

# Estructura temporal subyacente de los tipos de interés

Sara Bengoechea Rodríguez

11/8/2020

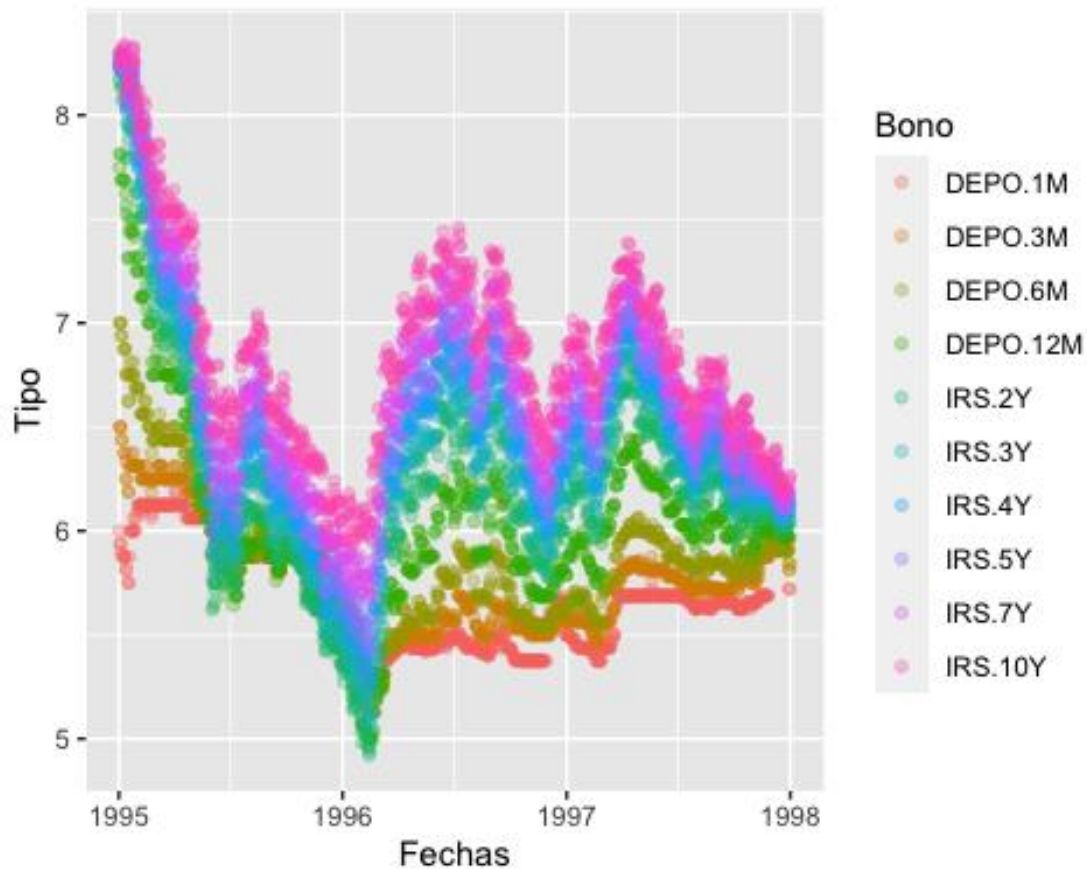
## Introducción

El objetivo de este trabajo es realizar una comprobación empírica del Análisis de Componentes Principales a un conjunto de 978 observaciones de los rendimientos de 10 bonos norteamericanos a distintos plazos entre el 2 de enero de 1995 y el 30 de septiembre de 1998.

## Análisis exploratorio y estudio de idoneidad del ACP

### Distribución

Mediante el siguiente gráfico observamos la distribución de las variables.



