

Tipos de interés

Matias Corredoira

8/11/2020

Esquema de la tarea requerida:

- Parte realizada anteriormente
- Análisis de correlaciones
- Análisis de componentes principales
- Varimax
- Predicción del bono a 10 años
- Conclusiones

Parte Realizada anteriormente

Cargamos algunas de las librerías más básicas que nos resultarán necesarias

```
library(factoextra)
library(FactoMineR)
library(readr)
library(MASS)
library(janitor)
library(magrittr)
```

Carga de la base de datos y visión Gral

```
TIUSD <- read_delim("~/CUNEF/Reduccion dimension/entrega_1/entrega_reduccion_bonos/ACPTIUSD.csv",
                  ";", escape_double = FALSE, trim_ws = TRUE)
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
##
## -- Column specification -----
## cols(
##   X1 = col_character(),
##   `DEPO 1M` = col_double(),
##   `DEPO 3M` = col_double(),
##   `DEPO 6M` = col_double(),
##   `DEPO 12M` = col_double(),
##   `IRS 2Y` = col_double(),
##   `IRS 3Y` = col_double(),
##   `IRS 4Y` = col_double(),
##   `IRS 5Y` = col_double(),
##   `IRS 7Y` = col_double(),
##   `IRS 10Y` = col_double()
## )
```

```
View(TIUSD)
```

Comprobamos tanto las primeras como las últimas observaciones

```
## # A tibble: 6 x 11
##   X1      `DEPO 1M` `DEPO 3M` `DEPO 6M` `DEPO 12M` `IRS 2Y` `IRS 3Y` `IRS 4Y`
##   <chr>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 02/0~         6         6.5         7         7.75      8.17      8.24      8.25
## 2 03/0~         5.94        6.5         7         7.81      8.22      8.28      8.29
## 3 04/0~         5.94        6.5         7         7.81      8.13      8.2       8.23
## 4 05/0~         5.90        6.44        6.94        7.69      8.14      8.22      8.26
## 5 06/0~         5.88        6.44        6.94        7.69      8.07      8.2       8.25
## 6 09/0~         5.88        6.38        6.88        7.69      8.14      8.25      8.29
## # ... with 3 more variables: `IRS 5Y` <dbl>, `IRS 7Y` <dbl>, `IRS 10Y` <dbl>
```

```
## # A tibble: 6 x 11
##   X1      `DEPO 1M` `DEPO 3M` `DEPO 6M` `DEPO 12M` `IRS 2Y` `IRS 3Y` `IRS 4Y`
##   <chr>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 23/0~      NA        5.5        5.41        5.25        5.11        5.14        5.22
## 2 24/0~      NA        5.44        5.34        5.16        5.03        5.08        5.16
## 3 25/0~      NA        5.31        5.25        5.06        4.93        4.97        5.07
## 4 28/0~      NA        5.31        5.25        5.06        4.92        4.96        5.04
## 5 29/0~      NA        5.31        5.25        5.06        4.94        4.99        5.08
## 6 30/0~      NA        5.31        5.25        5.06        4.82        4.86        4.92
## # ... with 3 more variables: `IRS 5Y` <dbl>, `IRS 7Y` <dbl>, `IRS 10Y` <dbl>
```

