
ESTRUCTURA TEMPORAL (SUBYACENTE) DE LOS TIPOS DE INTERÉS

Jose López Galdón

2020-11-08

ABSTRACT

En el presente trabajo hemos realizado una predicción del valor del bono estadounidense a 10 años, utilizando datos de los distintos bonos americanos entre los periodos del 2 de enero de 1995 y el 30 de septiembre de 1998. Para ello utilizaremos ACP (Análisis de Componentes Principales) de manera que podamos corroborar la siguiente cuestión: ¿puede establecerse una estructura subyacente que sintetice y agrupe los distintos plazos en virtud de sus características comunes?

PLANTEAMIENTO

El estudio de la estructura subyacente de los tipos de interés es reiterativa en la literatura financiera. Cabe citar, entre otros, el clásico artículo de Fama y Bliss de 1987 (The Information in Long-Maturity Forward Rates, The American Economic Review, Vol. 77, No. 4), el de Litterman y Scheinkman (Common factors affecting bond returns, Journal of fixed income, 1991) o, ya más recientes, el de Diebold, Piazzesi y Rudebusch (Modeling Bond Yields in Finance and Macroeconomics, American Economic Review 95.2, 2005), o el de Diebold y Li (Forecasting the term structure of government bond yields, Journal of Econometrics, 130.2, 2006). Puede consultarse también el paper de 2014 de Moody's Analytics Principal Component Analysis for Yield Curve Modelling (Enlaces a un sitio externo.) o el white paper de 2009 de Novosyolov y Satchkov, Global Term Structure Modeling Using Principal Component Analysis (Enlaces a un sitio externo.).

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, deberá trabajar con el archivo ACPTIUSD.csv. Vista previa del documento, disponible en la plataforma, del que deberá utilizar 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:

1. ¿Tiene sentido llevar a cabo, en este caso, un análisis de componentes principales? Para justificarlo, deberá llevar a cabo las pruebas que estime oportunas, como, por ejemplo el análisis de la matriz de correlaciones, el del determinante de dicha matriz, la prueba de esfericidad de Bartlett, el KMO o el MSA;
2. ¿Cuántos componentes permitirían explicar, adecuadamente, la estructura subyacente de los tipos de interés aquí analizados? Justifique su respuesta empleando, por ejemplo, las pruebas de la varianza explicada o del gráfico de sedimentación;
3. Finalmente, ¿tiene sentido llevar a cabo una rotación de las variables subyacentes? Para responder, lleva a cabo una rotación Varimax, por ejemplo.

Por último, deberá elaborar las oportunas conclusiones.

CÓDIGO UTILIZADO

Adjuntaremos un enlace a github con el código utilizado:

[HTTPS://GITHUB.COM/JLOPEZGALDON/TARD/BLOB/MAIN/01_ESTRUCTURA_TEMPORAL_TI/RMD/01_ESTRUCTURA_TEMPORAL_TI_V1.RMD](https://github.com/JLOPEZGALDON/TARD/blob/main/01_ESTRUCTURA_TEMPORAL_TI/RMD/01_ESTRUCTURA_TEMPORAL_TI_V1.RMD)

