Empirical Methods in Finance
TA Session

Mahyar Kargar

March 4, 2019

How many factors drive the term structure?

In this excerise, following Litterman and Scheinkman (1991), I try to answer the question above. A (non-zero) vector q of dimension N is an eigenvector of a square $(N \times N)$ matrix Σ iff

$$\Sigma q = \lambda q$$

where the scalar λ is the eigenvalue corresponding to eigenvector q.

Construct matrix Q so that its i-th column is the eigenvector q_i of $\Sigma = cov(y)$. Construct diagonal matrix Λ so that its diagonal elements are the corresponding eigenvalues, i.e. $\Lambda_{ii} = \lambda_i$. Then we have:

$$\begin{aligned} \Sigma &= Q \Lambda Q' \\ x_t &= Q'\, y_t \\ Cov(x) &= \Lambda, \end{aligned}$$

Note that since eigenvectors are orthonormal[1], $Q\,Q' = I$ and $Q' = Q^{-1}$.

As a result we have a factor model of the yield curve:

- Λ gives us the variances of the factors
- $\lambda_i / \sum_i \lambda_i$ gives the fraction of variance explained by the i-th factor
- Columns of Q (i.e. eigenvectors of Σ) tell us how y loads on principal components x

Let's go look at the data on zero-coupon bond yields. Download monthly zero-coupon 1- to 5-year yields from CRSP from June 1952 to December 2018 and look at their principal components.[2] PCA tries to use a few (< 5) linear combinations to explain the structure of Σ_{yields}.

[1] Eigenvectors have unit length and are orthogonal to each other: $q_i' q_i = 1$ and $q_i' q_j = 0$, $i \neq j$.

[2] CRSP TREASURIES - Fama-Bliss Discount Bonds from wrds.

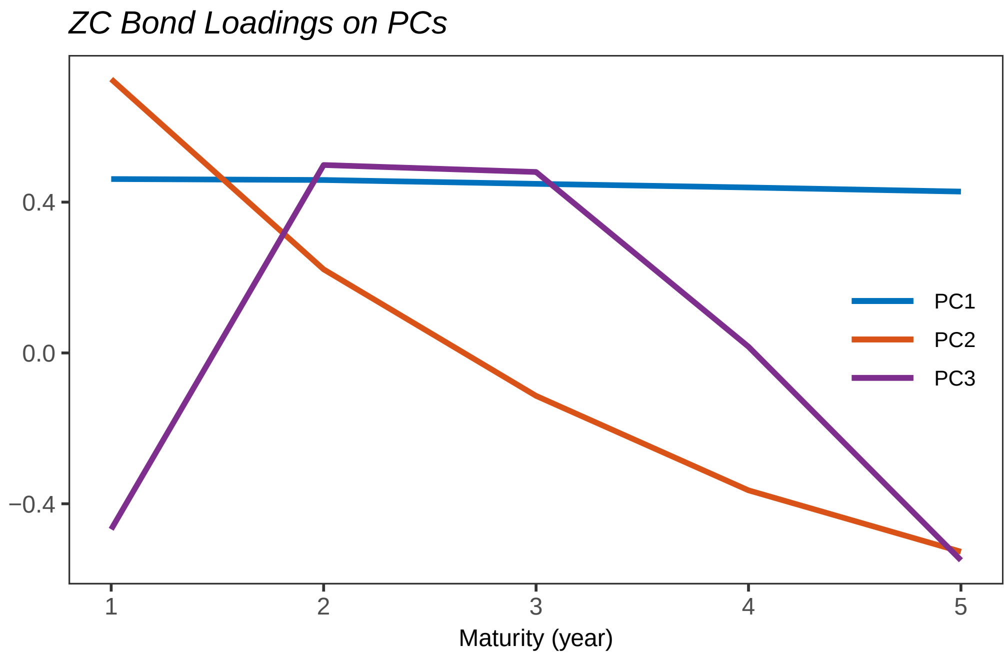
```r
yields <- fread("./ZC_FamaBliss.csv")
PCA <- prcomp(~ y1 + y2 + y3 + y4 + y5, data = yields)
summary(PCA)

## Importance of components:
##                           PC1     PC2     PC3     PC4     PC5
## Standard deviation     6.8355 0.58897 0.11255 0.07392 0.05774
## Proportion of Variance 0.9922 0.00737 0.00027 0.00012 0.00007
## Cumulative Proportion  0.9922 0.99954 0.99981 0.99993 1.00000

PCs <- data.table(indx=1:5,PC1=PCA$rotation[,1],PC2=PCA$rotation[,2],PC3=PCA$rotation[,3],
                  PC4=PCA$rotation[,4],PC5=PCA$rotation[,5])
PCs

##    indx       PC1        PC2         PC3        PC4         PC5
## 1:    1 0.4612263  0.7261333 -0.46775954  0.1928963  0.06318796
## 2:    2 0.4586389  0.2216448  0.49839055 -0.6075119 -0.35079935
## 3:    3 0.4484569 -0.1137787  0.47992589  0.2956614  0.68424871
## 4:    4 0.4388789 -0.3643691  0.01646777  0.5591711 -0.60139585
## 5:    5 0.4280098 -0.5271561 -0.54973433 -0.4400368  0.20754237
```