ANÁLISIS EXPLORATORIO

Transformaciones pertinentes para esta parte dentro de la base de datos

```
library(reshape2)

TIUSD2 = TIUSD[complete.cases(TIUSD), ]
TIUSD2$Fechas = as.Date(TIUSD2$X1, format = "%d/%m/%Y")

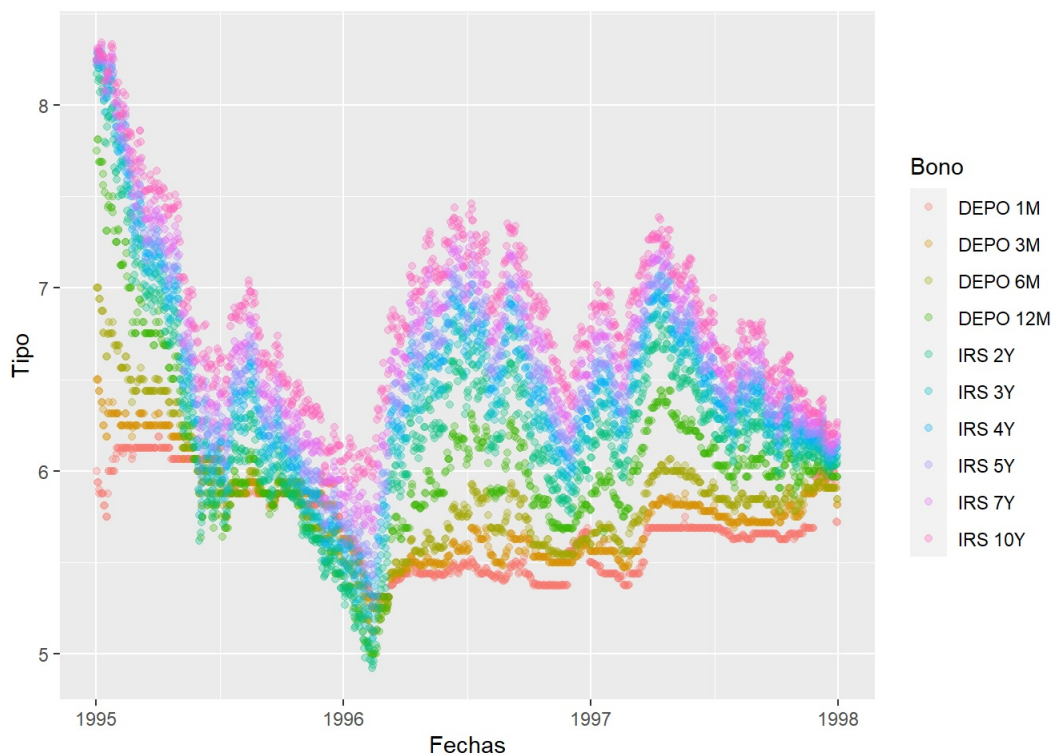
TIUSD2=TIUSD2[,2:12]
TIUSD2
```

```
## # A tibble: 783 x 11
##   `DEPO 1M` `DEPO 3M` `DEPO 6M` `DEPO 12M` `IRS 2Y` `IRS 3Y` `IRS 4Y` `IRS 5Y`
##   <dbl>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
## 1      6      6.5      7      7.75      8.17      8.24      8.25      8.22
## 2    5.94      6.5      7      7.81      8.22      8.28      8.29      8.29
## 3    5.94      6.5      7      7.81      8.13      8.2      8.23      8.24
## 4    5.90      6.44      6.94      7.69      8.14      8.22      8.26      8.27
## 5    5.88      6.44      6.94      7.69      8.07      8.2      8.25      8.27
## 6    5.88      6.38      6.88      7.69      8.14      8.25      8.29      8.31
## 7    5.88      6.38      6.88      7.69      8.09      8.21      8.25      8.25
## 8    5.88      6.34      6.88      7.62      8.07      8.2      8.23      8.24
## 9    5.81      6.25      6.76      7.56      8.02      8.15      8.19      8.22
## 10   5.81      6.25      6.75      7.52      7.8      7.96      8.03      8.05
## # ... with 773 more rows, and 3 more variables: `IRS 7Y` <dbl>, `IRS
## #   10Y` <dbl>, Fechas <date>
```

La Función melt de reshape2: “estira” el data frame

```
#Elegimos entre distintos tipos de representación.

data_long = melt(TIUSD2, id="Fechas")
ggplot(data=data_long, aes(x= Fechas, y=value, color=variable)) +
  #geom_line()
  geom_point(alpha = 0.3, position = position_jitter()) + #stat_smooth(method = "lm") +
  labs(y = "Tipo", colour="Bono")
```



