MTH6157

Lee Carter model example in R

We start with two .csv files that contain deaths and exposed to risk for ages 20 to 100 (81 ages) and for years 1961 to 2020 (60 years).

First, we set the working directory and load the data into two matrices named `deaths` and `exposures`. We also need to remove the first column from each data set.

```
> setwd("C:/6157 Survival Models 2023-24")

> deaths <- as.matrix(read.csv("Week 12 Mortality Projection Data
Deaths.csv"))

> deaths <- deaths[,-1]

> exposures <- as.matrix(read.csv("Week 12 Mortality Projection Data
Exposures.csv"))

> exposures <- exposures[,-1]

> head(deaths)

> head(exposures)
```

Now we calculate $m_{x,t} = \frac{deaths}{exposures}$ at each x, t and then the $\log(m_{x,t})$ for Lee-Carter.

```
> m_xt <- deaths / exposures

> head(m_xt)

> log_m_xt <- log(m_xt)

> head(log_m_xt)
```

First we calculate the a_x which is the average $\log(m_{x,t})$ at each age *x*.

Each age is represented by a row in the `log_m_xt` matrix.

```
> a_x <- rowMeans(log_m_xt)

> a_x

 [1] -7.1100645 -7.1329511 -7.1389894 -7.1534817 -7.1656507

 [6] -7.1477501 -7.1285656 -7.1208472 -7.0953392 -7.0613297

[11] -7.0114064 -6.9756979 -6.9151998 -6.8717821 -6.8079288

[16] -6.7328344 -6.6669970 -6.5864162 -6.5123869 -6.4139687

[21] -6.3189546 -6.2274895 -6.1317109 -6.0357455 -5.9400777

[26] -5.8263991 -5.7305774 -5.6261931 -5.5390774 -5.4208850

[31] -5.3183738 -5.2254998 -5.1178271 -5.0114198 -4.9127894

[36] -4.8120512 -4.7040798 -4.5993046 -4.4958682 -4.3947634

[41] -4.2894575 -4.1927209 -4.0870618 -3.9901655 -3.8849100
```

```
[46] -3.7911973 -3.7024702 -3.5967532 -3.5043274 -3.4060688
[51] -3.3156642 -3.2227274 -3.1161163 -3.0208076 -2.9209340
[56] -2.8314694 -2.7332195 -2.6473568 -2.5515024 -2.4550707
[61] -2.3614175 -2.2753491 -2.1714134 -2.0788587 -1.9800759
[66] -1.9029173 -1.8040241 -1.7149646 -1.6269088 -1.5329509
[71] -1.4466159 -1.3644631 -1.2618977 -1.1808638 -1.1023631
[76] -1.0170922 -0.9317227 -0.8674852 -0.8065266 -0.7514876
[81] -0.6542951
```

Now under Lee Carter $\log(m_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t}$ therefore to find b_x and k_t we need a form of regression on $\log(m_{x,t}) - a_x = b_x k_t + \varepsilon_{x,t}$

First we transpose the `log_m_xt` matrix naming the new matrix `LT`.

And then subtract the relevant a_x value from each of the 81 age columns of the transposed matrix, `LT`.

```
> LT <- t(log_m_xt)
> for(j in 1:81) LT[,j] <- LT[,j] - a_x[j]
```

We can now find b_x and k_t by a method called Singular Value Decomposition or SVD (outside of the syllabus of this module but shown here in R to complete the Lee Carter modelling). SVD is a matrix regression methodology. It takes a matrix such as our `LT` and re-writes it in the form,

LT = **U D V**

Where U and V' are orthogonal matrices and D is a diagonal matrix

Lee Carter has b_x equal to the first row of **V** in the SVD but normalised so that $\sum_{ages} b_x = 1$

The k_t are found by multiplying the first column of **U** by the (1,1) element of **D** and the sum of the first row of **V'**. This has the effect of of ensuring that $\sum_{years} k_t = 0$.