We see that the first PC explains 99.2% of the total variation in yields.

Let's look at eigenvalues and eigenvectors of the covariance matrix of yields and comapre them with what we get from PCA:

```r
Sigma <- cov(yields[,.(y1,y2,y3,y4,y5)])
eigs <- eigen(Sigma)
eigs

## eigen() decomposition
## $values
## [1] 46.723884863  0.346886164  0.012667254  0.005463963  0.003334349
##
## $vectors
##            [,1]       [,2]        [,3]       [,4]        [,5]
## [1,] -0.4612263  0.7261333 -0.46775954  0.1928963  0.06318796
## [2,] -0.4586389  0.2216448  0.49839055 -0.6075119 -0.35079935
## [3,] -0.4484569 -0.1137787  0.47992589  0.2956614  0.68424871
## [4,] -0.4388789 -0.3643691  0.01646777  0.5591711 -0.60139585
## [5,] -0.4280098 -0.5271561 -0.54973433 -0.4400368  0.20754237

eigs$values[1]/sum(eigs$values[1],eigs$values[2],eigs$values[3],
                   eigs$values[4],eigs$values[5])

## [1] 0.9921781

# same as above (not surprisingly)
```

Let's see how 1- to 5-year zero-coupon yields load on the first 3 PCs (*Level*, *Slope*, and *Curvature*). Recall these are just the first 3 eigenvectors of Σ.