## Summary

Visualizamos un resumen de los estadísticos más relevantes de las variables de nuestro dataset.

```
##      Min   Q1  Med Mean   SD   Q3  Max
## DEPO.1M  5.3  5.4  5.7   5.7   NA  5.9  6.2
## DEPO.3M  5.2  5.6  5.7   5.7  0.2  5.9  6.5
## DEPO.6M  5.1  5.7  5.8   5.8  0.3  5.9  7.0
## DEPO.12M 5.0  5.8  5.9   6.0  0.4  6.1  7.8
## IRS.2Y   4.9  5.9  6.1   6.1  0.5  6.3  8.2
## IRS.3Y   5.1  6.0  6.2   6.3  0.5  6.5  8.3
## IRS.4Y   5.3  6.0  6.3   6.4  0.5  6.6  8.3
## IRS.5Y   5.4  6.1  6.4   6.5  0.5  6.7  8.3
```

## Análisis de la matriz de correlación

La primera matriz indica el grado de correlación que hay entre las variables, que en este caso es alto, y la segunda matriz informa el número de observaciones. Por último, al observar la última matriz, no podemos decir que haya incorrelación entre las variables ya que se rechaza la H por  $nds = 0$ .

```
##      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
## DEPO.3M      0           0      0      0      0      0      0
## DEPO.6M      0      0           0      0      0      0      0
## DEPO.12M     0      0      0           0      0      0      0
## IRS.2Y       0      0      0      0           0      0      0
## IRS.3Y       0      0      0      0      0           0      0
```

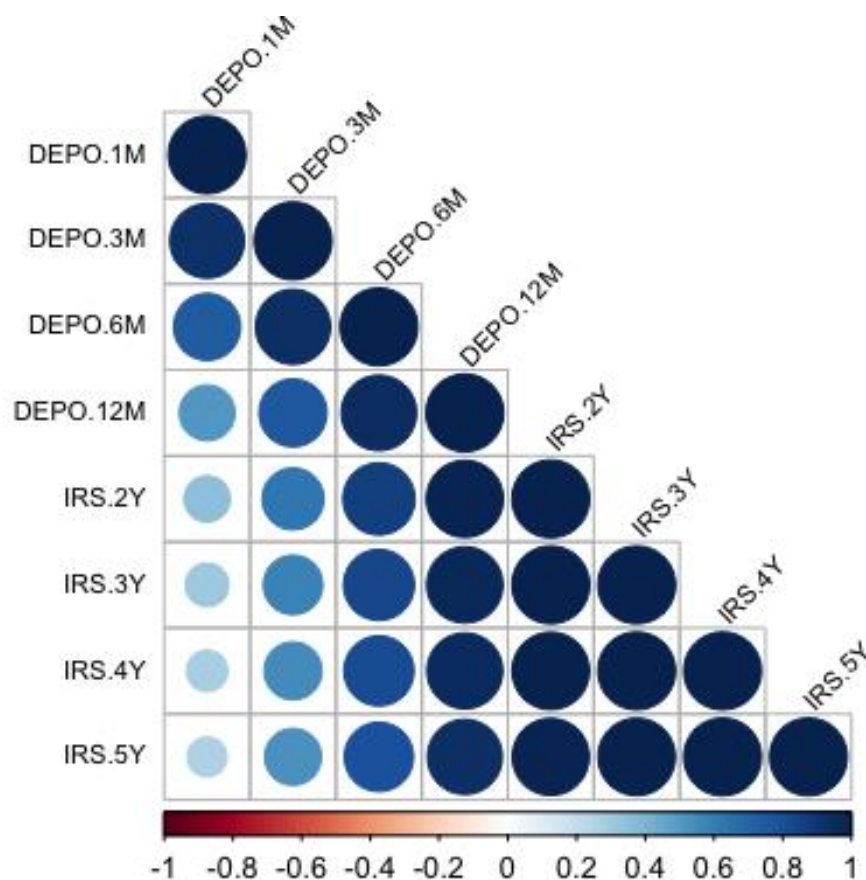
```
## IRS.4Y    0      0      0      0      0      0      0
## IRS.5Y    0      0      0      0      0      0      0
```

El siguiente paso es obtener el determinante de la matriz de correlaciones. Dado que dicho resultado está muy cercano a cero, significa que existe mucha asociación entre las variables del dataset.