In our example the decomposition matrix, which we will call decomp is made up of two vectors of length 60 which R will default to naming `d` and `v` and a vector of length 81 named `u`. We use values from these vectors to find our b_x and k_t.

```
> decomp <- svd(LT, 1, 1)
> b_x <- decomp$v / sum(decomp$v)
> b_x
            [,1]
 [1,] 0.013903767
 [2,] 0.013010429
 [3,] 0.012328419
 [4,] 0.009963432
```

```
 [5,] 0.009262974
 [6,] 0.007678748
 [7,] 0.007413045
 [8,] 0.006941794
 [9,] 0.005960373
[10,] 0.005406260
[11,] 0.004983449
[12,] 0.004394205
[13,] 0.004582124
[14,] 0.004166568
[15,] 0.004953114
[16,] 0.005268926
[17,] 0.005700848
[18,] 0.005532399
[19,] 0.006496751
[20,] 0.007136693
[21,] 0.007818349
[22,] 0.008475825
[23,] 0.009906504
[24,] 0.010159529
[25,] 0.010866634
[26,] 0.011898904
[27,] 0.012820267
[28,] 0.013570068
[29,] 0.014206135
[30,] 0.014854949
[31,] 0.015566125
[32,] 0.015793888
[33,] 0.016842778
[34,] 0.017355672
[35,] 0.018002633
[36,] 0.017926483
```

```
[37,] 0.018254107
[38,] 0.018795398
[39,] 0.019222053
[40,] 0.019532819
[41,] 0.019356244
[42,] 0.019540129
[43,] 0.019879473
[44,] 0.019983625
[45,] 0.020395140
[46,] 0.020449429
[47,] 0.019967346
[48,] 0.020534447
[49,] 0.020291737
[50,] 0.020191358
[51,] 0.019962558
[52,] 0.019321075
[53,] 0.019484384
[54,] 0.019268482
[55,] 0.018836110
[56,] 0.018174729
[57,] 0.017766824
[58,] 0.017084240
[59,] 0.016518872
[60,] 0.016006466
[61,] 0.015213803
[62,] 0.014287235
[63,] 0.014018851
[64,] 0.013207407
[65,] 0.012776004
[66,] 0.012182108
[67,] 0.011707149
[68,] 0.010933794
```

```
[69,] 0.010001395
[70,] 0.009273137
[71,] 0.008114189
[72,] 0.007284205
[73,] 0.007236794
[74,] 0.006144600
[75,] 0.005876817
[76,] 0.004848157
[77,] 0.004932189
[78,] 0.003576526
[79,] 0.003390228
[80,] 0.002093268
[81,] 0.002936034
> k_t <- decomp$u * sum(decomp$v) * decomp$d[1]
> k_t
             [,1]
 [1,]  26.8553800
 [2,]  26.5722421
 [3,]  27.7520758
 [4,]  23.5039724
 [5,]  24.1417029
 [6,]  25.0151972
 [7,]  21.3996298
 [8,]  23.6332872
 [9,]  23.8574801
[10,]  22.7903592
[11,]  20.7873162
[12,]  22.5389923
[13,]  21.8124750
[14,]  20.3935778
[15,]  19.4665575
[16,]  19.9129823
```

```
[17,]  17.1020353
[18,]  17.8435259
[19,]  17.2528878
[20,]  14.7274089
[21,]  13.2568738
[22,]  12.4361639
[23,]  11.4450054
[24,]   9.2387758
[25,]   9.8593216
[26,]   8.3016219
[27,]   6.1570147
[28,]   5.5405302
[29,]   4.4956481
[30,]   3.8507917
[31,]   2.7486435
[32,]   0.5362723
[33,]   1.0702416
[34,]  -2.2381855
[35,]  -1.5287609
[36,]  -3.0262347
[37,]  -4.4716193
[38,]  -5.5281141
[39,]  -7.0549626
[40,]  -9.0601182
[41,] -10.9460786
[42,] -11.7715225
[43,] -12.7876180
[44,] -16.2110113
[45,] -18.1399644
[46,] -19.7288088
[47,] -21.2441444
[48,] -21.8511599
```

```
[49,] -24.5815586
[50,] -26.6062788
[51,] -29.7689160
[52,] -31.2363929
[53,] -30.9025603
[54,] -31.9858987
[55,] -31.0679802
[56,] -31.3885141
[57,] -32.4735242
[58,] -31.3881899
[59,] -33.1938369
[60,] -26.1140362
```

We check the sums of the

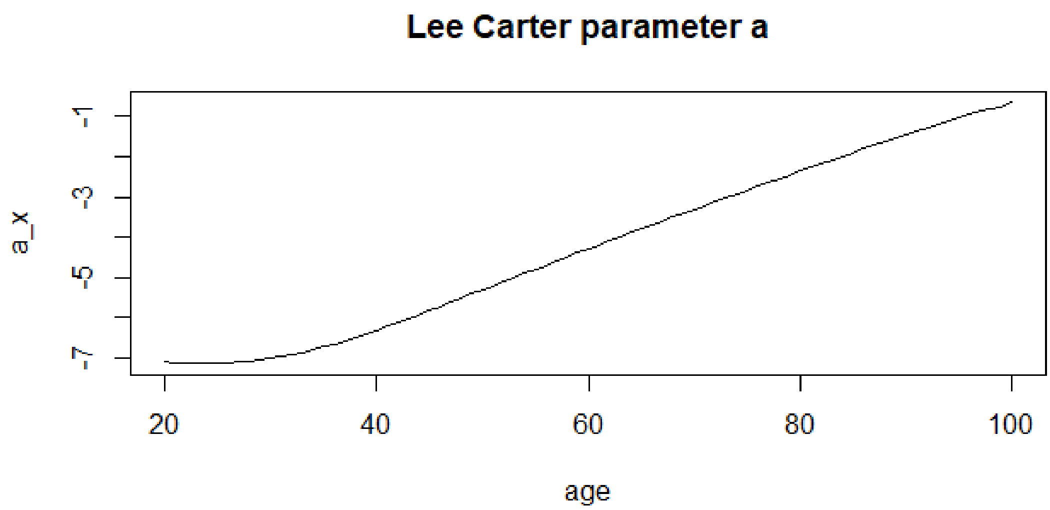
```
> sum(b_x)
[1] 1
> sum(k_t)
[1] 6.117329e-14
```