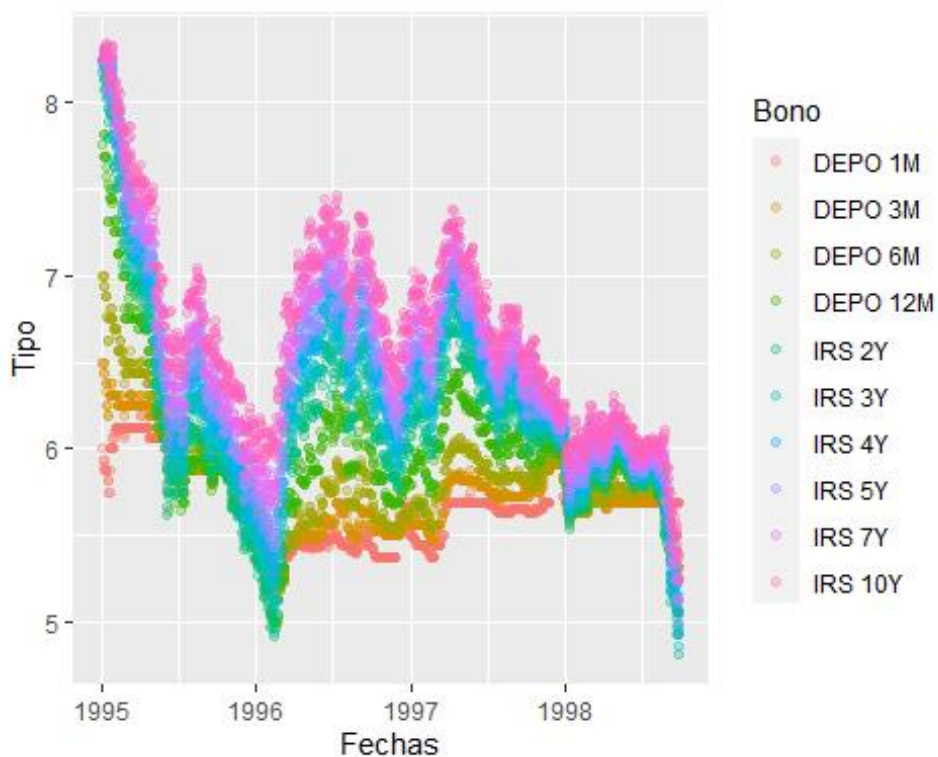
BASE DE DATOS

Tenemos un dataset con 978 filas y 11 variables, estas son:

- **X1**: Plazos entre el 2 de enero de 1995 y el 30 de septiembre de 1998
 - **DEPO_1M**: Bono norteamericano a 1 mes
 - **DEPO_3M**: Bono norteamericano a 3 meses
 - **DEPO_6M**: Bono norteamericano a 6 meses
 - **DEPO_12M**: Bono norteamericano a 12 meses
 - **IRS_2Y**: Bono norteamericano a 2 años
 - **IRS_3Y**: Bono norteamericano a 3 años
 - **IRS_4Y**: Bono norteamericano a 4 años
 - **IRS_5Y**: Bono norteamericano a 5 años
 - **IRS_7Y**: Bono norteamericano a 7 años
 - **IRS_10Y**: Bono norteamericano a 10 años
-
-

ANÁLISIS EXPLORATORIO

A continuación realizaremos un análisis exploratorio de nuestra base de datos, en el que analizaremos los estadísticos principales, correlaciones...



Como podemos observar los tipos de los bonos fluctuando a lo largo del tiempo debido a las distintas fases de los ciclos económicos, sin embargo, podemos ver cómo a partir de los bonos a 12 meses siguen una fluctuación distinta, lo que nos puede hacer pensar (a priori) que tenemos dos tipos de bonos, unos a corto plazo (por debajo de 12 meses) y a largo plazo (por encima de la anualidad).

RESUMEN DE LOS ESTADÍSTICOS

En este apartado incluiremos un resumen con los estadísticos principales de cada variable:

DATA SUMMARY

Name	TIUSD_activas
Number of rows	949
Number of columns	9



Column type frequency:

numeric	9
---------	---

Group variables	None
-----------------	------

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
DEPO 1M	0	1	5.69	0.22	5.31	5.48	5.69	5.87	6.19	
DEPO 3M	0	1	5.75	0.24	5.25	5.57	5.70	5.88	6.50	
DEPO 6M	0	1	5.82	0.30	5.12	5.66	5.78	5.91	7.00	
DEPO 12M	0	1	5.99	0.42	5.00	5.79	5.91	6.09	7.81	
IRS 2Y	0	1	6.15	0.51	4.92	5.86	6.08	6.32	8.22	
IRS 3Y	0	1	6.28	0.53	5.08	5.96	6.21	6.49	8.28	
IRS 4Y	0	1	6.37	0.53	5.26	6.02	6.32	6.62	8.29	