Modificamos de nuevo para poder trabajar mejor los datos

```
TIUSD.act=TIUSD[1:949, 1:9]
head(TIUSD.act)
```

```
## # A tibble: 6 x 9
##   X1      `DEPO 1M` `DEPO 3M` `DEPO 6M` `DEPO 12M` `IRS 2Y` `IRS 3Y` `IRS 4Y`
##   <chr>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 02/0~        6        6.5        7        7.75      8.17      8.24      8.25
## 2 03/0~      5.94      6.5        7        7.81      8.22      8.28      8.29
## 3 04/0~      5.94      6.5        7        7.81      8.13      8.2      8.23
## 4 05/0~      5.90      6.44      6.94      7.69      8.14      8.22      8.26
## 5 06/0~      5.88      6.44      6.94      7.69      8.07      8.2      8.25
## 6 09/0~      5.88      6.38      6.88      7.69      8.14      8.25      8.29
## # ... with 1 more variable: `IRS 5Y` <dbl>
```

```
str(TIUSD.act)
```

```
## tibble [949 x 9] (S3: tbl_df/tbl/data.frame)
## $ X1      : chr [1:949] "02/01/1995" "03/01/1995" "04/01/1995" "05/01/1995" ...
## $ DEPO 1M : num [1:949] 6 5.94 5.94 5.9 5.88 ...
## $ DEPO 3M : num [1:949] 6.5 6.5 6.5 6.44 6.44 ...
## $ DEPO 6M : num [1:949] 7 7 7 6.94 6.94 ...
## $ DEPO 12M: num [1:949] 7.75 7.81 7.81 7.69 7.69 ...
## $ IRS 2Y  : num [1:949] 8.17 8.22 8.13 8.14 8.07 ...
## $ IRS 3Y  : num [1:949] 8.24 8.28 8.2 8.22 8.2 8.25 8.21 8.2 8.15 7.96 ...
## $ IRS 4Y  : num [1:949] 8.25 8.29 8.23 8.26 8.25 8.29 8.25 8.23 8.19 8.03 ...
## $ IRS 5Y  : num [1:949] 8.22 8.29 8.24 8.27 8.27 8.31 8.25 8.24 8.22 8.05 ...
```

```
Dates=as.Date(TIUSD.act$X1, format = "%d/%m/%y") # Vector para los días
```

```
TIUSD.act=TIUSD.act[, -1] # Las fechas decidimos no incluirlas
str(Dates) # Estructura
```

```
## Date[1:949], format: "2019-01-02" "2019-01-03" "2019-01-04" "2019-01-05" "2019-01-06" ...
```

```
summary(TIUSD.act) # Estadísticos básicos
```

```
##      DEPO 1M      DEPO 3M      DEPO 6M      DEPO 12M
## Min.    :5.313   Min.    :5.250   Min.    :5.121   Min.    :4.996
## 1st Qu.:5.445   1st Qu.:5.570   1st Qu.:5.656   1st Qu.:5.793
## Median :5.656   Median :5.695   Median :5.781   Median :5.906
## Mean    :5.689   Mean    :5.749   Mean    :5.822   Mean    :5.986
## 3rd Qu.:5.875   3rd Qu.:5.875   3rd Qu.:5.906   3rd Qu.:6.094
## Max.    :6.188   Max.    :6.500   Max.    :7.000   Max.    :7.813
## NA's    :166
##      IRS 2Y      IRS 3Y      IRS 4Y      IRS 5Y
## Min.    :4.920   Min.    :5.080   Min.    :5.260   Min.    :5.430
## 1st Qu.:5.865   1st Qu.:5.955   1st Qu.:6.015   1st Qu.:6.055
## Median :6.085   Median :6.210   Median :6.320   Median :6.405
## Mean    :6.149   Mean    :6.277   Mean    :6.374   Mean    :6.454
## 3rd Qu.:6.320   3rd Qu.:6.490   3rd Qu.:6.620   3rd Qu.:6.720
## Max.    :8.220   Max.    :8.280   Max.    :8.290   Max.    :8.310
##
```

