "male", 77.05,  
 2018, "female", 81.01  
 )  
 )  
# n.b. does not include total

# Visualise e0 at birth

The following chunk shows life expectancy at birth over all available years

dta\_scot\_tidy %>%   
 ggplot(aes(x = year, y = e0, group = sex, colour = sex)) +   
 geom\_line() +   
 labs(  
 x = "Year",   
 y = "Life expectancy at birth",   
 title = "Life expectancy at birth by year in Scotland",   
 caption = "Source: HMD"  
 )



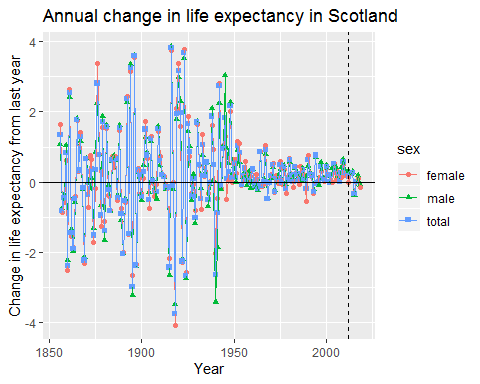
ggsave("figures/scot\_e0.png", height = 15, width = 25, dpi = 300, units = "cm")

Another, less familiar, way of visualising this data is in annual changes in e0. This will be more useful for informing our model

dta\_scot\_tidy %>%   
 group\_by(sex) %>%   
 arrange(year) %>%   
 mutate(delta\_e0 = e0 - lag(e0)) %>%  
 ungroup () %>%   
 ggplot(aes(x = year, y = delta\_e0, group = sex, colour = sex, shape = sex)) +   
 geom\_point() + geom\_line() +  
 labs(  
 x = "Year", y = "Change in life expectancy from last year",   
 title = "Annual change in life expectancy in Scotland"  
 ) +   
 geom\_hline(yintercept = 0) +  
 geom\_vline(xintercept = 2012, linetype = "dashed")

## Warning: Removed 3 rows containing missing values (geom\_point).

## Warning: Removed 3 rows containing missing values (geom\_path).



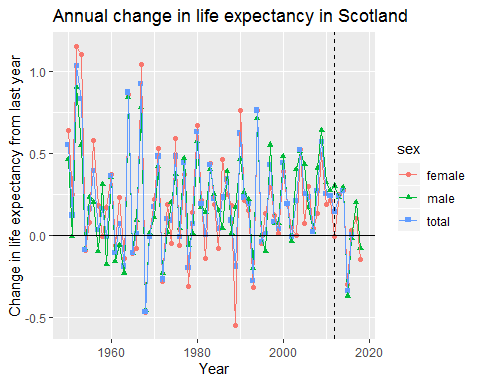
ggsave("figures/scot\_e0\_annual\_change.png", height = 15, width= 25, dpi = 300, units = "cm")

## Warning: Removed 3 rows containing missing values (geom\_point).  
  
## Warning: Removed 3 rows containing missing values (geom\_path).

The reason for not using data from before 1950 is clear from this. Before 1950 there was much more variation in life expectancy than there has been since.

Let’s look at the trend and life expectancy from 1950 onwards

dta\_scot\_tidy %>%   
 group\_by(sex) %>%   
 arrange(year) %>%   
 mutate(delta\_e0 = e0 - lag(e0)) %>%  
 ungroup() %>%   
 filter(year >= 1950) %>%   
 ggplot(aes(x = year, y = delta\_e0, group = sex, colour = sex, shape = sex)) +   
 geom\_point() + geom\_line() +  
 labs(  
 x = "Year", y = "Change in life expectancy from last year",   
 title = "Annual change in life expectancy in Scotland"  
 ) +   
 geom\_hline(yintercept = 0) +  
 geom\_vline(xintercept = 2012, linetype = "dashed")



ggsave("figures/scot\_e0\_annual\_change\_since\_1950.png", height = 15, width = 25, dpi = 300, units = "cm")

We can see there are similar, but not identical, patterns observed for both sexes. Previous research has demonstrated one year negative autocorrelation in the series, i.e. that ‘good’ years tend to precede ‘bad’ years and vice versa. [White (2002)](https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1728-4457.2002.00059.x) noted that, for changes in life expectancy in high income countries as a whole, there was a “small but significant negative relationship between change in average life expectancy in one year and change in the next”, and that:

The negative relationship between change in from one year to the next may represent a tendency for misestimates or random shocks to life expectancy to be corrected in the subsequent year. For example, a particularly virulant flu epidemic might cause a low rise in life expectancy that year. But if deaths are raised or overestimated in one year, a return to normal levels in the next would tend to cause a low rise in life expectancy to be followed by a high one.

[Lee (2016?)](https://link.springer.com/chapter/10.1007%2F978-3-030-05075-7_14) also suggested incoroporating 1 year lagged autocorrelation terms into population projections. This is something we could do but I don’t think will be necessary for our purposes.