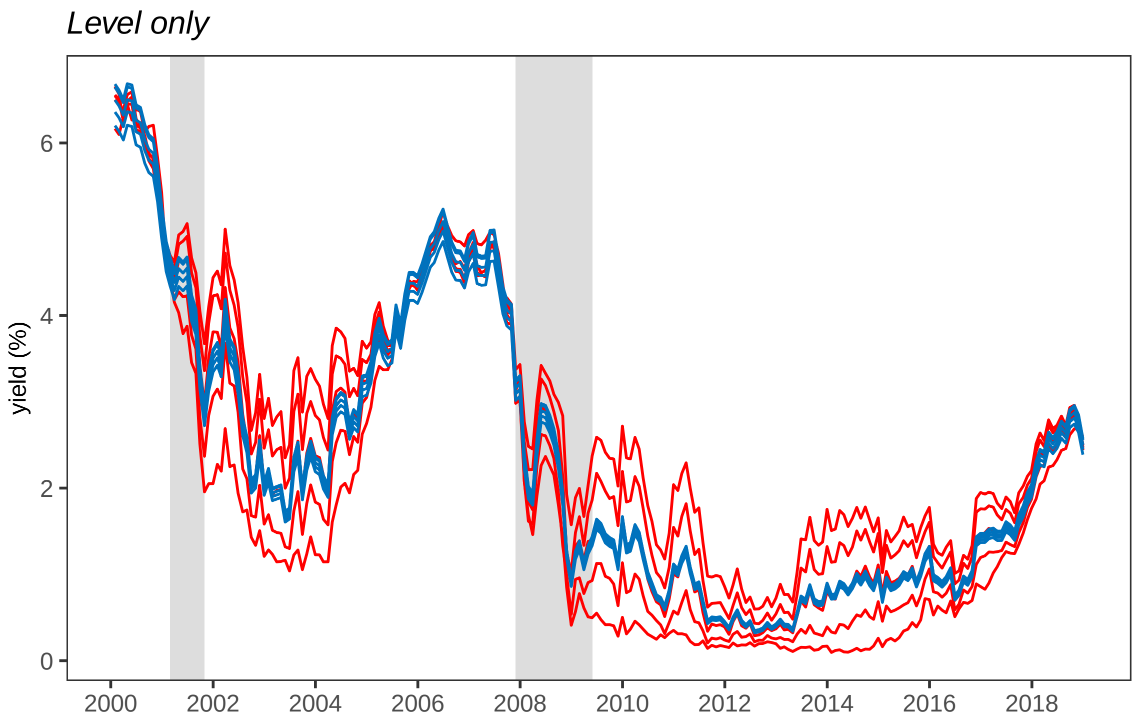


Factor models come from dropping the small eigenvalues, then a larger number of series are driven by a smaller number of factors. For example, what if we drop 4 and 5?

- What if we only have one factor, i.e. *Level*?

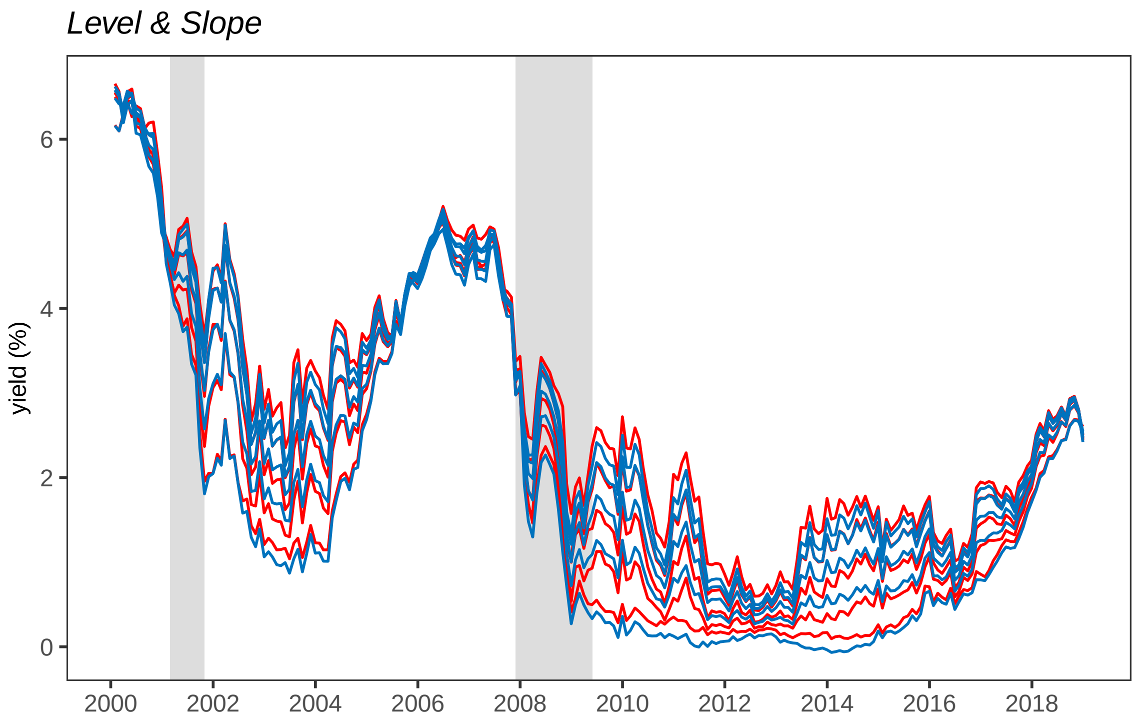
$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \\ y_{4,t} \\ y_{5,t} \end{bmatrix} \approx q_1 \, \text{level}_t$$

In the Figures below, red curves are zero-coupon yields with maturity of 1 to 5 years and the blue curves are the outputs of factor model for the corresponding maturities. The vertical shaded bars indicate NBER recessions.



