FPCA Weight and Maturity Investigation

Jonathan Babyn

March 16, 2019

1 FDA and FPCA bare minimum context

FDA is functional data analysis, basically data analysis performed on curves. FPCA is Functional Principal Component Analysis and is the extension of the classical statistical technique Principal Component Analysis used to find trends in the data. FPCA extends this to curves/functions, it can also be used for noise reduction and interpolation.

If you want more details on this I can point you to some books/articles or my MSc. thesis because it was basically centred around FDA.

Matt mentioned smoothing the maturity curves and I thought maybe FPCA would be a way to do it. I also tried it on the DFO stock weights and commercial weights and I think it did a pretty good job there too.

2 Smoothing Mats

FPCA was applied to the maturities with each age being considered a measurement along the years. The FPCA was performed in logit space so that when transformed back they were ensured to be between 0-1. FPCA breaks curves down into a mean function or mean curve and the corresponding eigenfunctions. I had to somewhat manually smooth the mean function to get something that seemed reasonable. Letting it pick would keep a lot of the peaks and valleys in the data here. Figure 1 is a dianostic plot of the maturity curve FPCA, showing the first three eigenfunctions, the mean function and the variability explained by the eigenfunctions.

The FPCA process removes a lot of noise in the data while keeping the parts that explain most of the variability. A comparision of the smoothed vs. unsmoothed appears in 2. Most of the noise is gone while the overall trend is captured.

When looking at the smoothed logistic regression maturity curves for each year, it also cleanly breaks into two time periods, pre-1978 and post-1978.
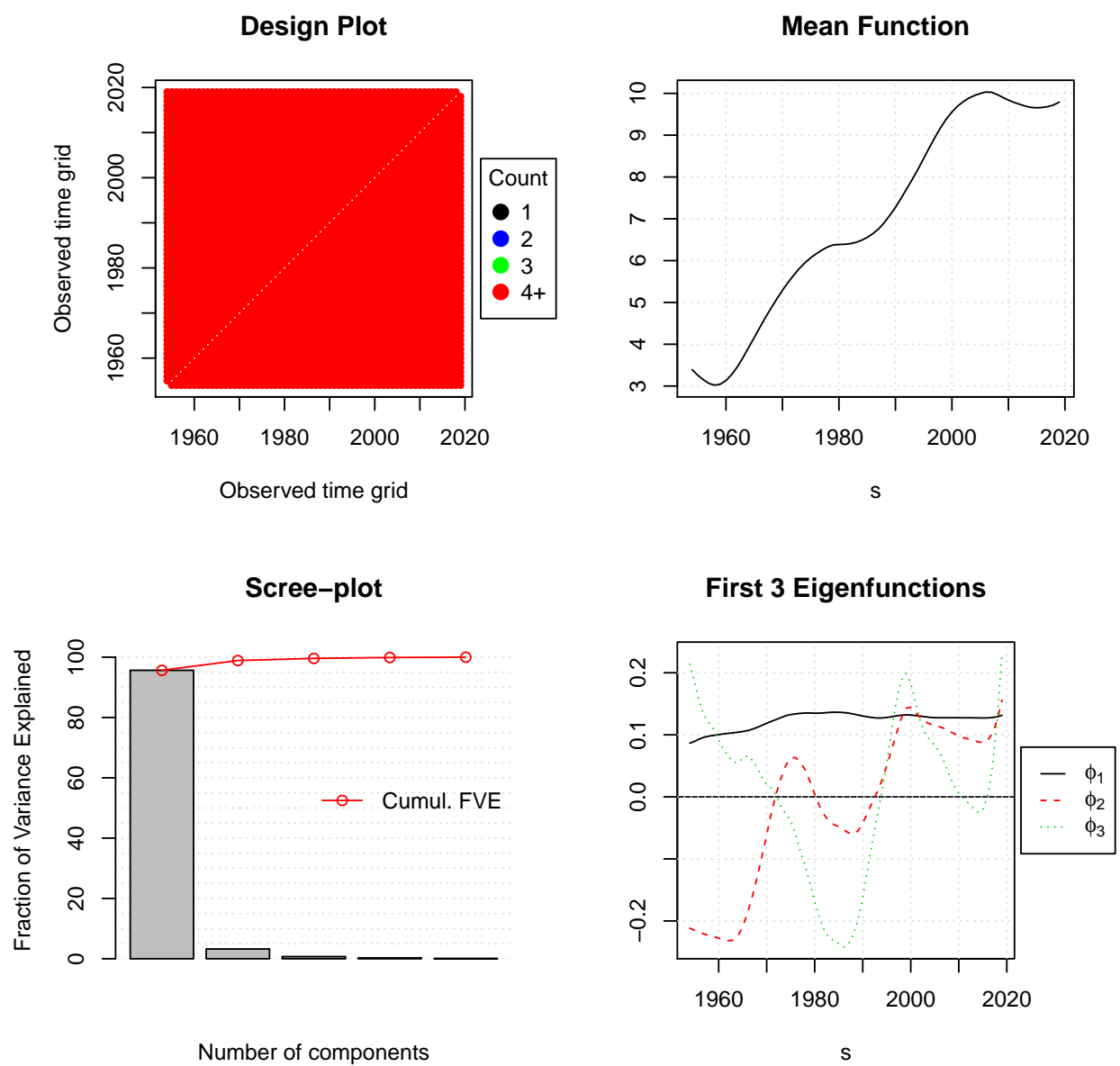
Design Plot