Análisis de correlaciones

Optamos por eliminar las filas que contienen valores NA

```
cor.mat = round(cor(TIUSD.act),2)

cor.mat = round(cor(TIUSD.act, use="complete.obs"),2)
cor.mat
```

```
##      DEPO 1M DEPO 3M DEPO 6M DEPO 12M IRS 2Y IRS 3Y IRS 4Y IRS 5Y
## DEPO 1M    1.00   0.92   0.74    0.51   0.35   0.31   0.28   0.26
## DEPO 3M    0.92   1.00   0.93    0.76   0.63   0.58   0.55   0.53
## DEPO 6M    0.74   0.93   1.00    0.94   0.86   0.83   0.81   0.79
## DEPO 12M   0.51   0.76   0.94    1.00   0.98   0.96   0.95   0.93
## IRS 2Y     0.35   0.63   0.86    0.98   1.00   1.00   0.99   0.98
## IRS 3Y     0.31   0.58   0.83    0.96   1.00   1.00   1.00   0.99
## IRS 4Y     0.28   0.55   0.81    0.95   0.99   1.00   1.00   1.00
## IRS 5Y     0.26   0.53   0.79    0.93   0.98   0.99   1.00   1.00
```

Como es lógico a medida que el bono aumenta en plazo su correlación se va reduciendo dado que estos de modo general se ven menos influidos por motivos presentes por su condición.

Queremos conocer los nds, para ello utilizaremos la librería Hmisc

```
require(Hmisc)

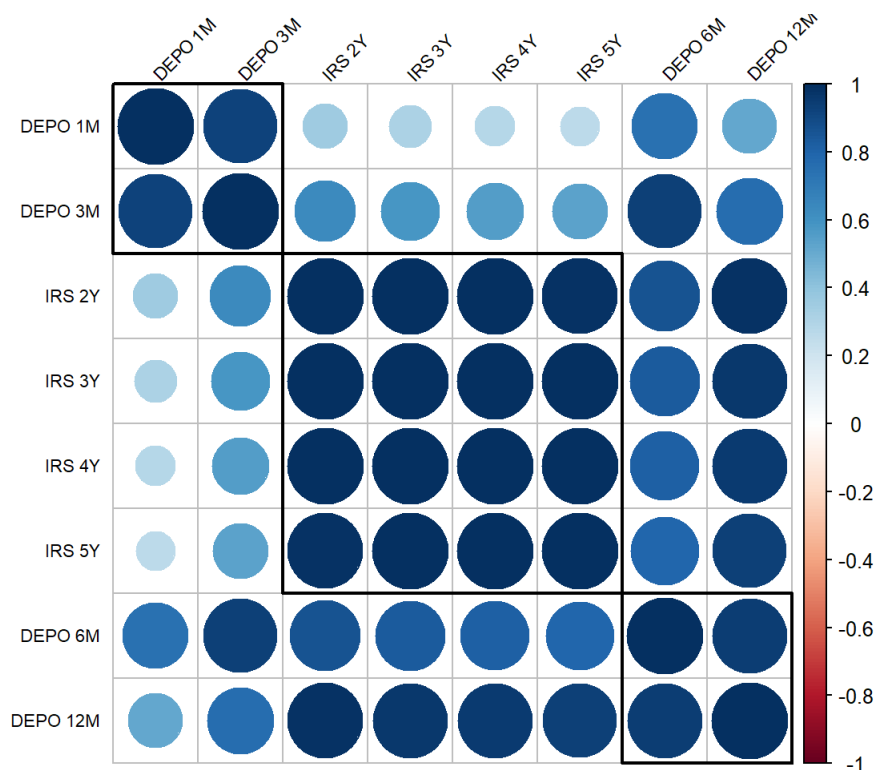
cor.mat.nds= rcorr(as.matrix(TIUSD.act))
cor.mat.nds
```

```
##          DEPO 1M DEPO 3M DEPO 6M DEPO 12M IRS 2Y IRS 3Y IRS 4Y IRS 5Y
## DEPO 1M      1.00    0.92    0.74    0.51    0.35    0.31    0.28    0.26
## DEPO 3M      0.92    1.00    0.93    0.76    0.63    0.59    0.56    0.53
## DEPO 6M      0.74    0.93    1.00    0.94    0.86    0.82    0.80    0.77
## DEPO 12M     0.51    0.76    0.94    1.00    0.97    0.95    0.93    0.91
## IRS 2Y       0.35    0.63    0.86    0.97    1.00    0.99    0.98    0.97
## IRS 3Y       0.31    0.59    0.82    0.95    0.99    1.00    1.00    0.99
## IRS 4Y       0.28    0.56    0.80    0.93    0.98    1.00    1.00    1.00
## IRS 5Y       0.26    0.53    0.77    0.91    0.97    0.99    1.00    1.00
##
## n
##          DEPO 1M DEPO 3M DEPO 6M DEPO 12M IRS 2Y IRS 3Y IRS 4Y IRS 5Y
## DEPO 1M      783    783    783    783    783    783    783    783
## DEPO 3M      783    949    949    949    949    949    949    949
## DEPO 6M      783    949    949    949    949    949    949    949
## DEPO 12M     783    949    949    949    949    949    949    949
## IRS 2Y       783    949    949    949    949    949    949    949
## IRS 3Y       783    949    949    949    949    949    949    949
## IRS 4Y       783    949    949    949    949    949    949    949
## IRS 5Y       783    949    949    949    949    949    949    949
##
## P
##          DEPO 1M DEPO 3M DEPO 6M DEPO 12M IRS 2Y IRS 3Y IRS 4Y IRS 5Y
## DEPO 1M      0      0      0      0      0      0      0      0
## DEPO 3M      0      0      0      0      0      0      0      0
## DEPO 6M      0      0      0      0      0      0      0      0
## DEPO 12M     0      0      0      0      0      0      0      0
## IRS 2Y      0      0      0      0      0      0      0      0
## IRS 3Y      0      0      0      0      0      0      0      0
## IRS 4Y      0      0      0      0      0      0      0      0
## IRS 5Y      0      0      0      0      0      0      0      0
```

