

COLEGIO UNIVERSITARIO DE ESTUDIOS FINANCIEROS CUNEF

Máster en Data Science para Finanzas

Técnicas de Agrupación y de Reducción de la Dimensión

Nombre: Paúl Córdova

Informe Ejecutivo

El presente informe tiene la finalidad de analizar la estructura temporal de los tipos de interés y gestionar la aplicación de la técnica de Análisis de Componentes Principales (ACP) para agrupar los plazos de los bonos norteamericanos según sus características comunes. Se determinó una relación creciente entre el promedio del tipo de interés con el tiempo de maduración, es decir, a una mayor temporalidad se incrementa el rendimiento de los bonos. Gráficamente, se demuestra que la maduración de los bonos a mediano y largo plazo mantiene un rendimiento superior al 6% en contraste con el corto plazo; además, este hecho se confirma con la visualización de la curva de rendimiento de los bonos con carácter creciente. Por su parte, el ACP predice que el primer componente contribuye en explicar a la variable suplementaria (IRS.10Y) en un 79.71%, siendo validada la aplicación de esta técnica por múltiples pruebas estadísticas. En consecuencia, el ACP determina que en conjunto el primer y segundo componente explican un 98.3% de la varianza; por lo cual, se logra plasmar de mejor forma la estructura subyacente en los tipos de interés.

Informe Nº 2 – Estructura temporal de los tipos de interés

Como punto de partida para el estudio de los rendimientos de los tipos de interés, se trabajó con una base de datos de series temporales, la cual contiene información diaria de los tipos de interés de los bonos norteamericanos¹ a partir del 2 de enero de 1995 al 30 de septiembre de 1998.

Al realizar un análisis exploratorio de la base de datos, en promedio se destaca una relación directa entre la tasa de interés y el periodo de maduración de los bonos, es decir, a medida que avanza la temporalidad, los tipos de interés de los bonos aumentan (ver Anexo A). Cabe recalcar que en maduración a un mes se encontraron valores desconocidos; sin embargo, posteriormente se presenta una solución a dicho inconveniente.

A continuación, se determinó que en la maduración a un mes la tasa mínima fue de 5,31%; mientras que, en la maduración a diez años la tasa máxima fue de 8.34%. Esto nos indica que el rango de valores o tipo de interés de los bonos se encontró entre un 5% y 9% aproximadamente.

¹ La base de datos contiene como variables a los plazos de los bonos (Ej. 1 mes, 3, meses, 6 meses, etc.)

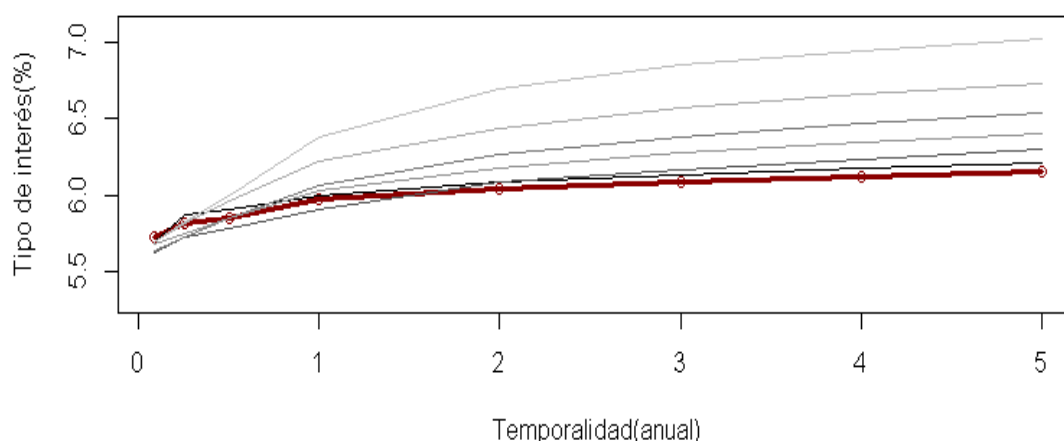
Con el objetivo de profundizar en la distribución de los tipos de interés según la maduración, se realizó un histograma de densidad. El Anexo B destaca que los bonos con plazos cortos, presentaron una distribución con asimetría positiva, lo que conllevó que la mayor parte de rendimientos mantuvieran tipos de interés bajos; simultáneamente, es evidente una alta concentración de observaciones en tipos de interés menores al 6% en el periodo de maduración de un mes y doce meses respectivamente.

