

Técnicas de agrupación y reducción de la dimensión

# Estructura temporal de los tipos de interés

Hugo César Octavio del Sueldo

---



## Introducción

El objetivo que perseguimos en el presente trabajo es, simplemente, efectuar una comprobación empírica mediante la aplicación del ACP 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. No pretendemos nada más que verificar si, tal y como plantean los estudios teóricos, puede establecerse una estructura subyacente que sintetice y agrupe los distintos plazos en virtud de sus características comunes. Para ello, trabajaremos con el archivo `ACPTIUSD.csv`, del que utilizaremos las 949 primeras observaciones (denominadas observaciones activas) y las 9 primeras variables (las variables activas); uno de los objetivos será emplear las observaciones 950 a 978 (llamadas observaciones suplementarias) para predecir el valor del bono a 10 años (IRS.10Y,

variable suplementaria). Aparte de cubrir este objetivo, queremos asimismo tener respuesta a las siguientes preguntas:

- ¿Tiene sentido llevar a cabo, en este caso, un análisis de componentes principales? Para justificarlo analizaremos la matriz de correlaciones, el determinante de dicha matriz, la prueba de esfericidad de Bartlett y el KMO;
- ¿Cuántos componentes permitirían explicar, adecuadamente, la estructura subyacente de los tipos de interés aquí analizados? Para responder a esta pregunta, realizaremos el gráfico de sedimentación;
- Finalmente, ¿tiene sentido llevar a cabo una rotación de las variables subyacentes? Para responder, llevaremos a cabo una rotación Varimax, quartimax y Promax y las compararemos con los componentes principales sin rotación.
- Por último, elaboraremos las oportunas conclusiones al trabajo.

## Desarrollo del análisis

En este apartado se explicará cómo se ha desarrollado el análisis y se dará respuesta a los diferentes objetivos planteados

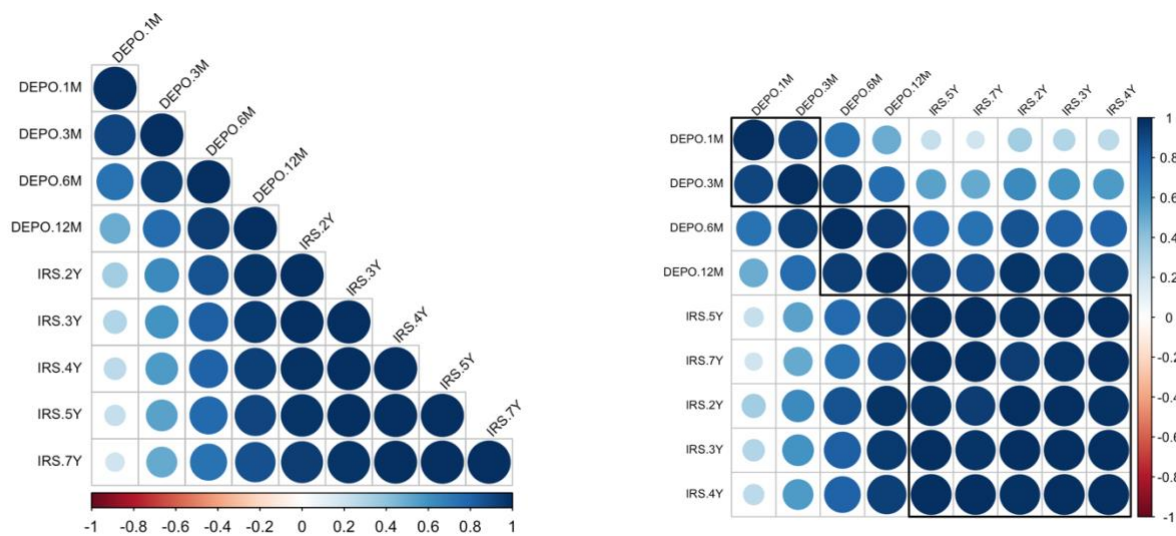
### *Análisis de la matriz de correlaciones*

#### **Matriz de correlación**

Para contestar al primer objetivo del trabajo debemos empezar con el análisis de la matriz de correlaciones de las observaciones activas (949 observaciones del fichero).