Mean Function

Scree−plot

First 3 Eigenfunctions
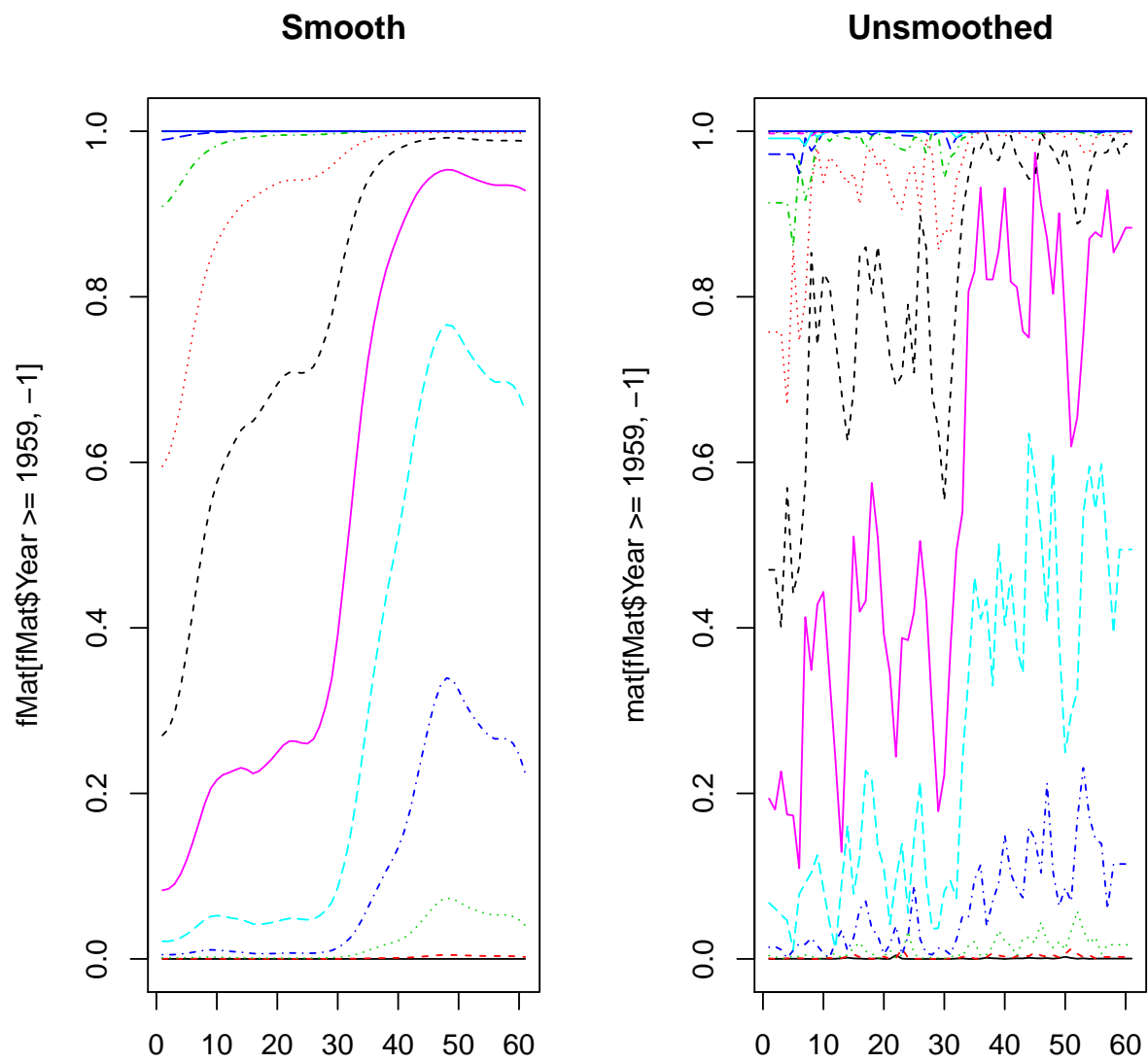
Figure 1: Maturity FPCA diagnostic plots