Realizamos una representación de los clúster, donde podemos distinguir claramente que entre determinados periodos de tiempo se da una diferenciación clara.

```
require(corrplot)

corrplot(cor.mat, type="full", order="hclust", addrect = 3,
         tl.col="black", tl.cex=0.7, tl.srt=45)
```

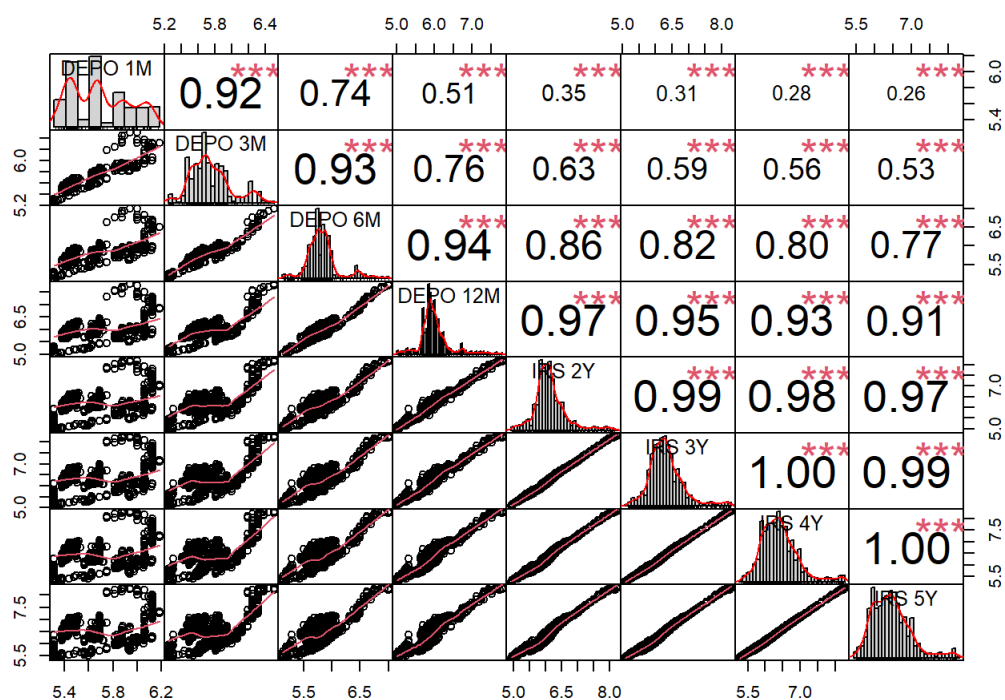


Se puede comprobar como las variables se dividen de forma clara en 3 grupos dependiendo de sus periodos de tiempo. Por ello, consideramos que es lícita la realización de un análisis de componentes principales.