- What if we have two factors, i.e. *Level* and *Slope*?

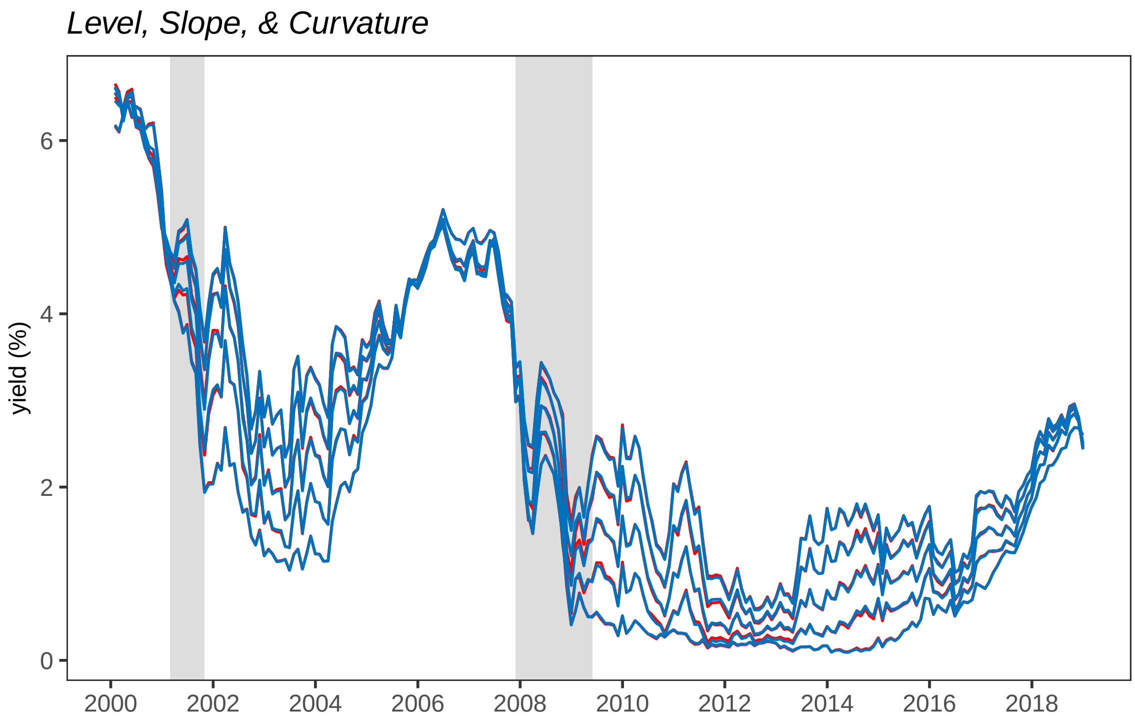
$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \\ y_{4,t} \\ y_{5,t} \end{bmatrix} \approx q_1 \,\text{level}_t + q_2 \,\text{slope}_t$$



- Finally three factors, i.e. *Level*, *Slope*, and *Curvature*:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \\ y_{4,t} \\ y_{5,t} \end{bmatrix} \approx q_1 \,\mathrm{level}_t + q_2 \,\mathrm{slope}_t + q_3 \,\mathrm{curve}_t$$

Almost a perfect fit!