The code below (not evaluated) tidies/arranges mortality risks by indivdiual ages and saves into a tidied file available locally

# move mx to tidy format  
mx\_scot\_tidy <-  
 dta\_scot\_mx %>%  
 as\_tibble() %>%  
 rename\_all(tolower) %>%  
 select(-openinterval) %>%  
 gather(-year, -age, key ="sex", value = "mx")  
  
# write mx to directory  
write\_csv(mx\_scot\_tidy, "data/mx\_scot.csv")

## Projecting based on (1950-2011)

The followign code calculates the average improvement rate and variation in improvement rates over the period 1950-2011

e0\_ch\_summary <-   
 dta\_scot\_tidy %>%   
 group\_by(sex) %>%   
 arrange(year) %>%   
 mutate(delta\_e0 = e0 - lag(e0)) %>%   
 ungroup() %>%   
 filter(between(year, 1950, 2011)) %>%   
 group\_by(sex) %>%   
 summarise(  
 mean\_de0 = mean(delta\_e0),  
 sd\_de0 = sd(delta\_e0)  
 ) %>%   
 ungroup()  
  
e0\_ch\_summary\_p1 <- e0\_ch\_summary  
  
e0\_ch\_summary

## # A tibble: 3 x 3  
## sex mean\_de0 sd\_de0  
## <chr> <dbl> <dbl>  
## 1 female 0.214 0.353  
## 2 male 0.207 0.285  
## 3 total 0.211 0.307

On average, between 1950 and 2011, mortality rates increased by around two years per decade, but with a standard deviation around 50% larger than this average. This highlights why it is important to incorporate measures of variation in life expectancy projections, and also why it can be tricky to identify a change in trend based on a small number of observations.

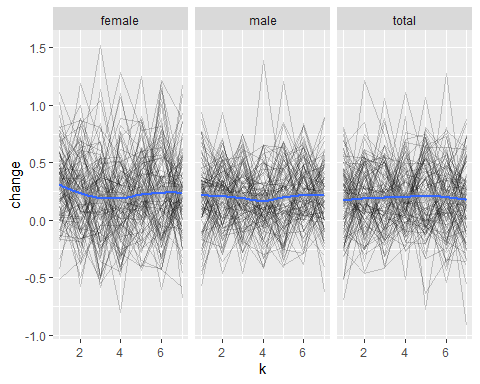
The code below creates the 10,000 simulations, each projecting forward seven periods (from 2012-2018 inclusive). A random number seed has been specified here, and the same random number seed is used later for (1990-2011) so that the only difference between the simulations are the summary statistics used.

set.seed(20)  
ten\_k\_runs <-   
 e0\_ch\_summary %>%   
 mutate(  
 draws = map2(  
 .x = mean\_de0, .y = sd\_de0,   
 .f = ~replicate(10000, rnorm(n = 7, mean = .x, sd = .y))  
 )  
 ) %>%   
 mutate(  
 draw\_df = map(  
 draws,  
 ~.x %>%   
 data.frame() %>%   
 mutate(k = 1:n()) %>%   
 gather(-k, key = "rep", value = "change") %>%   
 mutate(rep = str\_remove(rep, "X") %>%   
 as.numeric()  
 )   
 )  
 ) %>%   
 select(sex, draw\_df) %>%   
 unnest()  
  
  
# a 1% sample

The following code shows 100 of the 10,000 simulations. A smoothed curve is added to show the average change per period, which should be positive and broadly constant.

ten\_k\_runs %>%   
 filter(rep %in% sample(1:10000, 100)) %>%   
 ggplot(aes(x = k, y = change)) +  
 facet\_wrap(~sex) +   
 geom\_line(aes(group = rep), alpha = 0.2) +  
 stat\_smooth()

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



# to accumulate with 2011 observation as 0th value

The following code chunk changes these estimated annual changes into predicted changes, by accumulating the simulated values up until the current period , and adding the 2011 life expectancy to these accumulated changes. The first few rows of the dataset are then shown.

ten\_k\_all\_accumulated <-   
 ten\_k\_runs %>%   
 left\_join(dta\_scot\_tidy %>% filter(year == 2011) %>% select(-year)) %>%   
 group\_by(sex, rep) %>%   
 arrange(k) %>%   
 mutate(  
 cumulative\_change = cumsum(change),  
 prediction = e0 + cumulative\_change  
 ) %>%   
 ungroup() %>%   
 mutate(year = k + 2011) %>%   
 left\_join(  
 dta\_scot\_tidy %>% rename(obs\_e0 = e0)  
 )