	DEPO.1M	DEPO.3M	DEPO.6M	DEPO.12M	IRS.2Y	IRS.3Y	IRS.4Y	IRS.5Y	IRS.7Y
DEPO.1M	1.00	0.91	0.73	0.49	0.34	0.29	0.26	0.23	0.19
DEPO.3M	0.91	1.00	0.93	0.76	0.63	0.59	0.56	0.53	0.50
DEPO.6M	0.73	0.93	1.00	0.94	0.86	0.82	0.80	0.77	0.73
DEPO.12M	0.49	0.76	0.94	1.00	0.97	0.95	0.93	0.91	0.87
IRS.2Y	0.34	0.63	0.86	0.97	1.00	0.99	0.98	0.97	0.94
IRS.3Y	0.29	0.59	0.82	0.95	0.99	1.00	1.00	0.99	0.97
IRS.4Y	0.26	0.56	0.80	0.93	0.98	1.00	1.00	1.00	0.99
IRS.5Y	0.23	0.53	0.77	0.91	0.97	0.99	1.00	1.00	1.00
IRS.7Y	0.19	0.50	0.73	0.87	0.94	0.97	0.99	1.00	1.00

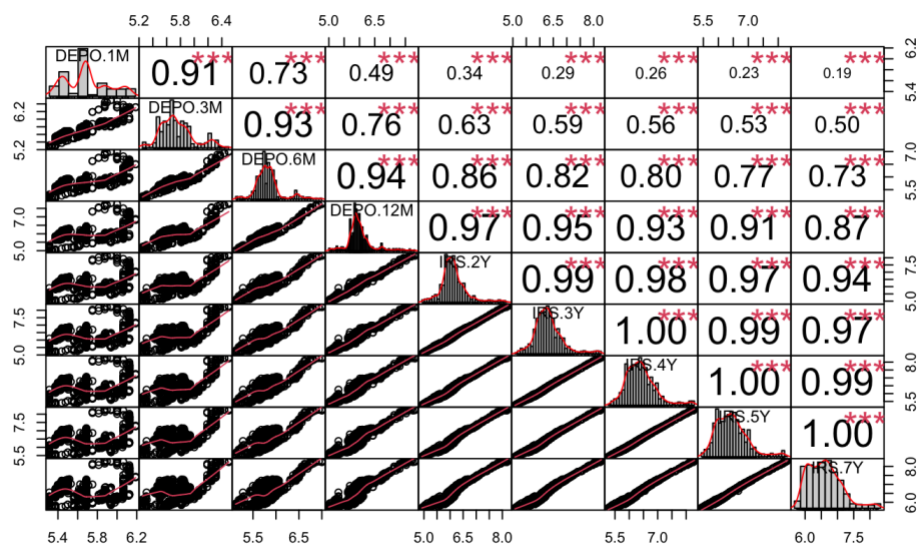
Se observa que a menor diferencia temporal entre los bonos, mayor es la correlación entre los mismos. Es decir, que el interés del bono a 10 años tendrá un comportamiento similar al bono a 7 años que el bono a un mes de plazo. Visualmente se pueden apreciar mejor estos comportamientos por lo que a continuación mostraremos la matriz de correlaciones en dos gráficos.



Realizando un análisis cluster podemos observar la correlacion existente en tres grupos

- los bonos de cortisimo plazo de 1 mes y 3 meses
- entre los bonos de 6 meses y 12 meses
- los bonos de mediano y largo plazo entre 2 años y 7 años

Otra forma de ver las correlaciones entre variables es la siguiente:



Por debajo de la diagonal, en los diagramas de dispersión por pares, podemos observar como a medida que el plazo es mayor la correlación aumenta. Esto mismo se confirma en el triángulo superior con el valor de la correlación y los niveles de significancia (con estrellas)

### **Análisis del determinante de la matriz**

Para conocer la presencia de multicolinealidad entre las variables objeto de estudio se debe analizar el determinante de la matriz. Cuanto más bajo sea, mayor asociación tendrán las variables entre sí, de forma que será adecuado llevar a cabo el ANFAC.

En esta oportunidad el resultado del determinante es  $3.47622 \times 10^{-13}$ , por lo tanto, al ser muy cercano a 0, lo que sugiere es un alto nivel de colinealidad en el conjunto de variables involucradas en la matriz.

### **Prueba de esfericidad de Bartlett**

Ahora aplicamos la prueba de Bartlett que se utiliza para probar la hipótesis nula que afirma que las variables no están correlacionadas en la población.

La prueba de esfericidad de Bartlett contrasta la hipótesis nula de que la matriz de correlaciones es una matriz identidad, en cuyo caso no existirían correlaciones significativas entre las variables y el modelo factorial no sería pertinente.