Mientras, los plazos de maduración de cinco y diez años mantuvieron simetría en sus distribuciones, aspecto que al contrastarlo con el análisis previo, se determina que a medida transcurre la temporalidad se obtiene rendimientos mayores en los tipos de interés de los bonos. No obstante, la maduración en el mediano y largo plazo presentó una menor concentración en los datos, aspecto que conlleva mayor variabilidad en la información.

Estructura temporal de los tipos de interés

En el mismo sentido, se analizó la curva de rendimiento o también denominada estructura temporal de los tipos de interés. Esto con la finalidad de confirmar el análisis previo de la información y poder determinar gráficamente la evolución de la tasa de interés. Ante ello, el Gráfico N° 1 resalta la existencia de una curva creciente², lo cual significa que los tipos de interés tienden a ser bajos en el corto plazo con respecto a las tasas en el largo plazo; lo que sugiere una reducción económica para dicho periodo de tiempo. Por otra parte, es fundamental pormenorizar en el uso de esta herramienta gráfica para la aplicación de modelos con las curvas de rendimiento dinámicas en miras de mejorar el análisis económico (Redfern y McLean, 2014).

Gráfico N° 1 Curva de rendimiento de los bonos norteamericanos



Fuente: Bond Yield Rates – U.S. Government

Elaborado por: Paúl Córdova

² La línea roja indica la curva de rendimiento de los bonos al 04/12/1997; mientras que, cada una de las líneas plomas representan una sucesión de los meses anteriores a la fecha de corte (seis meses).

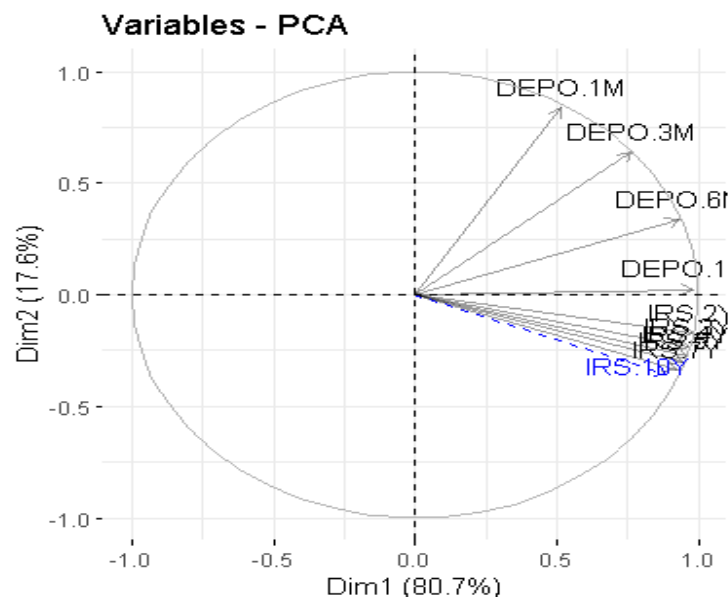
Adicionalmente, se observa que la estructura de tipos de interés es creciente para meses anteriores (líneas plomas) con respecto al punto de corte (línea roja); sin embargo, la curva dichos meses destaca un mayor concavidad; lo que indica la reducción del rendimiento esperado a medida transcurre el tiempo. Donde según BlackRock (2018) este aspecto puede llegar a ser perjudicial para las expectativas de inversión, dado el aplanamiento de la curva de rendimientos durante el segundo semestre de 1997.

Análisis de Componentes Principales (Predicción)

Antes de proceder con la técnica estadística, cabe aclarar que se imputaron los valores desconocidos (NA) de la variable de maduración correspondiente a un mes; este proceso se efectuó por medio del algoritmo iterativo del Análisis de Componentes Principales.