## Joining, by = "sex"

## Joining, by = c("sex", "year")

ten\_k\_all\_accumulated

## # A tibble: 210,000 x 9  
## sex k rep change e0 cumulative\_chan~ prediction year obs\_e0  
## <chr> <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 fema~ 1 1 0.624 80.8 0.624 81.4 2012 80.8  
## 2 fema~ 1 2 -0.0921 80.8 -0.0921 80.7 2012 80.8  
## 3 fema~ 1 3 -0.322 80.8 -0.322 80.5 2012 80.8  
## 4 fema~ 1 4 0.163 80.8 0.163 81.0 2012 80.8  
## 5 fema~ 1 5 -0.658 80.8 -0.658 80.2 2012 80.8  
## 6 fema~ 1 6 -0.442 80.8 -0.442 80.4 2012 80.8  
## 7 fema~ 1 7 0.410 80.8 0.410 81.2 2012 80.8  
## 8 fema~ 1 8 0.0799 80.8 0.0799 80.9 2012 80.8  
## 9 fema~ 1 9 0.683 80.8 0.683 81.5 2012 80.8  
## 10 fema~ 1 10 -0.0940 80.8 -0.0940 80.7 2012 80.8  
## # ... with 209,990 more rows

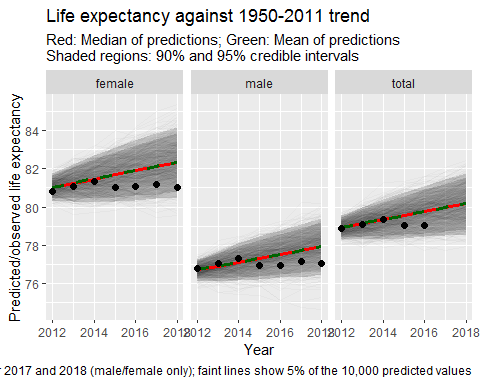
Next, summary statistics are produced for each sex and projection period showing the mean, median, and upper and lower 90% and 95% quantiles.

ten\_k\_all\_qis <-   
 ten\_k\_all\_accumulated %>%   
 group\_by(sex, year) %>%   
 summarise(  
 mean\_e0 = mean(prediction),  
 med\_e0 = median(prediction),  
 low\_95 = quantile(prediction, 0.025),  
 high\_95 = quantile(prediction, 0.975),  
 low\_90 = quantile(prediction, 0.050),  
 high\_90 = quantile(prediction, 0.950)  
 ) %>%   
 ungroup()

A 5% sample of of the simulated projections, along with the 90% and 95% intervals, are are shown in the image below. The observed life expectancies are also shown as points.

ten\_k\_5pc\_accumulated <-   
 ten\_k\_all\_accumulated %>%   
 filter(rep %in% sample(1:10000, 500))   
   
ten\_k\_5pc\_accumulated %>%   
 ggplot(aes(x = year, y = prediction, group = rep)) +   
 facet\_wrap(~sex) +   
 geom\_line(alpha = 0.025) +   
 geom\_ribbon(aes(x = year, ymin = low\_90, ymax = high\_90),  
 data = ten\_k\_all\_qis,  
 inherit.aes = FALSE,  
 alpha = 0.2  
 ) +   
 geom\_ribbon(aes(x = year, ymin = low\_95, ymax = high\_95),  
 data = ten\_k\_all\_qis,  
 inherit.aes = FALSE,  
 alpha = 0.2  
 ) +   
 geom\_line(  
 aes(x = year, y = med\_e0),  
 data = ten\_k\_all\_qis,  
 inherit.aes = FALSE,  
 colour = "red", size = 1.2  
 ) +   
 geom\_line(  
 aes(x = year, y = mean\_e0),  
 data = ten\_k\_all\_qis,  
 inherit.aes = FALSE,  
 colour = "darkgreen", size = 1.2, linetype = "dashed"  
 ) +   
 geom\_point(aes(y = obs\_e0), size = 2) +  
 labs(  
 x = "Year", y = "Predicted/observed life expectancy",  
 title = "Life expectancy against 1950-2011 trend",  
 subtitle = "Red: Median of predictions; Green: Mean of predictions\nShaded regions: 90% and 95% credible intervals",  
 caption = "Source: HMD; ONS/NRS for 2017 and 2018 (male/female only); faint lines show 5% of the 10,000 predicted values"  
 )

## Warning: Removed 1000 rows containing missing values (geom\_point).



ggsave("figures/obs\_vs\_pred\_e0\_1950\_2011.png", height = 15, width = 25, units = "cm", dpi = 300)

## Warning: Removed 1000 rows containing missing values (geom\_point).

From these figures it has become ever clearer that the the observed values have fallen ever further below the projected range, essentially flatlining after 2014, though broadly following the middle of the trendline for 2012-2014. The relative decline, against the range of expected values, appears to have been more severe for females than males.

The following figure shows how the observed life expectancy for each year compares with the distribution of projected estimates for that year.

ten\_k\_all\_accumulated %>%   
 group\_by(sex, year) %>%   
 summarise(empirical\_p = mean(prediction < obs\_e0)) %>%   
 spread(sex, empirical\_p)