Bartlett test of homogeneity of variances

data: TIUSD.act.C

Bartlett's K-squared = 1654, df = 8, p-value <  $2.2 \times 10^{-16}$

El resultado del p valor nos permite rechazar la hipótesis nula. Por lo tanto, podemos continuar con nuestro ANFAC.

### **Índice KMO**

En este paso nos tenemos que preguntar si existe la suficiente correlación entre las variables para efectuar el análisis factorial. El KMO lo hace a partir de la matriz de correlaciones parciales.

El KMO es una medida de adecuación de la muestra; este índice permite comparar las magnitudes de los coeficientes de correlación observados con las magnitudes de los coeficientes de correlación parcial.

Valores bajos del índice KMO desaconsejan el empleo del ANFAC (esto ocurriría cuando la suma de todos los coeficientes de determinación parciales sea pequeña en relación a la suma de todos los coeficientes de determinación), dado que las correlaciones entre pares de variables no pueden explicarse por el resto de variables.

Kaiser-Meyer-Olkin factor adequacy

Call: KMO(r = TIUSD.act.C)

Overall MSA = 0.87

MSA for each item =

DEPO.1M	DEPO.3M	DEPO.6M	DEPO.12M	IRS.2Y	IRS.3Y	IRS.4Y	IRS.5Y	IRS.7Y
0.81	0.80	0.85	0.92	0.88	0.85	0.89	0.87	0.90

El resultado es 0.87 lo que nos dice es que podemos continuar con el análisis Factorial.

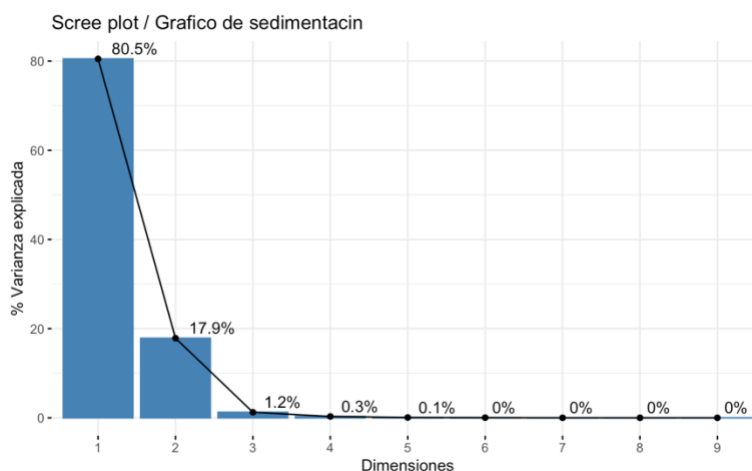
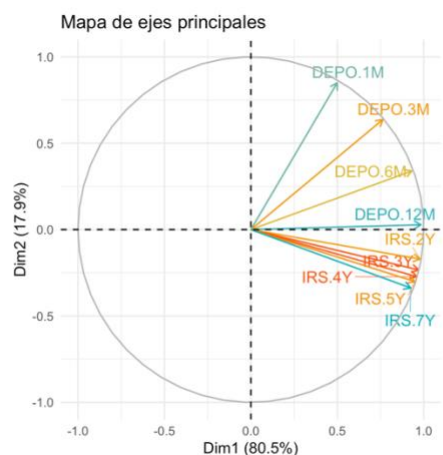
Por lo tanto, concluimos para nuestro primer objetivo que luego de realizar los cuatro análisis, podemos continuar con el análisis factorial.

Una vez contestado nuestro primer objetivo, vamos a responder al segundo objetivo planteado, el cual está ligado a la cantidad de componentes principales a extraer y al análisis de CP en sí mismo.

### *Análisis de componentes principales*

Para extraer los factores vamos a utilizar el método de componentes principales que se explica brevemente abajo.

**- Análisis de componentes principales.** Método para la extracción de factores utilizada para formar combinaciones lineales no correlacionadas de las variables observadas. El primer componente tiene la varianza máxima. Las componentes sucesivas explican progresivamente proporciones menores de la varianza y no están correlacionadas unas con otras. El análisis principal de las componentes se utiliza para obtener la solución factorial inicial. No se puede utilizar cuando una matriz de correlaciones es singular.



Tras realizarse dicho análisis, podemos observar que con nuestros primeros dos componentes principales podemos explicar el 98.4% de la variabilidad total de las variables. El 80.5 % es explicado por la dimensión 1 y el 17.9% por la dimensión 2.