Después de haber realizado el paso previo, se procede a gestionar la aplicación de la técnica de Análisis de Componentes Principales (ACP) mediante la determinación de los elementos (observaciones-variables) activos y pasivos en la base de datos. Este proceso tiene el objetivo de agrupar los plazos de maduración de los bonos en base a características comunes y posteriormente llegar a predecir los elementos suplementarios a través la determinación de los componentes principales de los elementos activos.³

Gráfico N° 2 Gráfico de posiciones



Fuente: Bond Yield Rates – U.S. Government

Elaborado por: Paúl Córdova

³ Para la elección de los elementos activos en la base de datos se filtró a partir de la fila [1:949] y las nueve primeras variables correspondientes al tiempo de maduración de los bonos. Mientras que, las variables suplementarias comprenden las filas [950:978] y la última columna (el valor del bono a 10 años).

Después de la aplicación de la técnica de ACP en los elementos activos, se predice que la primera dimensión contribuye en explicar mayormente a la variable suplementaria (*IRS.10Y*) en un 79.71% y en la segunda dimensión en un 12.59%; sumando en conjunto entre ambos componentes un 92.3%. Además, es evidente una correlación positiva de los elementos suplementarios con la primera dimensión y negativamente con la segunda (ver Gráfico 2).

Pruebas Estadísticas

Para comprobar la necesidad de realizar el Análisis de Componentes Principales se ejecutaron varias pruebas a los elementos activos en la base de datos. En una primera instancia, se identificó que en la matriz de correlación existe una alta relación entre las variables cercanas en sus plazos de maduración, esto se debe a la relación temporal existente entre los tipos de interés. En este sentido, se identificó que el determinante en la matriz de correlación fue muy bajo; lo que indica altas intercorrelaciones entre las variables. Consecuentemente se aplicó la prueba de esfericidad de Bartlett la cual determinó el rechazo de la hipótesis nula ($p.value < 0.05$), en consecuencia se estableció la no independencia de las variables; por lo cual conviene aplicar esta técnica. Por último, se ejecutó la prueba de Kaiser-Meyer-Olkin dictaminando un valor de KMO de 0.87; lo que conlleva una fuerte asociatividad en el conjunto de variables y se recomienda la aplicación del ACP para esta base de datos (López y Fachelli, 2015).

Elección óptima de componentes principales y rotación Varimax

En el Anexo C, se encuentra la llamada gráfica de sedimentación en la cual se puede apreciar el porcentaje de varianza explicada por cada uno de los componentes previstos en base a los elementos activos de la base de datos (López, 2018). Es evidente los dos primeros componentes explican en conjunto un 98,3% de la varianza; por lo tanto se determina que ambos permiten explicar la estructura subyacente de los tipos de interés en los bonos norteamericanos.

El objetivo de la rotación Varimax consiste en minimizar la cantidad de variables que tienen un factor de saturación sobre otra variable, en búsqueda de ganar mayor interpretabilidad de los factores (López y Fachelli, 2015). En nuestro caso se analizó si es conveniente realizar la dicha rotación; sin embargo el p-value fue igual a cero; por lo cual, se descartó su aplicación para las variables subyacentes.

Conclusiones:

- Se evidencia en la base de datos una relación directa entre la tasa de interés y el periodo de maduración de los bonos; por lo cual, al transcurrir los plazos de los bonos sus tipos de interés se incrementan. Esta noción se confirma al analizar la estructura temporal de los tipos de interés que mantiene una forma creciente.

- Es evidente una alta concentración de observaciones en tipos de interés menores al 6% en el periodo de maduración de un mes y doce meses respectivamente; siendo mayor el rendimiento al mediano y largo plazo.
- La predicción del ACP destaca que la primera dimensión contribuye en explicar mayormente a la variable suplementaria (*IRS.10Y*) en un 79.71% y en la segunda dimensión en un 12.59%.
- El primer y segundo componente del ACP explican en conjunto un 98,3% de la varianza; siendo la aplicación de esta técnica probada mediante herramientas estadísticas, las cuales confirmaron la robustez en el uso del Análisis de Componentes Principales.