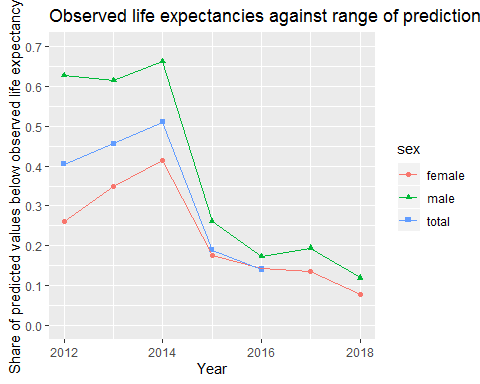
## # A tibble: 7 x 4  
## year female male total  
## <dbl> <dbl> <dbl> <dbl>  
## 1 2012 0.26 0.627 0.405  
## 2 2013 0.349 0.615 0.456  
## 3 2014 0.414 0.662 0.511  
## 4 2015 0.176 0.262 0.189  
## 5 2016 0.144 0.174 0.140  
## 6 2017 0.136 0.194 NA   
## 7 2018 0.0786 0.119 NA

By 2018, only 8% of the model projections for female life expectancy were below the observed life expectancy. For males only 12% of model projections were below the observed values. The relative decline, and sharp fall against projected values, is even more apparent when presenting the above data as a graph.

ten\_k\_all\_accumulated %>%   
 group\_by(sex, year) %>%   
 summarise(empirical\_p = mean(prediction < obs\_e0)) %>%   
 ggplot(aes(x = year, y = empirical\_p, group = sex, colour = sex, shape = sex)) +   
 geom\_point() + geom\_line() +  
 scale\_y\_continuous(limits = c(0, 0.70), breaks = seq(0, 0.7, 0.1)) +  
 labs(  
 x = "Year",   
 y = "Share of predicted values below observed life expectancy",  
 title = "Observed life expectancies against range of predictions based on 1950-2011 trend"  
  
 )

## Warning: Removed 2 rows containing missing values (geom\_point).

## Warning: Removed 2 rows containing missing values (geom\_path).



ggsave("figures/trend\_observed\_percentiles\_1950\_2011.png", height = 10, width = 12, units = "cm", dpi = 300)

## Warning: Removed 2 rows containing missing values (geom\_point).  
  
## Warning: Removed 2 rows containing missing values (geom\_path).

obs\_percentiles\_1950\_2011 <-   
 ten\_k\_all\_accumulated %>%   
 group\_by(sex, year) %>%   
 summarise(empirical\_p = mean(prediction < obs\_e0)) %>%   
 ungroup() %>%   
 mutate(fit\_period = "1950-2011")

## Interrim summary

The analyses above have compared observed life expectancies in Scotland against projected life expectancies based on annual changes observed between 1950 and 2011 inclusive (). The following analyses repeat this exercise using the more recent pre-slowdown period 1990-2011 (). The analyses have shown that life expectancies gains have fallen increasingly far down the range of values that could be plausibly expected based on these earlier trends.

# Analyses using (1990-2011)

The average annual improvement and standard deviation in these annual changes are calculated below

e0\_ch\_summary <-   
 dta\_scot\_tidy %>%   
 group\_by(sex) %>%   
 arrange(year) %>%   
 mutate(delta\_e0 = e0 - lag(e0)) %>%   
 ungroup() %>%   
 filter(between(year, 1990, 2011)) %>% # NOTE THE CHANGE HERE  
 group\_by(sex) %>%   
 summarise(  
 mean\_de0 = mean(delta\_e0),  
 sd\_de0 = sd(delta\_e0)  
 ) %>%   
 ungroup()  
  
e0\_ch\_summary\_p2 <- e0\_ch\_summary  
e0\_ch\_summary

## # A tibble: 3 x 3  
## sex mean\_de0 sd\_de0  
## <chr> <dbl> <dbl>  
## 1 female 0.211 0.254  
## 2 male 0.265 0.249  
## 3 total 0.237 0.245

Considering , the average improvement rates were slightly higher than over the longer period , mainly due to faster rates of improvement for males. Over both 1950-2011 and 1990-2011 female life expectancies increased at around 2.1 years per decade on average, for males the rates of increase were 2.1 years per decade over 1950-2011, and 2.7 years per decade over 1990-2011, leading to catch-up in life expectancies between the sexes. In the latter period 1990-2011 the annual variation was smaller than in the longer time period 1950-2011, with greater reductions in standard deviations for females (standard deviations of 0.25 years per year in the shorter period compared with 0.35 years per year in the longer period) than in males (from 0.29 years per year to 0.25 years per year). Both the reduced variation (especially for females) and the increased average improvement rates (especially for females) suggest that the observed life expectancies will fall proportionately further below the range of projections when using as the reference period than .

e0\_ch\_summary\_p1

## # A tibble: 3 x 3  
## sex mean\_de0 sd\_de0  
## <chr> <dbl> <dbl>  
## 1 female 0.214 0.353  
## 2 male 0.207 0.285  
## 3 total 0.211 0.307

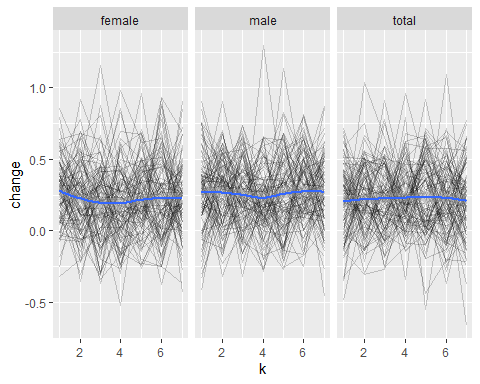
Draws from distributions calibrated on will now be calculated.