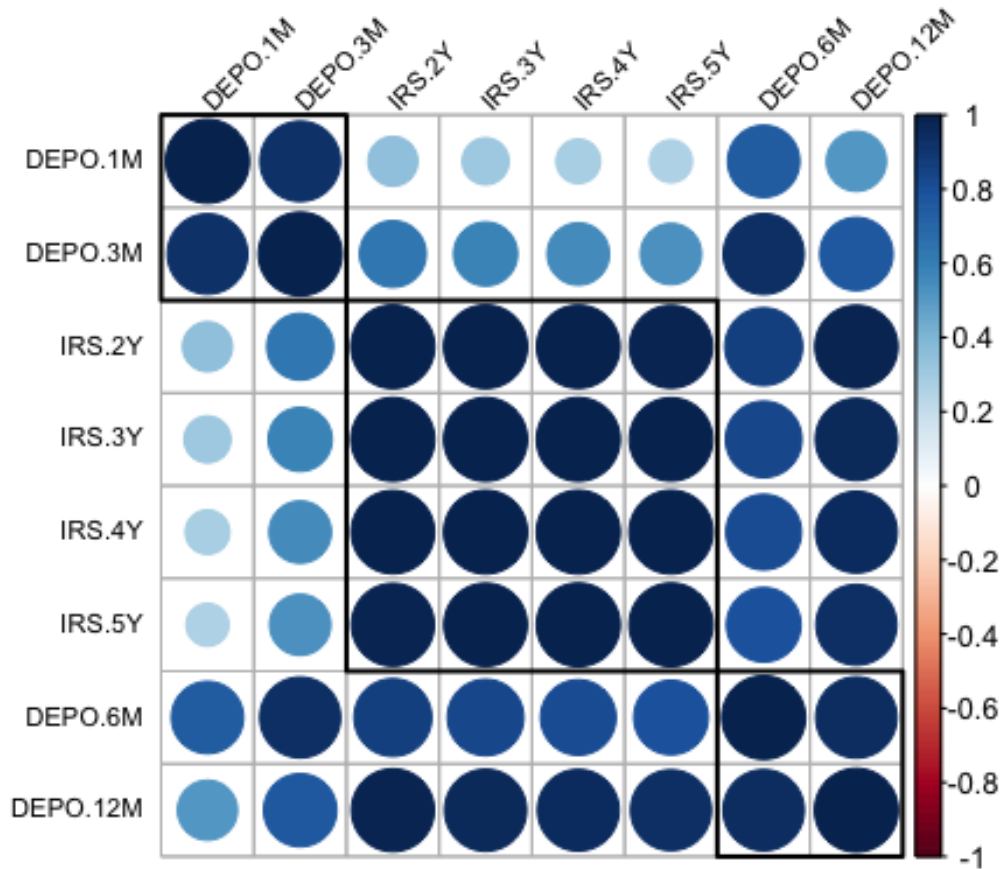
```
## [1] 5.5021e-12
```

## Correlogramas

A continuación se muestra un correlograma para así poder apreciar de manera más visual qué variables están más correlacionadas entre sí. Lo más destacable es la correlación existente entre las variables que se encuentran entre DEPO.12M e IRS.5Y.