Comprobamos tanto dispersiones, distribuciones por pares además de los coeficientes de las variables.

```
require(PerformanceAnalytics)
```

```
chart.Correlation(TIUSD.act, histogram=TRUE, pch=19)
```



Para KMO nos resultará necesaria la inversa de la matriz de correlaciones

```
invR = solve(cor.mat)
invR
```

```
##          DEPO 1M    DEPO 3M    DEPO 6M    DEPO 12M    IRS 2Y    IRS 3Y
## DEPO 1M    6.553861  -1.608477  -9.987096    9.045637  -25.057705  17.04985
## DEPO 3M   -1.608477  14.121881 -19.096345   -11.621017   86.099853  -49.69194
## DEPO 6M   -9.987096 -19.096345  68.435688   -35.997165  -89.131423   82.42126
## DEPO 12M   9.045637 -11.621017 -35.997165    90.438196   13.085913 -116.81722
## IRS 2Y   -25.057705  86.099853 -89.131423    13.085913  -37.334836   67.34156
## IRS 3Y    17.049854 -49.691936  82.421257 -116.817215   67.341560   66.73452
## IRS 4Y    27.453154 -90.181931  73.390160   27.158721    8.431326   15.92846
## IRS 5Y   -21.150106  73.826357 -75.507533    23.803639  -19.385326  -82.55939
##          IRS 4Y    IRS 5Y
## DEPO 1M    27.453154 -21.15011
## DEPO 3M   -90.181931  73.82636
## DEPO 6M    73.390160 -75.50753
## DEPO 12M   27.158721  23.80364
## IRS 2Y     8.431326 -19.38533
## IRS 3Y    15.928464 -82.55939
## IRS 4Y   -160.793152  94.18404
## IRS 5Y    94.184039  11.43200
```

Realizaremos ahora las correlaciones parciales. Para ello se exige la no existencia de los valores NA por lo que habrá que modificar.

```
require(ppcor)
```

```
TIUSD.act.C=TIUSD.act[complete.cases(TIUSD.act),]
p.cor.mat=pcor(TIUSD.act.C)
```

```
# Realización de las parcia
```

```
les
```

Realizamos KMO

```
library(psych)
```

```
KMO(TIUSD.act)
```

```
## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = TIUSD.act)
## Overall MSA = 0.84
## MSA for each item =
## DEPO 1M DEPO 3M DEPO 6M DEPO 12M IRS 2Y IRS 3Y IRS 4Y IRS 5Y
## 0.80 0.80 0.84 0.91 0.84 0.82 0.82 0.85
```

Podemos ver que nos da un valor de 0.84 lo cual indica que las correlaciones entre las variables es bastante alta mientras que las correlaciones parciales son bajas, ya que en esta operación se divide entre las parciales. Con esto, concluimos también que la realización de un análisis factorial sobre estas variables sería mas que razonable.

```
cortest.bartlett(TIUSD.act)
```

```
## $chisq
## [1] 31782.42
##
## $p.value
## [1] 0
##
## $df
## [1] 28
```

Queremos saber si la matriz de correlaciones es identidad por lo que realizamos este test. Viendo los resultados que nos ofrece la chi-cuadrado (Que en este caso es muy alto) nos permite concluir que se rechaza la hipótesis nula, es decir, hay relación entre las variables. Por ello extraemos también que la realización de ACP es lícita.