set.seed(20)  
ten\_k\_runs <-   
 e0\_ch\_summary %>%   
 mutate(  
 draws = map2(  
 .x = mean\_de0, .y = sd\_de0,   
 .f = ~replicate(10000, rnorm(n = 7, mean = .x, sd = .y))  
 )  
 ) %>%   
 mutate(  
 draw\_df = map(  
 draws,  
 ~.x %>%   
 data.frame() %>%   
 mutate(k = 1:n()) %>%   
 gather(-k, key = "rep", value = "change") %>%   
 mutate(rep = str\_remove(rep, "X") %>%   
 as.numeric()  
 )   
 )  
 ) %>%   
 select(sex, draw\_df) %>%   
 unnest()

As before, a 1% sample of samples from the distribution, this time using , will now be shown, along with the average improvement rate based on this limited sample. (The average improvement rate based on the complete sample will be much closer and less variable than the observed values)

ten\_k\_runs %>%   
 filter(rep %in% sample(1:10000, 100)) %>%   
 ggplot(aes(x = k, y = change)) +  
 facet\_wrap(~sex) +   
 geom\_line(aes(group = rep), alpha = 0.2) +  
 stat\_smooth()

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



And as before, these simulations will now be used to produced projected values against which the observed values can be compared

ten\_k\_all\_accumulated <-   
 ten\_k\_runs %>%   
 left\_join(dta\_scot\_tidy %>% filter(year == 2011) %>% select(-year)) %>%   
 group\_by(sex, rep) %>%   
 arrange(k) %>%   
 mutate(  
 cumulative\_change = cumsum(change),  
 prediction = e0 + cumulative\_change  
 ) %>%   
 ungroup() %>%   
 mutate(year = k + 2011) %>%   
 left\_join(  
 dta\_scot\_tidy %>% rename(obs\_e0 = e0)  
 )

## Joining, by = "sex"

## Joining, by = c("sex", "year")

The summary statistics (mean, median, upper and lower 90% and 95% quantiles)

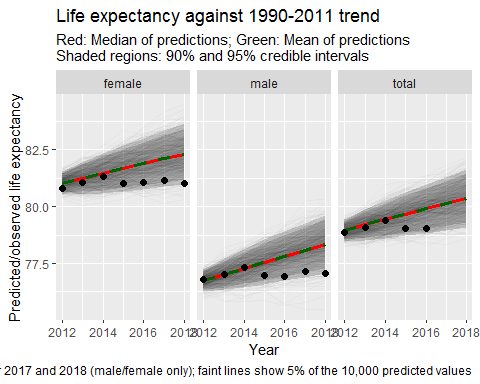
ten\_k\_all\_qis <-   
 ten\_k\_all\_accumulated %>%   
 group\_by(sex, year) %>%   
 summarise(  
 mean\_e0 = mean(prediction),  
 med\_e0 = median(prediction),  
 low\_95 = quantile(prediction, 0.025),  
 high\_95 = quantile(prediction, 0.975),  
 low\_90 = quantile(prediction, 0.050),  
 high\_90 = quantile(prediction, 0.950)  
 ) %>%   
 ungroup()  
ten\_k\_all\_qis

## # A tibble: 21 x 8  
## sex year mean\_e0 med\_e0 low\_95 high\_95 low\_90 high\_90  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 female 2012 81.0 81.0 80.5 81.5 80.6 81.5  
## 2 female 2013 81.2 81.2 80.5 82.0 80.7 81.8  
## 3 female 2014 81.5 81.5 80.6 82.3 80.7 82.2  
## 4 female 2015 81.7 81.7 80.7 82.7 80.8 82.5  
## 5 female 2016 81.9 81.9 80.7 83.0 80.9 82.8  
## 6 female 2017 82.1 82.1 80.9 83.3 81.1 83.1  
## 7 female 2018 82.3 82.3 81.0 83.6 81.2 83.4  
## 8 male 2012 76.8 76.8 76.3 77.2 76.4 77.2  
## 9 male 2013 77.0 77.0 76.3 77.7 76.5 77.6  
## 10 male 2014 77.3 77.3 76.5 78.1 76.6 78.0  
## # ... with 11 more rows