Al mismo tiempo, podemos observar que la Dim.1 está fuertemente explicada por los bonos por encima de 6 meses, mientras que la Dim.2 está explicada por los bonos a 1 y 3 meses.

Ahora nos deberíamos preguntar ¿Cuántos componentes permitirían explicar, adecuadamente, la estructura subyacente de los tipos de interés aquí analizados?

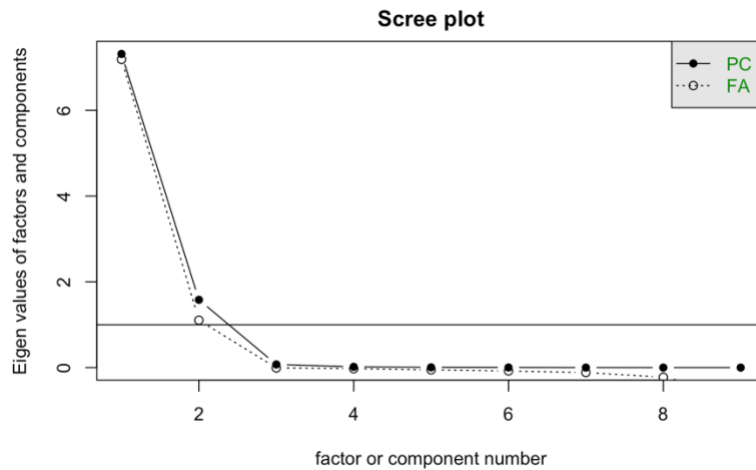
Respondemos a esta pregunta con las tres pruebas siguientes:

**- Kaiser Criterion (Guttman, 1954):** esta regla sugiere que se deben retener todos los factores que tengan un eigenvalue de 1.0 o mayor; con el razonamiento de que un factor no debe explicar menos que la varianza equivalente que hubiera explicado una sola de las variables incluidas en el análisis. La regla sin embargo no es estricta y debe analizarse en conjunto con otros criterios.

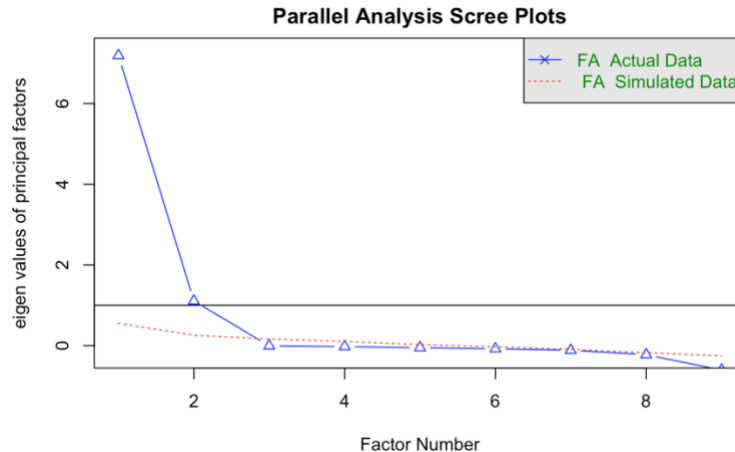
eigenvalue	percentage of variance	cumulative percentage of variance
comp 1 7.315910e+00	8.128789e+01	81.28789
comp 2 1.581944e+00	1.757715e+01	98.86504
comp 3 7.402414e-02	8.224904e-01	99.68753
comp 4 1.837135e-02	2.041262e-01	99.89166
comp 5 6.103129e-03	6.781254e-02	99.95947
comp 6 2.973791e-03	3.304212e-02	99.99251
comp 7 4.505266e-04	5.005851e-03	99.99752
comp 8 1.422854e-04	1.580949e-03	99.99910
comp 9 8.099756e-05	8.999729e-04	100.00000

Segun la regla del kaiser eligiríamos solo las dos primeras dimensiones al ser mayores a uno

**- Análisis del Scree Plot (Cattell, 1966):** este método complementa al anterior y se basa también el análisis de la magnitud de los eigenvalues pero a partir de la tendencia que se observa en el Scree Plot. Se procuran seleccionar un grupo reducido de factores que tengan eigenvalues significativamente superiores a los demás, para lo cual se identifica el punto de inflexión en la curva del scree plot (también referido como el codo por su semejanza con un brazo) a partir del cual la curva se transforma a una línea “plana” o relativamente recta. En el ejemplo que se presenta hay un claro punto de inflexión después de dos factores.