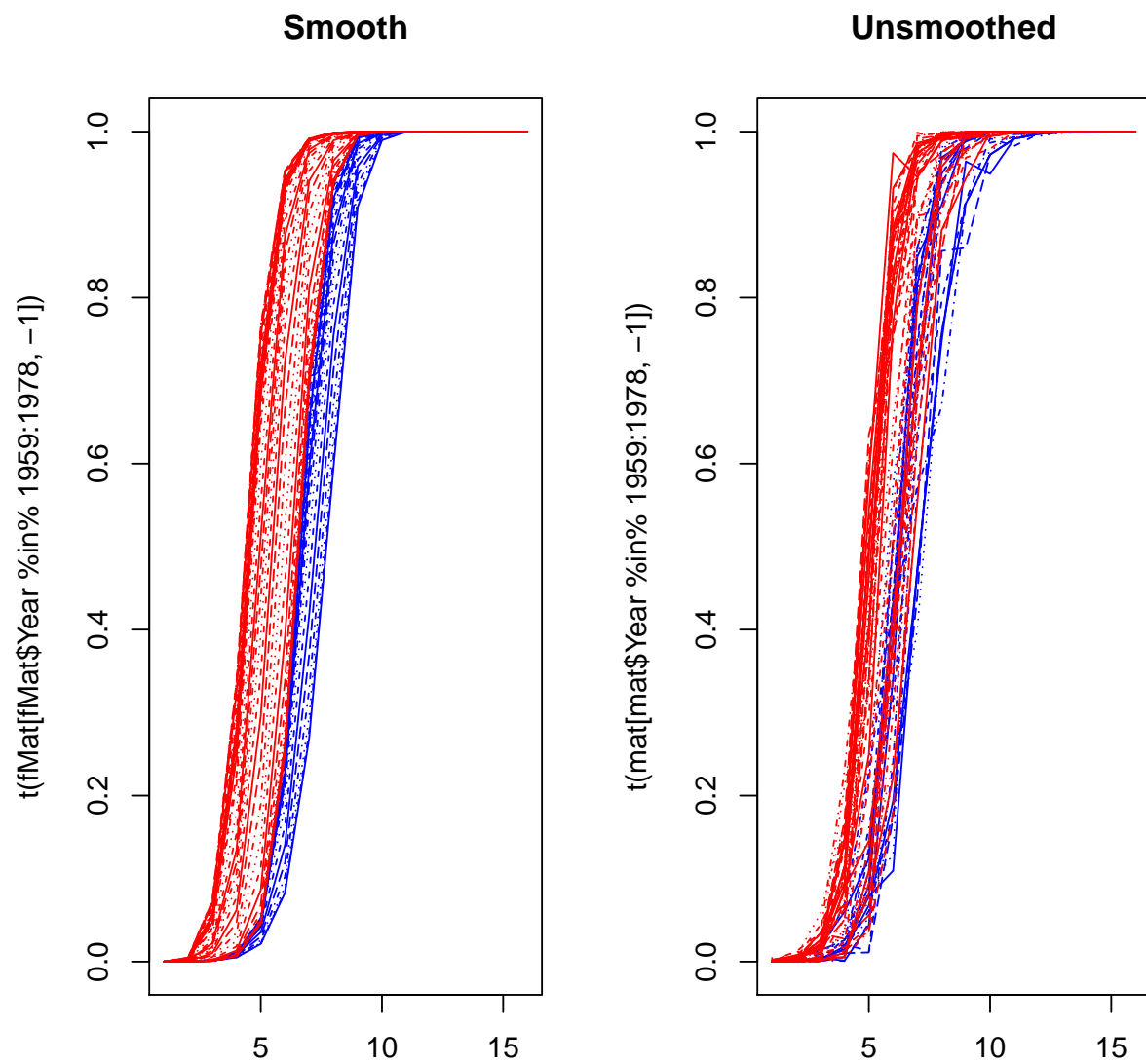Figure 2: Smoothed Maturitys by year vs. not

Figure 3: The logistic regression curves.Red is post 1978.