Level, Slope, & Curvature
yield (%)
0
2
4
6
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018

R code

```
library(data.table)
library(zoo)
library(ggplot2)
library(scales)

rm(list=ls())
options(max.print=999999)

yields <- fread("./ZC_FamaBliss.csv")
PCA <- prcomp(~ y1 + y2 + y3 + y4 + y5, data = yields)
summary(PCA)
PCs <- data.table(indx=1:5,PC1=PCA$rotation[,1],PC2=PCA$rotation[,2],PC3=PCA$rotation[,3],
                  PC4=PCA$rotation[,4],PC5=PCA$rotation[,5])
PCs

Sigma <- cov(yields[,.(y1,y2,y3,y4,y5)])
eigs <- eigen(Sigma)
eigs
eigs$values[1]/sum(eigs$values[1],eigs$values[2],eigs$values[3],
                   eigs$values[4],eigs$values[5])

ggplot(PCs) + geom_line(aes(x=indx,y=PC1,color = "PC1"),size=1.0) +
  geom_line(aes(x=indx,y=PC2,color = "PC2"),size=1.0) +
  geom_line(aes(x=indx,y=PC3,color = "PC3"),size=1.0) +
  theme_bw() + xlab("Maturity (year)") + ylab("") + ggtitle("ZC Bond Loadings on PCs") +
  scale_colour_manual("",breaks = c("PC1", "PC2","PC3"),
                      values = c("PC1"= rgb(0,0.447,0.741), "PC2"=rgb(0.85,0.325,0.098),
                                 "PC3"=rgb(0.494,0.184,0.5560)),
                      labels = c("PC1","PC2","PC3")) +
  theme(title = element_text(size=10,face="italic"),
        axis.text.x = element_text(size=9,face="plain"),
        axis.text.y = element_text(size=9,face="plain"),
        axis.title.x = element_text(size=9,face="plain"),
        axis.title.y = element_text(size=9,face="plain"),
        legend.text = element_text(size=7.5,face="plain"),
        legend.position = c(0.9,0.5)) +
  theme(legend.key.size =  unit(0.4, "in"),legend.key.height =  unit(0.2, "in"),
        legend.key.width =  unit(0.4, "in")) +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        legend.background = element_rect(fill=alpha(0.4)))

y <- yields[,.(y1,y2,y3,y4,y5)]
y <- as.matrix(y)
```

```r
# rotation gives the matrix of eigenvectors (Q)
X <- t(PCA$rotation)%*%t(y)
X1 <- as.matrix(X[1,])
X2 <- as.matrix(X[2,])
X3 <- as.matrix(X[3,])
T <- length(y[,1])
F1 <- X1%*%t(as.matrix(PCA$rotation[,1]))
F2 <- X1%*%t(as.matrix(PCA$rotation[,1]))+X2%*%t(as.matrix(PCA$rotation[,2]))
F3 <- X1%*%t(as.matrix(PCA$rotation[,1]))+X2%*%t(as.matrix(PCA$rotation[,2]))+
  X3%*%t(as.matrix(PCA$rotation[,3]))
data <- data.table(date=yields$date,y1=y[,1],y2=y[,2],y3=y[,3],y4=y[,4],y5=y[,5],
                   F1=F1,F2=F2,F3=F3)

recessions.dt <- fread("./NBER_recessions.csv",colClasses=c('Date', 'Date'),
                       header=TRUE,na.strings = "..")
recessions.dt[,`:=`(Peak=as.Date(Peak, "%Y-%m-%d"),Trough=as.Date(Trough, "%Y-%m-%d"))]
recessions.trim  <- recessions.dt[Trough>=min(data[year(date)>2000]$date)]

ggplot() +
  geom_rect(data=recessions.trim, aes(xmin=Peak, xmax=Trough,ymin=-Inf, ymax=Inf), alpha=0.2) +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=y1),color="red") +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=y2),color="red") +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=y3),color="red") +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=y4),color="red") +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=y5),color="red") +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=F1.y1),color=rgb(0,0.447,0.741)) +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=F1.y2),color=rgb(0,0.447,0.741)) +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=F1.y3),color=rgb(0,0.447,0.741)) +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=F1.y4),color=rgb(0,0.447,0.741)) +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=F1.y5),color=rgb(0,0.447,0.741)) +
  theme_bw() + ylab("yield (%)") + xlab("") + ggtitle("Level only") +
  scale_x_date(breaks = seq(as.Date("2000-01-01"), as.Date("2018-12-31"), by="2 years"),
               labels=date_format("%Y")) +
  theme(title = element_text(size=10,face="italic"),
        axis.text.x = element_text(size=9,face="plain"),
        axis.text.y = element_text(size=9,face="plain"),