And now the equivalent visualisation showing observed against range of projected values will now be produced

ten\_k\_2pc\_accumulated <-   
 ten\_k\_all\_accumulated %>%   
 filter(rep %in% sample(1:10000, 500))   
  
ten\_k\_2pc\_accumulated %>%   
 ggplot(aes(x = year, y = prediction, group = rep)) +   
 facet\_wrap(~sex) +   
 geom\_line(alpha = 0.025) +   
 geom\_ribbon(aes(x = year, ymin = low\_90, ymax = high\_90),  
 data = ten\_k\_all\_qis,  
 inherit.aes = FALSE,  
 alpha = 0.2  
 ) +   
 geom\_ribbon(aes(x = year, ymin = low\_95, ymax = high\_95),  
 data = ten\_k\_all\_qis,  
 inherit.aes = FALSE,  
 alpha = 0.2  
 ) +   
 geom\_line(  
 aes(x = year, y = med\_e0),  
 data = ten\_k\_all\_qis,  
 inherit.aes = FALSE,  
 colour = "red", size = 1.2  
 ) +   
 geom\_line(  
 aes(x = year, y = mean\_e0),  
 data = ten\_k\_all\_qis,  
 inherit.aes = FALSE,  
 colour = "darkgreen", size = 1.2, linetype = "dashed"  
 ) +   
 geom\_point(aes(y = obs\_e0), size = 2) +  
 labs(  
 x = "Year", y = "Predicted/observed life expectancy",  
 title = "Life expectancy against 1990-2011 trend",  
 subtitle = "Red: Median of predictions; Green: Mean of predictions\nShaded regions: 90% and 95% credible intervals",  
 caption = "Source: HMD; ONS/NRS for 2017 and 2018 (male/female only); faint lines show 5% of the 10,000 predicted values"  
 )

## Warning: Removed 1000 rows containing missing values (geom\_point).



ggsave("figures/obs\_vs\_pred\_e0\_1990\_2011.png", height = 15, width = 25, units = "cm", dpi = 300)

## Warning: Removed 1000 rows containing missing values (geom\_point).

As implied by the higher average improvement rates and reduced annual variation over compared with , the observed values fall even lower against the distribution of projected values, with observed life expectancies for both males and females in 2018 falling to around the bottom 5% (lower end of the lighter grey credible interval band) of the distribution of projections. The table below, and subsequent figure, show where the observed life expectancies fall against the range of projected values for each year.

ten\_k\_all\_accumulated %>%   
 group\_by(sex, year) %>%   
 summarise(empirical\_p = mean(prediction < obs\_e0)) %>%   
 spread(sex, empirical\_p)

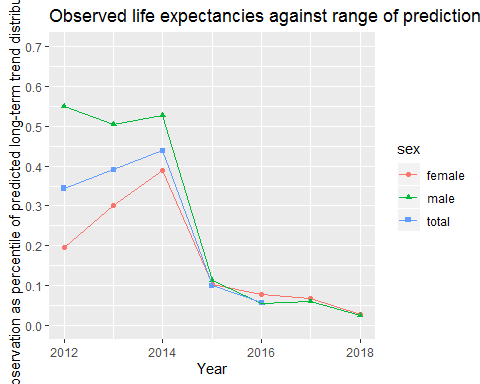
## # A tibble: 7 x 4  
## year female male total  
## <dbl> <dbl> <dbl> <dbl>  
## 1 2012 0.196 0.550 0.344   
## 2 2013 0.302 0.504 0.391   
## 3 2014 0.389 0.527 0.438   
## 4 2015 0.102 0.114 0.0993  
## 5 2016 0.0777 0.0548 0.0573  
## 6 2017 0.0686 0.0595 NA   
## 7 2018 0.0286 0.024 NA

By 2018, less than 3% of the projected estimates for life expectancy in 2018 were below the observed life expectancy. As with the range of distributions based on , values fell against expectations from 2014 onwards and have not recovered since.

ten\_k\_all\_accumulated %>%   
 group\_by(sex, year) %>%   
 summarise(empirical\_p = mean(prediction < obs\_e0)) %>%   
 ggplot(aes(x = year, y = empirical\_p, group = sex, colour = sex, shape = sex)) +   
 geom\_point() + geom\_line() +  
 scale\_y\_continuous(limits = c(0, 0.70), breaks = seq(0, 0.7, 0.1)) +  
 labs(  
 x = "Year",   
 y = "Observation as percentile of predicted long-term trend distribution",  
 title = "Observed life expectancies against range of predictions based on 1990-2011 trend"  
 )

## Warning: Removed 2 rows containing missing values (geom\_point).

## Warning: Removed 2 rows containing missing values (geom\_path).



ggsave("figures/trend\_observed\_percentiles\_1990\_2011.png", height = 10, width = 12, units = "cm", dpi = 300)

## Warning: Removed 2 rows containing missing values (geom\_point).  
  
## Warning: Removed 2 rows containing missing values (geom\_path).

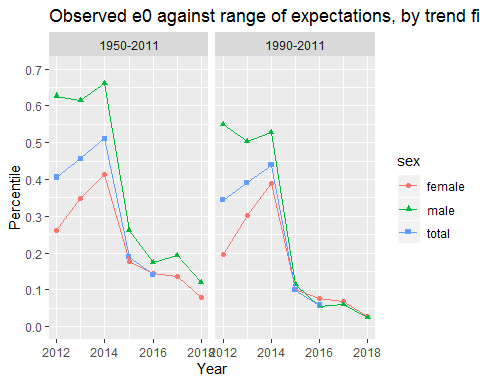
obs\_percentiles\_1990\_2011 <-   
 ten\_k\_all\_accumulated %>%   
 group\_by(sex, year) %>%   
 summarise(empirical\_p = mean(prediction < obs\_e0)) %>%   
 ungroup() %>%   
 mutate(fit\_period = "1990-2011")