3   Smoothing Weights

FPCA was also applied on the commerical weights and the DFO stock weights. Over 90% of the variability is explained by a single eigenfunction and really reduces the noise when using data with only a single functional principal component. When looking at the data represented by a single FPC, the noise is much reduced and a couple things become visible. To me, it looks as though the weights break into 3 periods, the really wonky late stuff with 14 year olds under 6kg, those 14 year olds under 8Kg and those above. Looking at the years this is pretty cleanly pre-1994 for those above 8Kg, then post 1994, then the last 3 years of the data as seen in fig. 4. I tried using a clustering algorithm on the FPC scores generated and it found the same thing on commercial data, only pre and post 1994 for the stock weights. Since the DFO stock and commercial weight data is only ages 3 to 14 I needed to extend it to 1 through 16. This was done by extrapolating the mean function and the FPCs using splines. I don't really like extrapolating splines at the edges because they often do really wonky things but it worked pretty well here IMO since this are somewhat simple curves. The extrapolation is seen in fig 5.

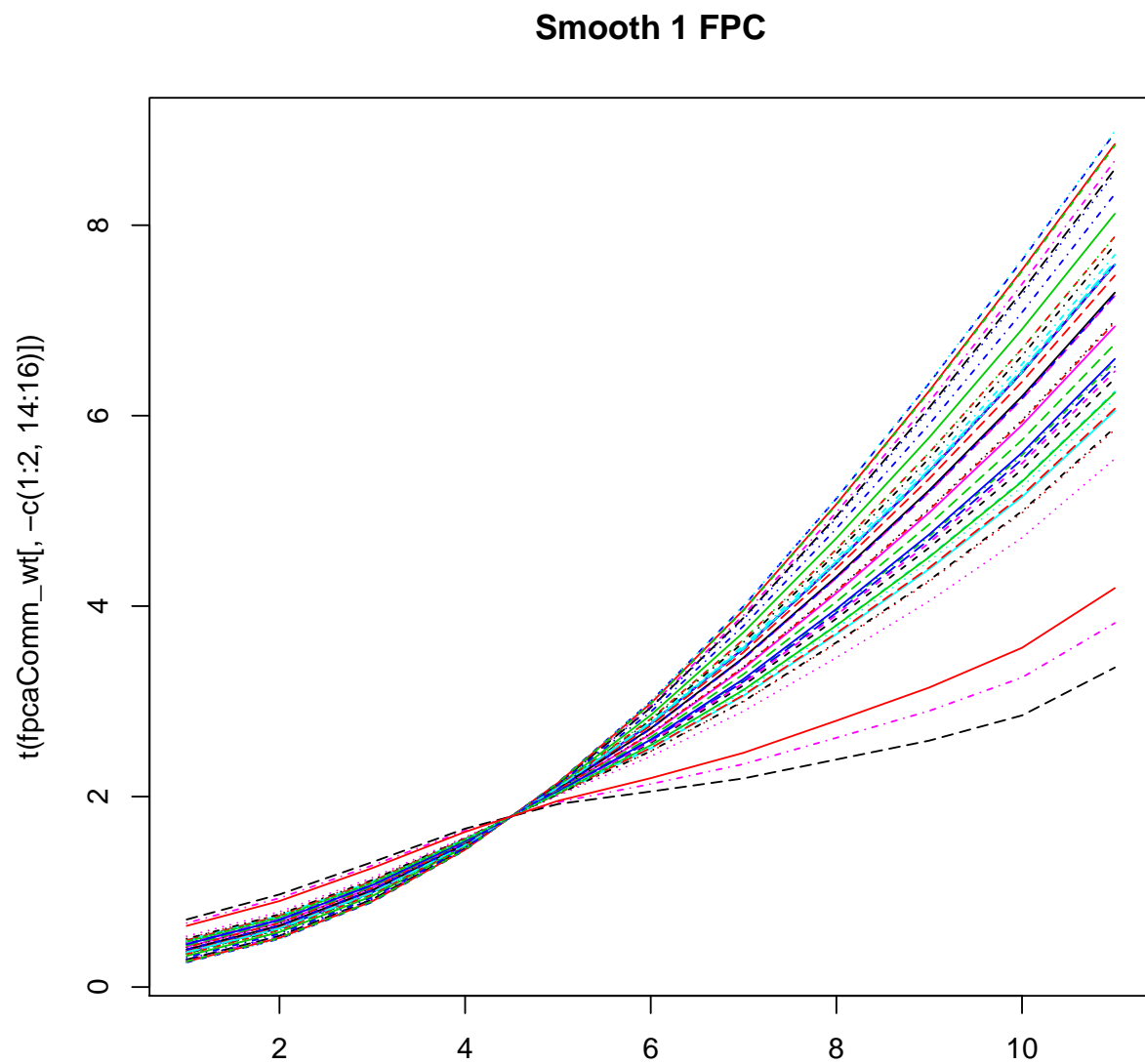I think these smoothed weights compare pretty favourably with Noel's midy_wt and stock_wt stuff.