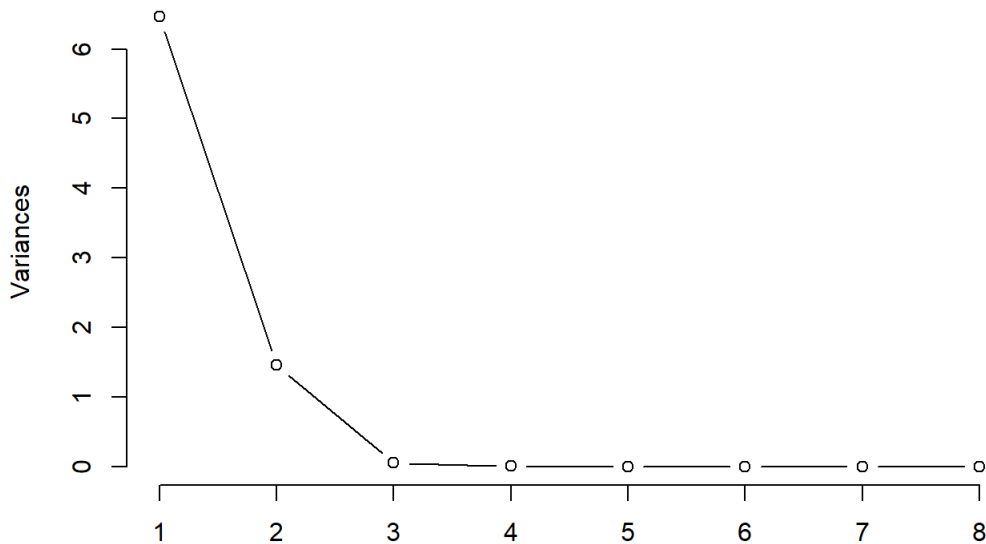
Análisis de componentes principales

```
acp <- prcomp(TIUSD.act.C,
              center = TRUE, scale = TRUE) #Escalamos para evitar problemas con los pesos
acp
```

```
## Standard deviations (1, ..., p=8):
## [1] 2.54195802 1.20794908 0.24087793 0.11551854 0.06947986 0.05349123 0.01311355
## [8] 0.00900036
##
## Rotation (n x k) = (8 x 8):
## PC1 PC2 PC3 PC4 PC5
## DEPO 1M -0.2224648 -0.67165893 -0.58679440 0.38019372 0.10209777
## DEPO 3M -0.3129580 -0.49562942 0.23148694 -0.57622986 -0.44405793
## DEPO 6M -0.3762142 -0.22268754 0.43173386 -0.08508011 0.46280345
## DEPO 12M -0.3900404 0.04378184 0.41806976 0.43616824 0.26850625
## IRS 2Y -0.3810834 0.19944501 0.05265247 0.35050105 -0.56005325
## IRS 3Y -0.3758681 0.24236117 -0.12013900 0.09938056 -0.26330079
## IRS 4Y -0.3712997 0.26766795 -0.26674633 -0.17695781 0.06237662
## IRS 5Y -0.3667877 0.28557973 -0.39057701 -0.40235429 0.34529432
## PC6 PC7 PC8
## DEPO 1M -0.008272761 0.001399062 -0.003795108
## DEPO 3M 0.271202504 -0.005626290 0.003438263
## DEPO 6M -0.633031294 -0.012532451 0.012948214
## DEPO 12M 0.639309249 -0.009688755 -0.004556345
## IRS 2Y -0.279242624 0.473928858 -0.270424166
## IRS 3Y -0.126455772 -0.452361373 0.696939868
## IRS 4Y 0.011535747 -0.531666634 -0.633505674
## IRS 5Y 0.151113366 0.536476160 0.199001283
```

```
plot(acp, type= 'l')
```

acp



Se ve de forma clara como las 2 primeras variables resultarían más que suficiente dado que las mismas en su conjunto son capaces de calcular la gran parte de la varianza, debajo mostramos las cifras.

```
summary(acp)
```

```
## Importance of components:
##               PC1      PC2      PC3      PC4      PC5      PC6      PC7
## Standard deviation  2.5420  1.2079  0.24088  0.11552  0.06948  0.05349  0.01311
## Proportion of Variance 0.8077  0.1824  0.00725  0.00167  0.00060  0.00036  0.00002
## Cumulative Proportion 0.8077  0.9901  0.99734  0.99901  0.99961  0.99997  0.99999
##               PC8
## Standard deviation   9e-03
## Proportion of Variance 1e-05
## Cumulative Proportion 1e+00
```

Entre esas 2 variables ya se consigue explicar un 99.01 % de la varianza explicada por lo que se decide optar por reducir a solamente a esas 2 para tener así un modelo efectivo y a la vez lo más simplificado posible.