The following figure shows how the observed life expectancies fell compared with the range of projected life expectancies for each year using both the and projections.

obs\_percentiles\_1950\_2011 %>%   
 bind\_rows(  
 obs\_percentiles\_1990\_2011  
 ) %>%   
 ggplot(aes(x = year, y = empirical\_p, group = sex, colour = sex, shape = sex)) +   
 facet\_wrap(~fit\_period) +   
 geom\_point() + geom\_line() +  
 scale\_y\_continuous(limits = c(0, 0.70), breaks = seq(0, 0.7, 0.1)) +  
 labs(  
 x = "Year", y = "Percentile",  
 title = "Observed e0 against range of expectations, by trend fit period"  
 )

## Warning: Removed 4 rows containing missing values (geom\_point).

## Warning: Removed 2 rows containing missing values (geom\_path).



ggsave("figures/trend\_observed\_percentiles\_bothperiods.png", height = 15, width = 20, units = "cm", dpi = 300)

## Warning: Removed 4 rows containing missing values (geom\_point).  
  
## Warning: Removed 2 rows containing missing values (geom\_path).

# Projections against expected for mortality rates at single ages

The same process to projecting forward life expectancies at birth can also be applied to age specific mortality rates . To have a consistent data source, with associated consistency in methods applied, let’s use the single year lifetable data used to extract the 2017 and 2018 life expectancies previously.

readxl::excel\_sheets("data/singleyearlifetablesscotland1.xls")

## [1] "Contents" "Terms and Conditions" "Notation"   
## [4] "2018" "2017" "2016"   
## [7] "2015" "2014" "2013"   
## [10] "2012" "2011" "2010"   
## [13] "2009" "2008" "2007"   
## [16] "2006" "2005" "2004"   
## [19] "2003" "2002" "2001"   
## [22] "2000" "1999" "1998"   
## [25] "1997" "1996" "1995"   
## [28] "1994" "1993" "1992"   
## [31] "1991" "1990" "1989"   
## [34] "1988" "1987" "1986"   
## [37] "1985" "1984" "1983"   
## [40] "1982" "1981" "1980"

#Data are available from 1980. Each year is a worksheet. Within each worksheet the data are in the same cell locations   
  
clean\_data <- function(X){  
 out <- X %>%   
 select(-junk) %>%   
 gather(-x, key = "sex\_quant", value = "value") %>%   
 separate(sex\_quant, into = c("sex", "variable")) %>%   
 mutate(sex = case\_when(sex == 'm' ~ "male", sex == 'f' ~ "female", TRUE ~ NA\_character\_))  
 return(out)   
}  
  
mx\_data <-   
 tibble(  
 year = as.character(1980:2018)  
 ) %>%   
 mutate(dirty\_data = map(  
 year,   
 ~read\_excel(  
 path = "data/singleyearlifetablesscotland1.xls",  
 range = "A8:L108",  
 col\_names = c("x", "m\_mx", "m\_qx", "m\_lx", "m\_dx", "m\_ex", "junk", "f\_mx", "f\_qx", "f\_lx", "f\_dx", "f\_ex"),  
 sheet = .x  
 ))  
 ) %>%   
 mutate(  
 tidy\_data = map(dirty\_data, clean\_data)  
 ) %>%   
 select(-dirty\_data) %>%   
 unnest() %>%   
 spread(variable, value) %>%   
 mutate(year = as.numeric(year))  
  
mx\_data

## # A tibble: 7,878 x 8  
## year x sex dx ex lx mx qx  
## <dbl> <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1980 0 female 1046. 75.2 100000 0.0105 0.0105   
## 2 1980 0 male 1362. 69.0 100000 0.0137 0.0136   
## 3 1980 1 female 82 75.0 98954. 0.000829 0.000828  
## 4 1980 1 male 103. 68.9 98638. 0.00105 0.00105   
## 5 1980 2 female 23.4 74.1 98872. 0.000236 0.000236  
## 6 1980 2 male 67.2 68.0 98535. 0.000682 0.000682  
## 7 1980 3 female 40.9 73.1 98848. 0.000414 0.000414  
## 8 1980 3 male 57.6 67.0 98468. 0.000585 0.000585  
## 9 1980 4 female 18.8 72.1 98807. 0.00019 0.00019   
## 10 1980 4 male 55.8 66.1 98410 0.000567 0.000567  
## # ... with 7,868 more rows

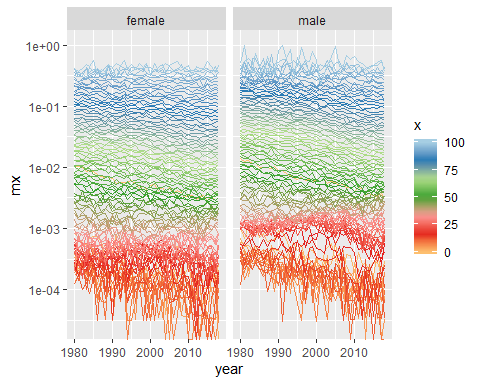
Now to save this

mx\_data %>% write\_csv("data/mx\_data\_scot\_1980\_2018.csv")