Referencias bibliográficas o digitales:

- López, Juan M. (11 de octubre de 2018). "El análisis de componentes principales". *Notas de clase*. Colegio Universitario de Estudios Financieros.
- López, Pedro y Fachelli, Sandra. (2015). *Metodología de la Investigación Social Educativa*. Barcelona: Universitat Autònoma de Barcelona.
- *Lo que la curva de rendimiento puede decir a los inversionistas*. (2018). BlackRock. Recuperado de <https://www.blackrock.com/cl/vision-de-mercado/what-the-yield-curve-can-tell-equity-investors>
- Redfern, David y McLean, Douglas. (2014). Principal Component Analysis for Yield Curve Modelling Reproduction of out-of-sample yield curves. MOODY's Analytics. Recuperado de <https://www.moodyanalytics.com/-/media/whitepaper/2014/2014-29-08-pca-for-yield-curve-modelling.pdf>

ANEXOS

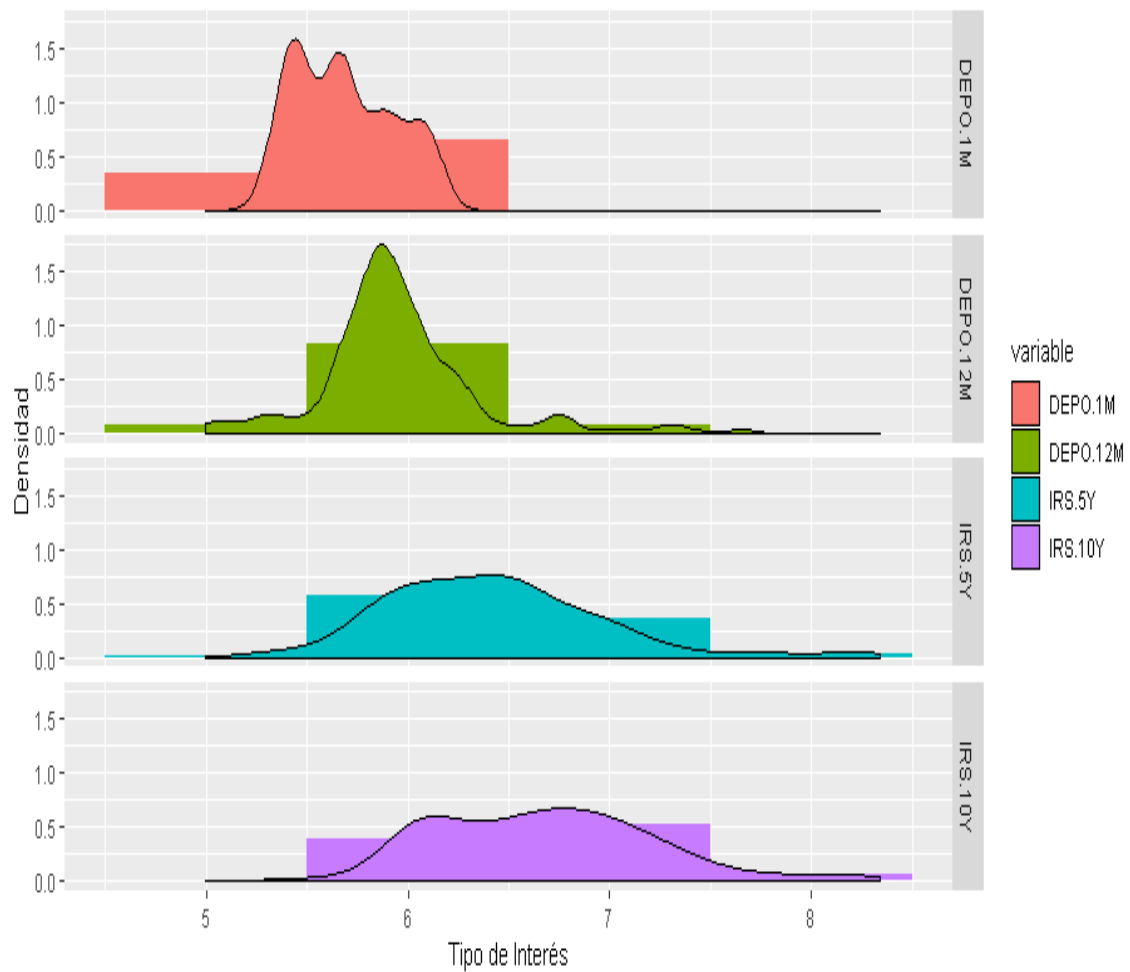
A. Tabla del promedio de los tipos de interés por variable (periodo de maduración)