- **Análisis paralelo (Horn, 1965):** Esta regla suele complementar las anteriores cuando el numero de variables iniciales y factores resultantes es elevado. El procedimiento es basado en el principio de que los factores a extraer deben dar cuenta de mas varianza que la que es esperada de manera aleatoria. El procedimiento reordena las observaciones de manera aleatoria entre cada variable y los eigenvalues son recalculados a partir de esta nueva base de datos aleatoriamente ordenada. Los factores con eigenvalues mayores a los valores aleatorios son retenidos para interpretación.



Los graficos de sedimentacion nos confirman que el numero de factores debe ser 2

Finalmente, para responder al tercer objetivo debemos responder a la siguiente pregunta ¿tiene sentido llevar a cabo una rotación de las variables subyacentes? Para responder, llevaremos a cabo una rotación Varimax y la compararemos con los componentes principales sin rotación.

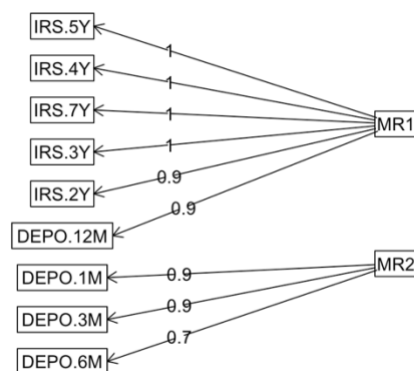
*Rotación factorial*

La matriz de saturaciones factoriales, o matriz factorial, indica la relacion entre los factores y las variables. Sin embargo, del resultado que finalmente obtenemos puede ser dificil extraer una interpretaci3n sencilla de los factores.

La rotacion factorial pretende seleccionar la soluci3n mas sencilla e interpretable, siempre siguiendo el criterio de parsimonia. En s3ntesis, consiste en hacer girar los ejes de coordenadas, que representan a los factores, hasta conseguir que se aproximen al maximo a las variables en que estan saturados.

Loadings:  
MR1 MR2  
DEPO.1M 0.067 0.946  
DEPO.3M 0.351 0.939  
DEPO.6M 0.661 0.741  
DEPO.12M 0.862 0.486  
IRS.2Y 0.948 0.312  
IRS.3Y 0.965 0.259  
IRS.4Y 0.975 0.224  
IRS.5Y 0.976 0.199  
IRS.7Y 0.974 0.166  
  
MR1 MR2  
SS loadings 5.990 2.843  
Proportion Var 0.666 0.316  
Cumulative Var 0.666 0.981

#### Factor Analysis



Con la rotacion Varimax los dos factores estan explicando el 98.1% de la variabilidad total. Por lo tanto, no conviene la rotacion ya que sin ella estabamos explicando el 98.4% de la variabilidad total.

El primer factor representa una proporcion mayor de la variabilidad total con el 66% y el segundo factor el 31%, sin embargo, como ya hemos comentado, no merece la pena realizar la rotacion ortogonal.



Una vez hecho esto, vamos a predecir el bono a 10 años para terminar con nuestro cuarto y último objetivo.

### *Predicción*

Para realizar la predicción, separamos una muestra que será utilizada de forma exclusiva para el test. Al mismo tiempo utilizaremos la función pcr y el número de componentes principales que hemos identificado como necesarios. Abajo los resultados del promedio de error respecto a los datos originales.

```
``{r}
#Predecimos sobre la muestra de test

modelo_test_pcr <- pcr(formula = IRS.10Y ~ ., data = TIUSD.sup.test, scale. = TRUE, ncomp = 2)

# Test-MSE
predicciones_test <- predict(modelo_test_pcr, newdata = TIUSD.sup.test, ncomp = 2)
test_mse <- mean((predicciones_test - TIUSD.sup.test$IRS.10Y)^2)
test_mse
```



```
[1] 0.0003958593
```


```

Podemos observar que las predicciones sobre la muestra de test son muy buenas ya que el promedio de error es solo de 0.0003 incluso mejor que los resultados expresados en la muestra train.

### *Conclusiones finales*

Luego de los análisis realizados podemos concluir que los análisis factoriales y de componentes principales pueden ser de gran utilidad para la predicción de los tipos de interés a 10 años u otros sucesos similares. Al mismo tiempo, compensa realizar un ajuste de la dimensión de los datos, quedandonos solo con aquellos que nos aporta valor, haciendonos más fácil el análisis.