Varimax

En primer lugar realizaremos la rotación entre variables

Aquí buscamos la explicación que nos puede proporcionar cada una de las dimensiones de las variables

```
act.R <- principal(TIUSD.act, rotate='varimax', nfactors=2,scores=TRUE)
act.R
```



```
## Principal Components Analysis
## Call: principal(r = TIUSD.act, nfactors = 2, rotate = "varimax", scores = TRUE)
## Standardized loadings (pattern matrix) based upon correlation matrix
##
##      RC1  RC2   h2    u2 com
## DEPO 1M  0.07 0.99 0.98 0.0239 1.0
## DEPO 3M  0.39 0.92 0.99 0.0082 1.3
## DEPO 6M  0.68 0.72 0.98 0.0174 2.0
## DEPO 12M 0.87 0.47 0.98 0.0241 1.5
## IRS 2Y   0.95 0.29 0.99 0.0059 1.2
## IRS 3Y   0.97 0.24 1.00 0.0014 1.1
## IRS 4Y   0.98 0.20 0.99 0.0083 1.1
## IRS 5Y   0.97 0.17 0.98 0.0217 1.1
##
##
##      RC1  RC2
## SS loadings      5.12 2.77
## Proportion Var    0.64 0.35
## Cumulative Var    0.64 0.99
## Proportion Explained 0.65 0.35
## Cumulative Proportion 0.65 1.00
##
## Mean item complexity = 1.3
## Test of the hypothesis that 2 components are sufficient.
##
## The root mean square of the residuals (RMSR) is 0.01
## with the empirical chi square 5.35 with prob < 0.97
##
## Fit based upon off diagonal values = 1
```

Se puede ver como los bonos que son de plazo más corto vienen explicados por la dimensión 2 mientras que los que son de plazo más largo lo hacen por la 1. Finalmente los de 6M y 12M se ve que se explican por las 2, sin embargo en los de 6 M vemos que la explicación de las 2 dimensiones está mucho más equilibrada.

Predicción del bono a 10 años

Realizamos las modificaciones pertinentes en los datos

```
TIUSD %<>% clean_names()
```

Dividimos la base de datos en 2 partes, training, a la que daremos un porcentaje de datos mucho mayor y test.

```
library(imputeTS)
TIUSD_NA <- na_mean(TIUSD)

training <- TIUSD_NA[1:949, 2:11]
test <- TIUSD_NA[950:978, 2:11]
```

La predicción la realizamos para un modelo de 2 dimensiones como hemos concluido anteriormente

```
library(pls)
modelo_pcr <- pcr(irs_10y ~ ., data = training, scale. = TRUE, ncomp = 2)
```

Ya habiendo modelizado el modelo realizamos la predicción con el mismo

```
prediccion_pcr <- predict(modelo_pcr, newdata = test, ncomp = 2)
test_pcr_mse <- mean((prediccion_pcr - test$irs_10y)^2)
test_pcr_mse
```

```
## [1] 0.003004387
```

Podemos comprobar que en este caso la predicción nos ofrece bastante fiabilidad.

Conclusiones

- Observando las correlaciones entre las distintas variables hemos podido ver que se pueden distinguir 3 grupos dentro de las mismas las cuales se separan en función del periodo de tiempo que trabajan.
- KMO también nos da soluciones que nos hacen concluir que la relación entre las variables es muy clara.
- Lads variables que trabajan un periodo de tiempo más breves se ven mucho más afectadas por la 2º dimensión. A medida que

aumenta el periodo de tiempo de los bonos, se van viendo cada vez más afectados por la 1ª dimensión

- El modelo de predicción realizado para un plazo de 10 años nos da una fiabilidad considerable.