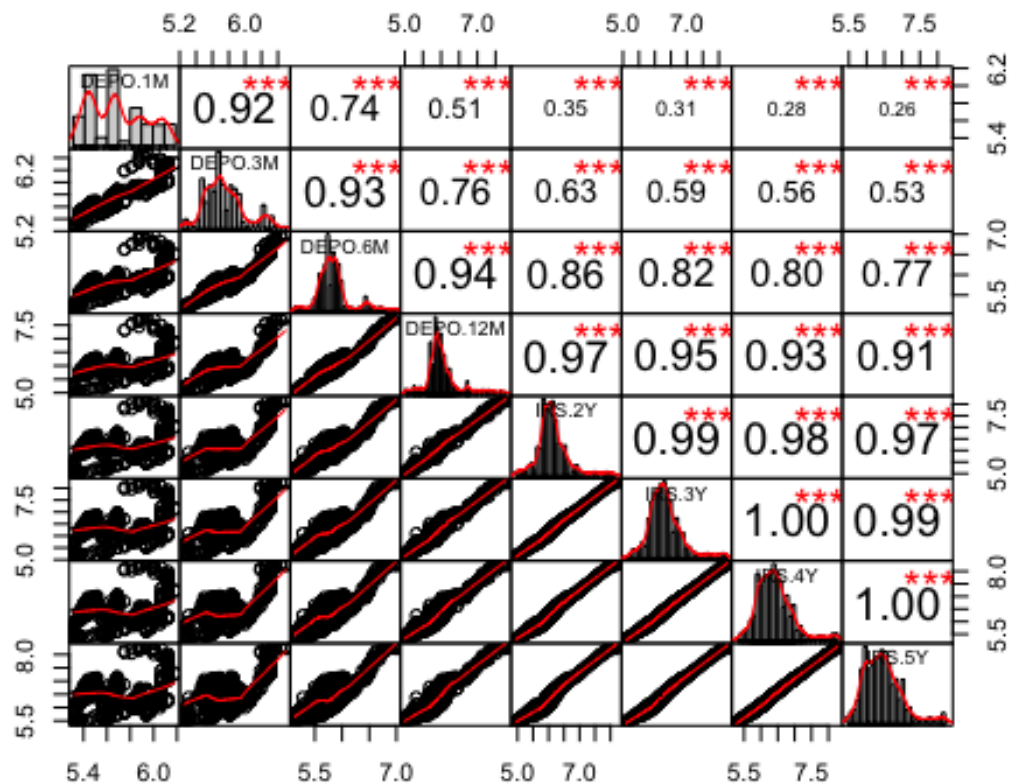


Por otro lado, si estamos interesados en visualizar clusters, utilizaríamos el siguiente correlograma:



Por último, podemos estudiar la correlación de manera más completa con el siguiente chart.

En la diagonal podemos observar la distribución de cada variable. Por debajo de la diagonal, se encuentran diagramas de dispersión por pares con línea de ajuste. Con ello se puede apreciar si existe una relación lineal o no para cada caso. Por encima de la diagonal aparecen los valores del coeficiente de correlación y mediante estrellas, indica el p-valor, siendo a mayor número de estrellas, menor p-valor. Lo que indica que se rechazaría la hipótesis nula, como se ha explicado anteriormente.



### Esfericidad de Barlett

Realizamos la prueba de esfericidad de Bartlett para verificar la idoneidad del ACP. Puesto que el p.valor es muy bajo y  $\chi^2$  muy alto, se rechaza la hipótesis nula, lo que indica la presencia de asociación entre las variables.

En este caso, la no esfericidad es clara, sin embargo no es recomendable llevar a cabo la prueba de esfericidad de Bartlett cuando hay más de 100 observaciones.

```
## R was not square, finding R from data
## $chisq
## [1] 31782.42
##
## $p.value
## [1] 0
##
## $df
## [1] 28
```

### Indice KMO de Kaiser-Meyer-Olkin

El valor obtenido está cercano a 1, lo que implica que la relación entre las variables es alta.

```
## [1] 0.83799
```

### Matriz de Adecuación de la Muestra (MSA)

Similar al KMO, tenemos el MSA, que estudia la relación entre las variables pero de manera más específica. Puesto que todos los valores son altos, tenemos la misma conclusión que anteriormente.

```
##          MSA
## DEPO.1M 0.79261
## DEPO.3M 0.78144
## DEPO.6M 0.84063
## DEPO.12M 0.91650
## IRS.2Y 0.84798
## IRS.3Y 0.81641
## IRS.4Y 0.81881
## IRS.5Y 0.85404
```

### Análisis de componentes principales

Tras el estudio previo realizado, se ha demostrado la alta correlación entre las variables. Es por ello que el Análisis de Componentes Principales es recomendado.