	D.1M	D.3M	D.6M	D.12M	D.2Y	D.3Y	D.4Y	D.5Y	D.7Y	D.10Y
Promedio	5.69	5.74	5.81	5.97	6.12	6.25	6.35	6.43	6.55	6.69

Fuente: Bond Yield Rates – U.S. Government

Elaborado por: Paúl Córdova

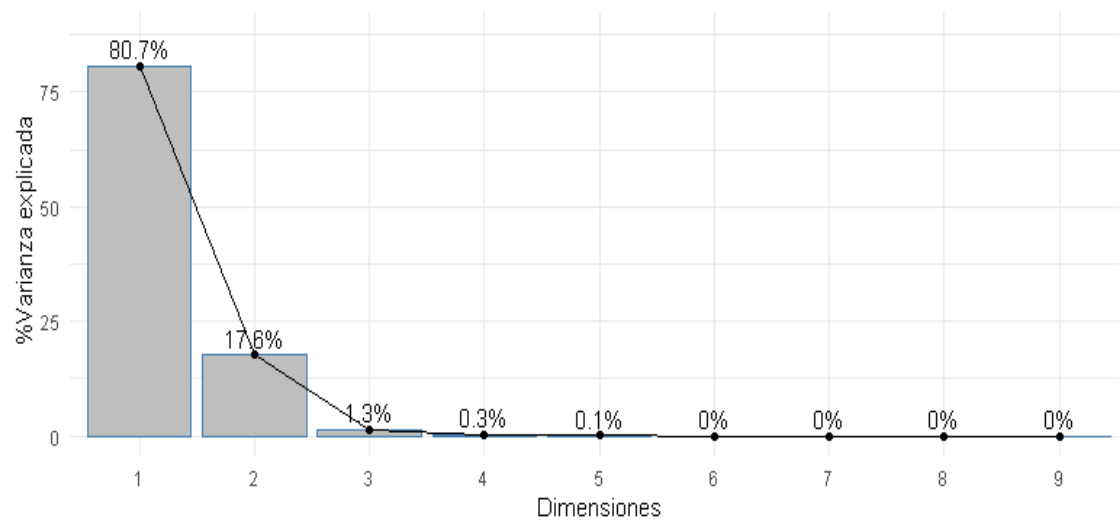
B. Histograma de densidad según variable