        axis.title.x = element_text(size=9,face="plain"),
        axis.title.y = element_text(size=9,face="plain"),
        legend.text = element_text(size=7.5,face="plain"),
        legend.position = c(0.9,0.5)) +
  theme(legend.key.size =  unit(0.4, "in"),legend.key.height =  unit(0.2, "in"),
        legend.key.width =  unit(0.4, "in")) +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        legend.background = element_rect(fill=alpha(0.4)))
```

```
ggplot() +
  geom_rect(data=recessions.trim, aes(xmin=Peak, xmax=Trough,ymin=-Inf, ymax=Inf), alpha=0.2) +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=y1),color="red") +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=y2),color="red") +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=y3),color="red") +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=y4),color="red") +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=y5),color="red") +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=F2.y1),color=rgb(0,0.447,0.741)) +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=F2.y2),color=rgb(0,0.447,0.741)) +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=F2.y3),color=rgb(0,0.447,0.741)) +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=F2.y4),color=rgb(0,0.447,0.741)) +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=F2.y5),color=rgb(0,0.447,0.741)) +
  theme_bw() + ylab("yield (%)") + xlab("") + ggtitle("Level & Slope") +
  scale_x_date(breaks = seq(as.Date("2000-01-01"), as.Date("2018-12-31"),
                            by="2 years"), labels=date_format("%Y")) +
  theme(title = element_text(size=10,face="italic"),
        axis.text.x = element_text(size=9,face="plain"),
        axis.text.y = element_text(size=9,face="plain"),
        axis.title.x = element_text(size=9,face="plain"),
        axis.title.y = element_text(size=9,face="plain"),
        legend.text = element_text(size=7.5,face="plain"),
        legend.position = c(0.9,0.5)) +
  theme(legend.key.size =  unit(0.4, "in"),legend.key.height =
          unit(0.2, "in"),legend.key.width =  unit(0.4, "in")) +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        legend.background = element_rect(fill=alpha(0.4)))

ggplot() +
  geom_rect(data=recessions.trim, aes(xmin=Peak, xmax=Trough,ymin=-Inf, ymax=Inf), alpha=0.2) +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=y1),color="red") +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=y2),color="red") +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=y3),color="red") +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=y4),color="red") +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=y5),color="red") +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=F3.y1),color=rgb(0,0.447,0.741)) +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=F3.y2),color=rgb(0,0.447,0.741)) +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=F3.y3),color=rgb(0,0.447,0.741)) +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=F3.y4),color=rgb(0,0.447,0.741)) +
  geom_line(data=data[year(date)>=2000],aes(x=as.Date(date),y=F3.y5),color=rgb(0,0.447,0.741)) +
  theme_bw() + ylab("yield (%)") + xlab("") + ggtitle("Level, Slope, & Curvature") +
  scale_x_date(breaks = seq(as.Date("2000-01-01"), as.Date("2018-12-31"),
                            by="2 years"), labels=date_format("%Y")) +
  theme(title = element_text(size=10,face="italic"),
        axis.text.x = element_text(size=9,face="plain"),
```

```
      axis.text.y = element_text(size=9,face="plain"),
      axis.title.x = element_text(size=9,face="plain"),
      axis.title.y = element_text(size=9,face="plain"),
      legend.text = element_text(size=7.5,face="plain"),
      legend.position = c(0.9,0.5)) +
theme(legend.key.size =  unit(0.4, "in"),legend.key.height =  unit(0.2, "in"),
      legend.key.width =  unit(0.4, "in")) +
theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
      legend.background = element_rect(fill=alpha(0.4)))
```