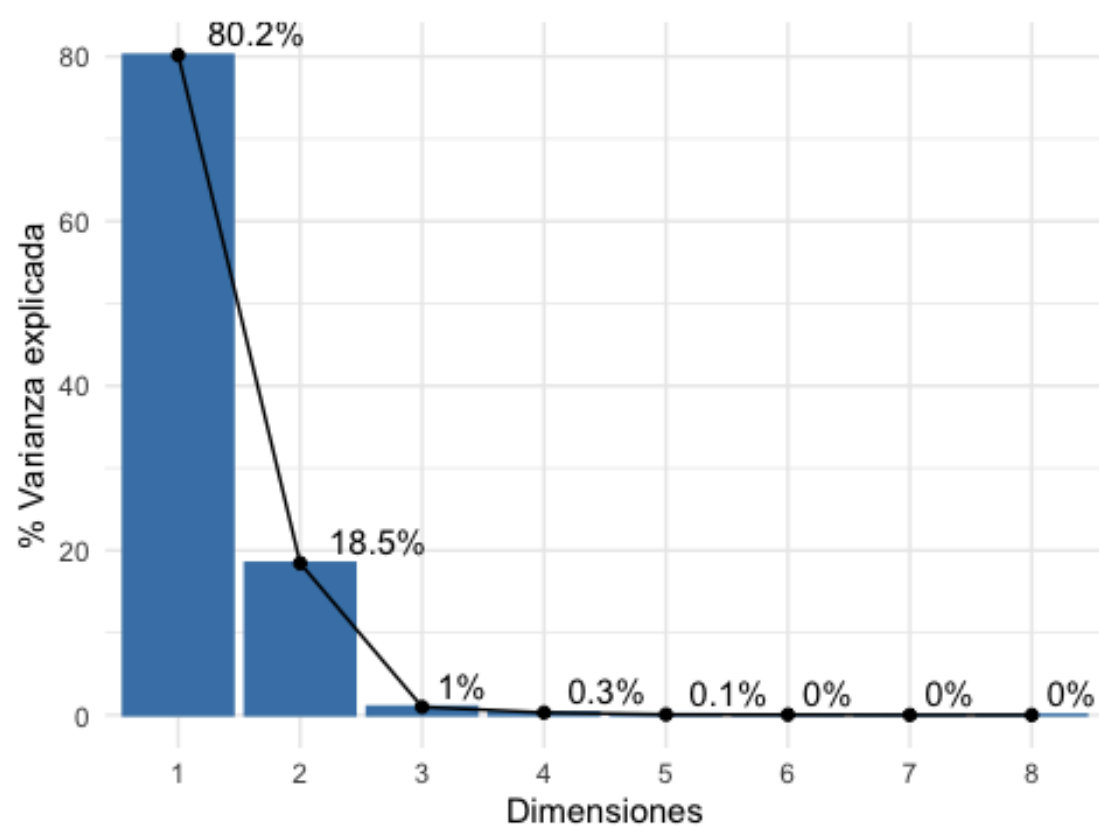
Realizamos el análisis de componentes principales y obtenemos un total de ocho componentes principales. Debemos prestar atención a los autovalores, que aportan información sobre la varianza explicada por cada componente.

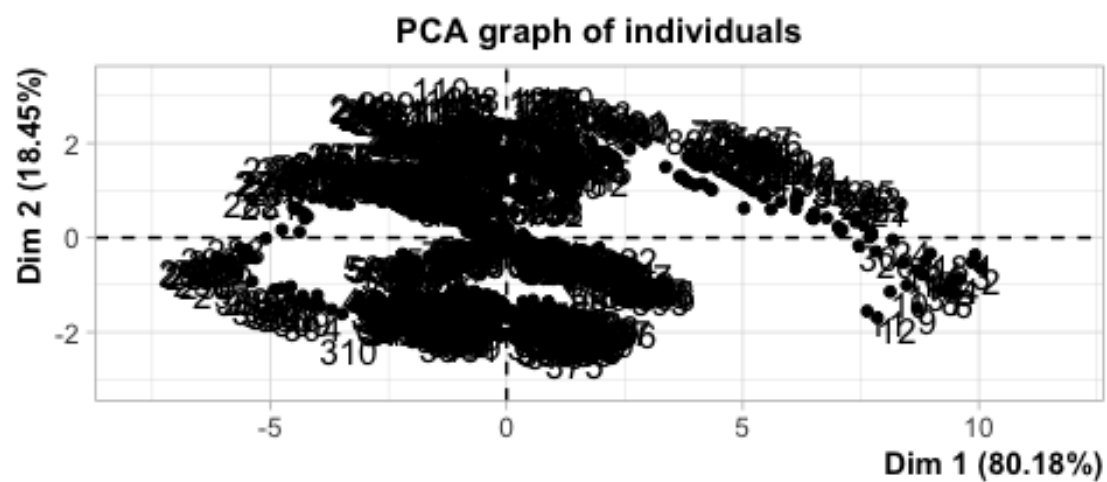
Para explicar el 100% de la varianza necesitaríamos 6 componentes, sin embargo, solo con dos componentes ya explicamos el 98.63% de la varianza.