Smooth 1 FPC

Figure 4: The fitted commercial weights with a single FPC, ages 3 to 14(ignore axis)
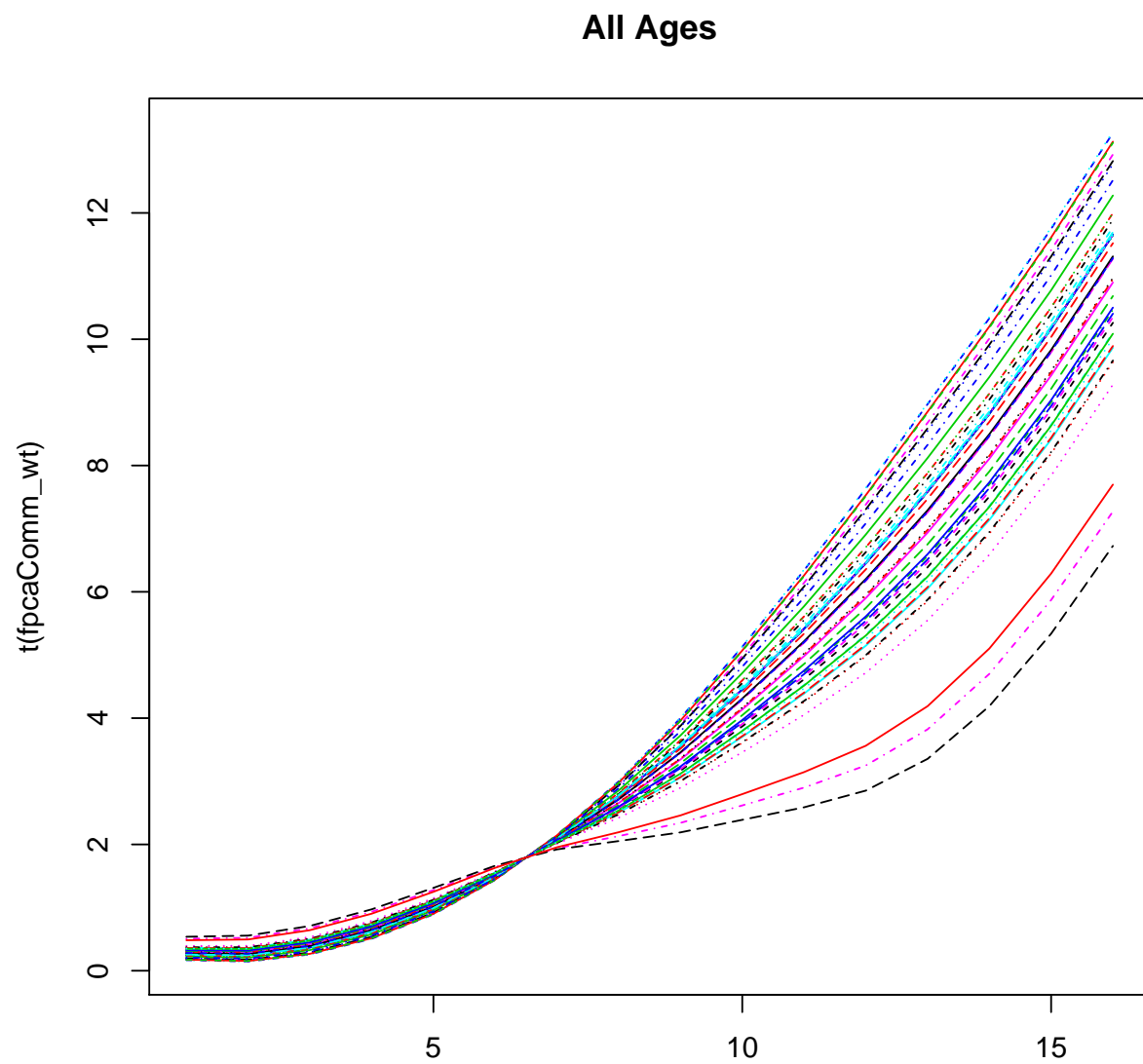
All Ages