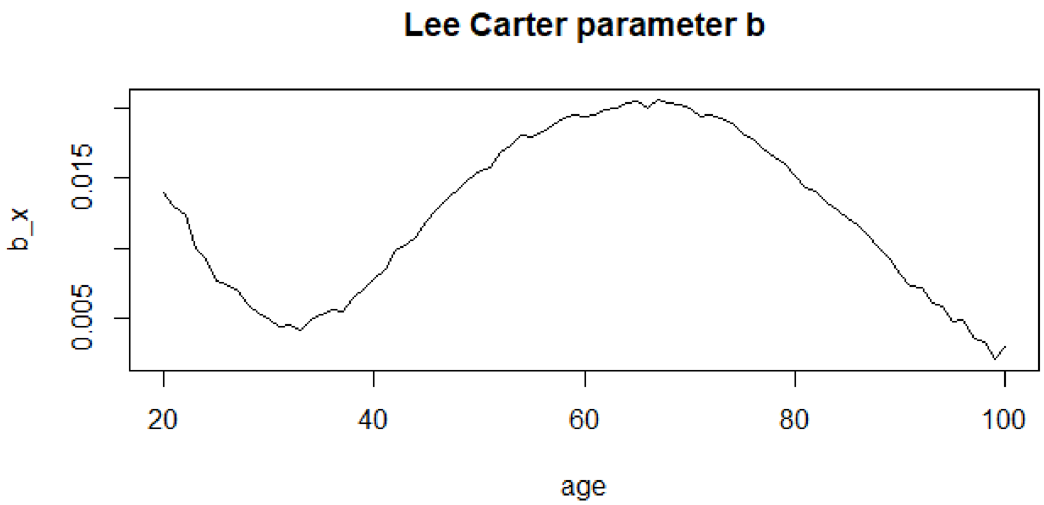
Now we can plot the three Lee Carter parameters

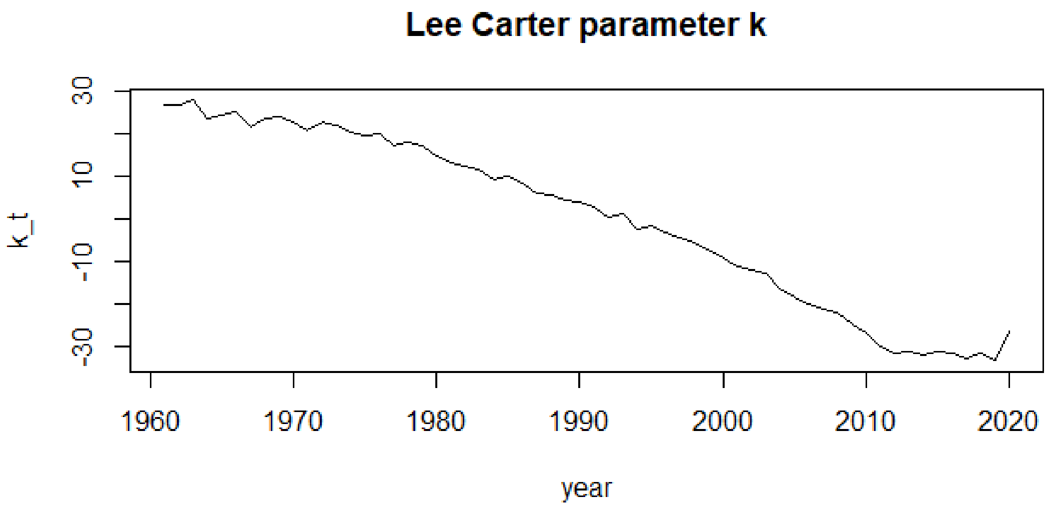
```
> x = c(20:100)
> t = c(1961:2020)
> plot(x, a_x, type="l", main = "Lee Carter parameter a", xlab =
"age")
> plot(x, b_x, type="l", main = "Lee Carter parameter b", xlab =
"age")
> plot(t, k_t, type="l", main = "Lee Carter parameter k", xlab =
"year")
```

<u>Question</u>

How would you interpret these three plots below?


Lee Carter parameter a
a_x
-1
-3
-5
-7
20
40
60
80
100
age


Lee Carter parameter b
b_x
0.015
0.005
20
40
60
80
100
age


Lee Carter parameter k
k_t
30
10
-10
-30
1960
1970
1980
1990
2000
2010
2020
year