##	eigenvalue	percentage of variance	cumulative percentage of variance
## comp 1	6.41	80.18	80.18
## comp 2	1.48	18.45	98.63
## comp 3	0.08	0.98	99.61
## comp 4	0.02	0.29	99.90
## comp 5	0.00	0.06	99.96
## comp 6	0.00	0.04	100.00
## comp 7	0.00	0.00	100.00
## comp 8	0.00	0.00	100.00

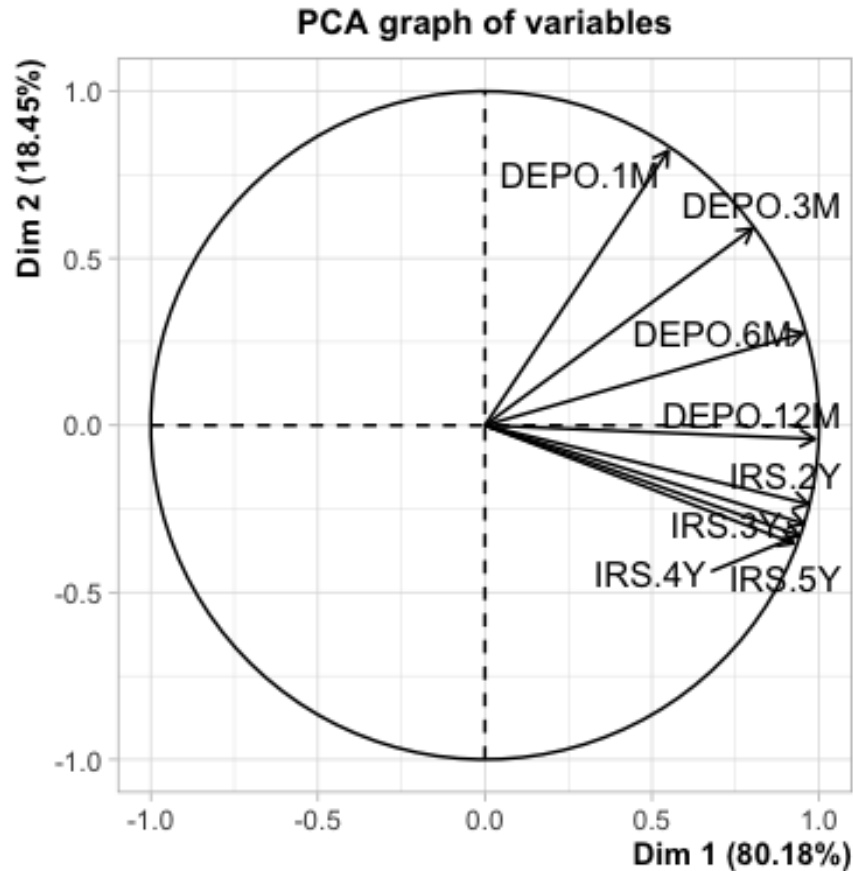
Para comprender de manera más visual el análisis de componentes podemos acudir a los siguientes gráficos.

Gráfico de sedimentación









Como se puede apreciar en las siguientes tabla de coordenadas y contribuciones, DEPO.1M está mejor explicada por la dimensión 2, mientras que las otras están mas explicadas por la dimensión 1.

##	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
## DEPO.1M	0.54862	0.821995	0.130116	0.079753	-0.0066298
## DEPO.3M	0.80096	0.589378	-0.012309	-0.099286	0.0309955
## DEPO.6M	0.95263	0.275756	-0.116548	-0.024351	-0.0348241
## DEPO.12M	0.98747	-0.040796	-0.140687	0.042459	-0.0142635
## IRS.2Y	0.96863	-0.235506	-0.042662	0.051970	0.0381459
## IRS.3Y	0.95491	-0.294376	0.026550	0.019150	0.0169030
## IRS.4Y	0.94017	-0.329298	0.085638	-0.014236	-0.0053927
## IRS.5Y	0.92459	-0.353262	0.133986	-0.041318	-0.0237396