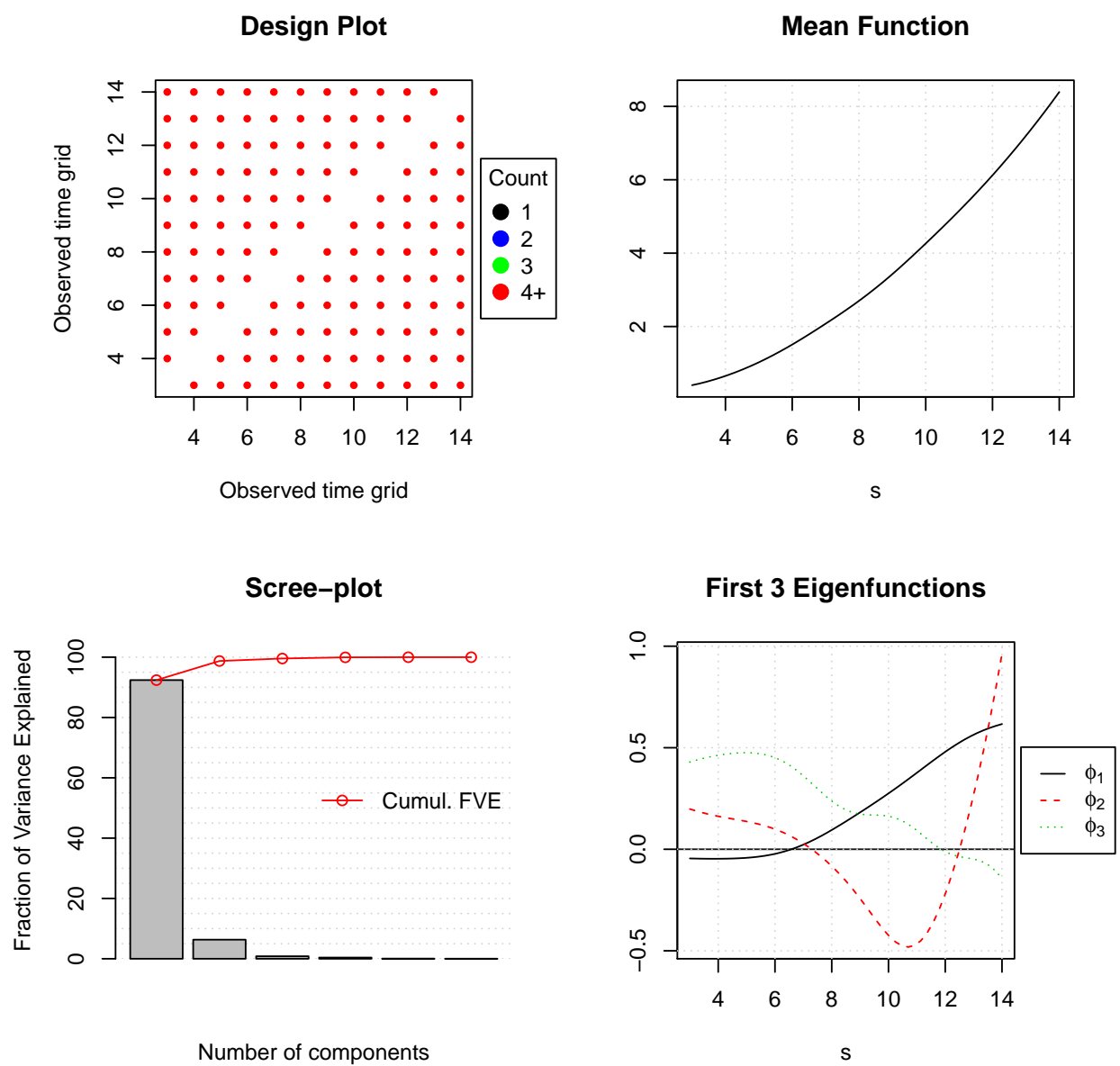Figure 5: Extrapolating the smoothed data to more ages

Figure 6: Diagnostic FPCA plots for the commercial weights. Stock weights more or less is identical.