Let’s visualise how has changed at various ages

ggplot(mx\_data, aes(x = year, y = mx, colour = x, group = x)) +   
 geom\_line() +   
 facet\_wrap(~sex) +  
 scale\_y\_log10() +  
 scale\_colour\_distiller(palette = "Paired")

## Warning: Transformation introduced infinite values in continuous y-axis

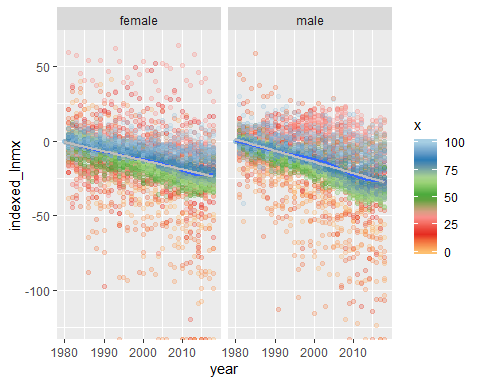


Indexing all values to initial year, what’s the percentage trend across all ages, and how much variation (implicitly this is about understanding how reasonable the drift assumption in the Lee-Carter spec seems to be)

mx\_data %>%   
 select(year, x, sex, mx) %>%   
 mutate(lnmx = log10(mx)) %>%   
 group\_by(x, sex) %>%   
 mutate(indexed\_lnmx = 100 \* (lnmx - lnmx[year == 1980] )) %>%   
 ggplot(aes(x = year, y = indexed\_lnmx)) +   
 geom\_point(aes(colour = x), alpha = 0.25) +  
 facet\_wrap(~sex) +   
 scale\_colour\_distiller(palette = "Paired") +  
 stat\_smooth() +   
 stat\_smooth(method = "lm", colour = "grey")

## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

## Warning: Removed 26 rows containing non-finite values (stat\_smooth).  
  
## Warning: Removed 26 rows containing non-finite values (stat\_smooth).



The nonlinear smoother, indicated by the blue line, is similar to the linear regression line of best fit, indicating that the linear drift assumption made in Lee-Carter modelling approach is not unreasonable. However it is clear that mortality rates in young children (orange colours) seem to be improving proportionately faster than the general trend, improvement rates are somewhat faster in early retirement ages (light greens), and mortality rate trends are persistently worse than the trend in young adulthood (red/pinky colours).

To avoid getting too sucked into the fascinating enormity of mortality data, let’s define the task very carefully:

* For each sex and age in single years
  + For 1990-2011
  + Calculate mean and SD in log
  + Generate 1000 simulations for 7 forward periods
  + Show against observed for 2012-2018 for a few specific ages
    - 0
    - 25
    - 45
    - 65
    - 80
    - 90
  + Calculate empirical quantiles for all ages
    - show as heatmap and/or linemaps

set.seed(20)  
  
  
mx\_data %>%  
 filter(between(year, 1990, 2011)) %>%   
 mutate(mx = mx + 1E-6) %>% #continuity correction  
 select(x, sex, year, mx) %>%   
 group\_by(x, sex) %>%   
 arrange(year) %>%   
 mutate(lnmx = log(mx)) %>%   
 mutate(d\_lnmx = lnmx - lag(lnmx)) %>%   
 summarise(  
 mean\_ch = mean(d\_lnmx, na.rm = TRUE),  
 sd\_ch = sd(d\_lnmx, na.rm = TRUE)  
 ) %>%   
 mutate(  
 predblock = map2(.x = mean\_ch, .y = sd\_ch, ~replicate(1000,rnorm(n = 7, mean = .x, sd = .y)))   
 )

## # A tibble: 202 x 5  
## # Groups: x [101]  
## x sex mean\_ch sd\_ch predblock   
## <dbl> <chr> <dbl> <dbl> <list>   
## 1 0 female -0.0310 0.137 <dbl[,1000] [7 x 1,000]>  
## 2 0 male -0.0307 0.0905 <dbl[,1000] [7 x 1,000]>  
## 3 1 female -0.00895 0.360 <dbl[,1000] [7 x 1,000]>  
## 4 1 male -0.0384 0.526 <dbl[,1000] [7 x 1,000]>  
## 5 2 female -0.0418 0.651 <dbl[,1000] [7 x 1,000]>  
## 6 2 male -0.0570 0.537 <dbl[,1000] [7 x 1,000]>  
## 7 3 female -0.0477 0.500 <dbl[,1000] [7 x 1,000]>  
## 8 3 male -0.0717 0.902 <dbl[,1000] [7 x 1,000]>  
## 9 4 female 0.00655 0.684 <dbl[,1000] [7 x 1,000]>  
## 10 4 male -0.0851 0.531 <dbl[,1000] [7 x 1,000]>  
## # ... with 192 more rows