##	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
## DEPO.1M	4.6927	45.77440	21.63083	26.96128	0.92452
## DEPO.3M	10.0020	23.53270	0.19357	41.78506	20.20732
## DEPO.6M	14.1487	5.15151	17.35483	2.51353	25.50781
## DEPO.12M	15.2027	0.11275	25.28811	7.64167	4.27924
## IRS.2Y	14.6280	3.75740	2.32541	11.44865	30.60606
## IRS.3Y	14.2166	5.87070	0.90059	1.55439	6.00950
## IRS.4Y	13.7811	7.34620	9.37011	0.85902	0.61168
## IRS.5Y	13.3282	8.45433	22.93654	7.23640	11.85387

## Rotación de variables subyacentes

Al efectuar una rotación Varimax obtenemos un nuevo modelo mejor, donde los pesos de las variables se diferencian mejor, y los componentes principales explican la varianza de forma más equilibrada.

```
## Factor Analysis using method = minres
## Call: fa(r = TIUSD.act, nfactors = 2, rotate = "varimax", fm = "minres")
## Standardized loadings (pattern matrix) based upon correlation matrix
##           MR1  MR2  h2      u2 com
## DEPO.1M  0.08 0.95 0.92  0.0837 1.0
## DEPO.3M  0.37 0.93 1.00 -0.0036 1.3
## DEPO.6M  0.67 0.73 0.98  0.0175 2.0
## DEPO.12M 0.86 0.48 0.97  0.0310 1.6
## IRS.2Y   0.95 0.30 1.00  0.0048 1.2
## IRS.3Y   0.97 0.24 1.00 -0.0001 1.1
## IRS.4Y   0.97 0.21 0.99  0.0076 1.1
## IRS.5Y   0.97 0.18 0.97  0.0307 1.1
##
##                               MR1  MR2
## SS loadings                   5.07 2.76
## Proportion Var                 0.63 0.34
## Cumulative Var                 0.63 0.98
## Proportion Explained           0.65 0.35
## Cumulative Proportion          0.65 1.00
##
## Mean item complexity = 1.3
## Test of the hypothesis that 2 factors are sufficient.
##
## The degrees of freedom for the null model are 28 and the objective
function was 33.65 with Chi Square of 31782
## The degrees of freedom for the model are 13 and the objective function
was 9.23
##
## The root mean square of the residuals (RMSR) is 0.01
## The df corrected root mean square of the residuals is 0.01
##
## The harmonic number of observations is 904 with the empirical chi square
3.45 with prob < 1
## The total number of observations was 949 with Likelihood Chi Square =
8705.1 with prob < 0
##
## Tucker Lewis Index of factoring reliability = 0.41
## RMSEA index = 0.839 and the 90 % confidence intervals are 0.825 0.855
## BIC = 8616
## Fit based upon off diagonal values = 1
```

## Predicción del valor del bono a 10 años

A continuación predecimos el valor del bono a 10 años con las observaciones activas para las observaciones suplementarias. Las predicciones obtenidas son las que se muestran en la tabla a continuación.

```
## Principal component regression , fitted with the singular value
decomposition algorithm.
## Call:
## pcr(formula = IRS.10Y ~ ., ncomp = 2, data = training_data)

## , , 2 comps
##
##      IRS.10Y
## 756  6.5839
## 757  6.5849
## 758  6.5904
## 759  6.5686
## 760  6.5741
## 761  6.5765
## 762  6.5792
## 763  6.5470
## 764  6.5136
## 765  6.5894
## 766  6.6219
## 767  6.6272
## 768  6.5868
## 769  6.5167
## 770  6.4916
## 771  6.4768
## 772  6.5121
## 773  6.5174
## 774  6.5029
## 775  6.4493
## 776  6.4802
## 777  6.4789
## 778  6.4753
## 779  6.4753
## 780  6.4753
## 781  6.5022
## 782  6.5408
## 783  6.5464
```

## Conclusiones:

- Es adecuado realizar un análisis de componentes sobre el dataset propuesto, dado el alto grado de correlación entre variables.

- Para explicar adecuadamente la estructura subyacente de los tipos de interés de interés es suficiente con dos componentes principales, que en total explican un 98.63 % de la varianza.
- Una rotación de variables subyacentes mediante varimax daría lugar a un modelo mejor.