Fuente: Bond Yield Rates – U.S. Government

Elaborado por: Paúl Córdova

C. Gráfico de Sedimentación



Fuente: Bond Yield Rates – U.S. Government

Elaborado por: Paúl Córdova

D. Script

```
#### Análisis rendimiento de los bonos norteamericanos ####

# Delimitamos el WD
setwd("C:/Users/Edwin/Desktop/CUNEF/Master Data Science/Tecnicas de Agrupacion y reduccion de la
dimension/Datos")

# Cargamos base de datos
bd <- read.csv("ACPTIUSD.csv", header = T, sep = ";", dec = ".")

# Identificar la estructura y dimensión de los datos
str(bd)

# Visualizamos los primeros y últimos datos
head(bd, 10)
tail(bd, 10)

## ANÁLISIS EXPLORATORIO
library(tidyverse)

# Identificación de datos NA
sum(is.na.data.frame(bd))==T # 195 valores NA
sum(is.na(bd$DEPO.1M))==T # Todos los NA son de la 1era variable

# Análisis del promedio
bd %>%
  summarise(
    D.1M= round(mean(DEPO.1M, na.rm = T),2),
    D.3M= round(mean(DEPO.3M),2),
    D.6M= round(mean(DEPO.6M),2),
    D.12M= round(mean(DEPO.12M),2),
    D.2Y= round(mean(IRS.2Y),2),
    D.3Y= round(mean(IRS.3Y),2),
    D.4Y= round(mean(IRS.4Y),2),
    D.5Y= round(mean(IRS.5Y),2),
    D.7Y= round(mean(IRS.7Y),2),
    D.10Y= round(mean(IRS.10Y),2)
  )

# Rangos de la tasa de interés
min(bd$DEPO.1M, na.rm = T) # Mínimo a 1 mes
max(bd$IRS.10Y) # Máximo a 10 años

# Histograma de densidad según variables
library("reshape2")

nbd <- melt(bd)
nbd1 <- filter(nbd, variable==c('DEPO.1M','DEPO.12M','IRS.5Y','IRS.10Y'))
ggplot(nbd1, aes(x=value, fill=variable, xlab('a'))) +
  geom_histogram(aes(y = ..density..), na.rm = F, binwidth=1) +
  geom_density() +
  facet_grid(variable~.) +
  labs(x='Tipo de Interés', y='Densidad') +
  ggtitle('Histograma de densidad según maduración')

# Curva de redimiento de los bonos
temporalidad <- c(1/12,3/12,6/12,1,2,3,4,5)
g <- gray.colors(6,start=0.001,end=0.75,gamma=2.2)
plot(temporalidad,
      bd[783,2:9], #no se incluyen NA
      type="l",
      xlab="Temporalidad(anual)",
      ylab="Tipo de interés(%)",
      lwd=3.5,
      ylim=c(5.3,7.1),
      col="darkred")
```

```

points(temporalidad,
       as.numeric(bd[783,2:9]),
       col="darkred")
for ( i in 1:6 ) {
  lines(temporalidad,
       as.numeric(bd[783-i*30,2:9]),
       type="l",
       col=g[i])
}

## ANÁLISIS COMPONENTES PRINCIPALES

# Carga de las librerías
library(factoextra)
library(FactoMineR)

## Análisis de Componentes Principales (ACP)

# Imputamos los valores NA en la variable DEPO.1M
library("missMDA")
nb <- estim_ncpPCA(bd[, -1], ncp.max=10)
nb #Para el siguiente paso el número de componentes debe ser igual a 7
bdm <- as.data.frame(imputePCA(bd[, -1], ncp=7))
#Base de datos sin NA
bd1 <- bdm[, 1:10]
names(bd1) <- sub(c("completeObs."), "", names(bd1))

#ACP- Observaciones y variables activas
ACP <- PCA(bd1[1:949, 1:9], graph=T)
ACP

#ACP- Observaciones y variables suplementarias
ACPs <- PCA(bd1, ind.sup = 950:978,
           quanti.sup = 10, graph=FALSE)

# Predección para la variable cuantitativa IRS.10Y
ACPs$quanti.sup

# Gráfico de posición de los variable suplementaria
fviz_pca_var(ACPs, alpha.var = 0.4)

## Aplicación de pruebas - Justificación uso de ACP
mcorr <- cor(na.omit(bd[1:945, 2:10]))

#Matriz de correlación
mcorr
#Determinante de la matriz de correlación
det(mcorr)
#Test de esfericidad de Bartlett
library(psych)
cortest.bartlett(mcorr, n=nrow(bd))
#Test de Kaiser-Meyer-Olkin
KMO(mcorr)

# Gráfico de sedimentación
fviz_eig(ACP, addlabels=T, hjust=0.5, barfill = 'gray', main=NULL) +
  labs(x="Dimensiones", y="%Varianza explicada", title = "Gráfico Sedimentación") +
  ylim(0, 90)+
  theme_minimal()

# Rotación VARIMAX
fa.parallel(bd1[1:945, 1:9])
varimax1 <- factanal(bd1[1:945, 1:9], factors=2, rotation="varimax")
varimax1

```