IRS 5Y	0	1	6.45	0.53	5.43	6.06	6.40	6.72	8.31	
IRS 7Y	0	1	6.58	0.53	5.67	6.14	6.54	6.88	8.33	

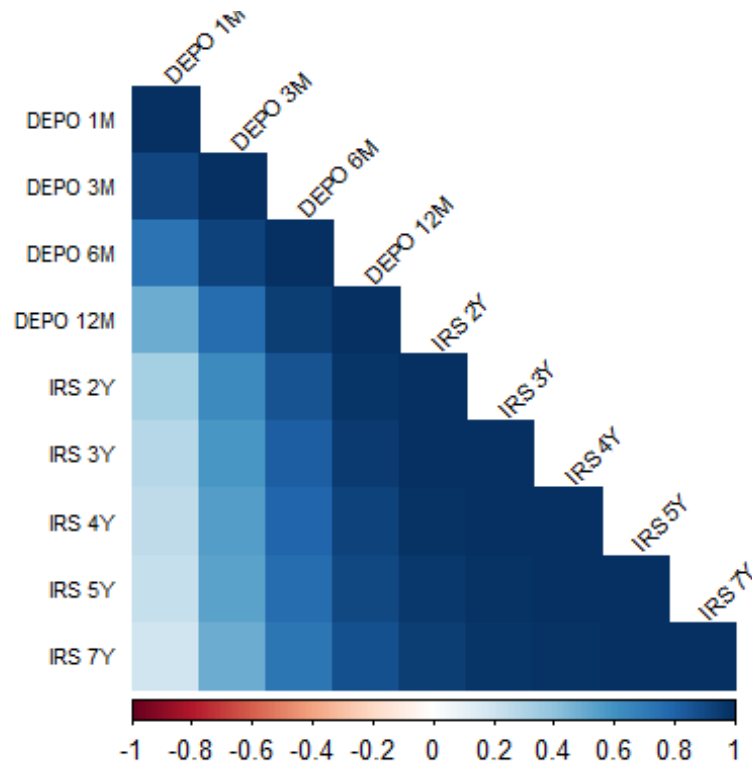
Tenemos 949 observaciones y 9 variables, de los depósitos a 1 mes tenemos 166 valores nulos. La media de los depósitos se sitúa entre el 5.6% y el 6.4%, conforme aumenta el plazo la desviación es mayor, esto tiene sentido porque aumenta el riesgo.

CORRELACIONES

##	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

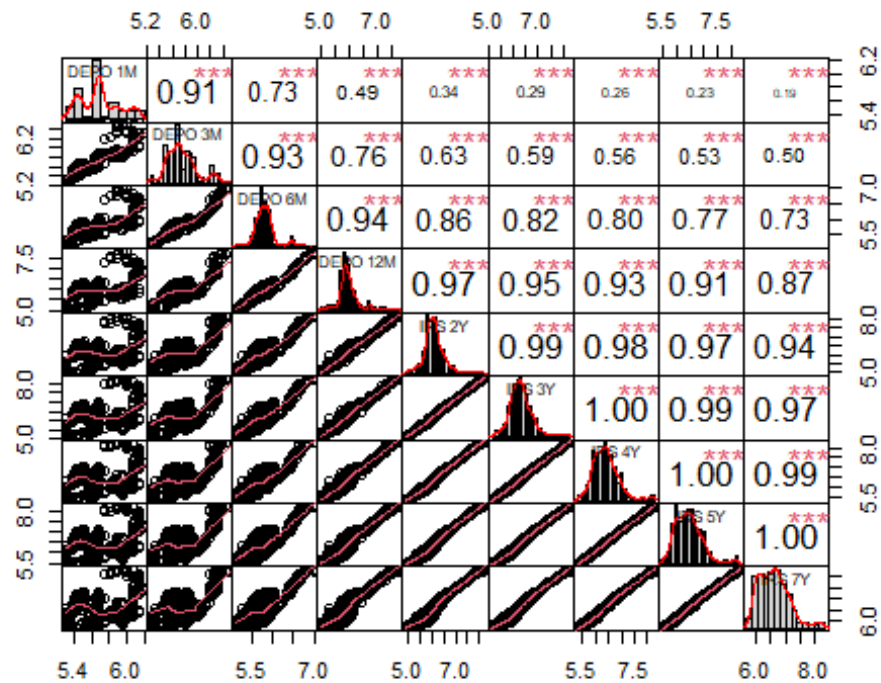
Como podemos observar en la primera salida, vemos como existe una alta correlación entre los bonos a 1 mes y los de 3 y conforme subimos el plazo, esta correlación disminuye... Sin embargo, en los bonos a más de 1 año está correlación es mucho más fuerte, por lo tanto, podemos seguir pensando en los dos grupos de bonos, los de corto plazo y los de largo.

Para lograr un mejor entendimiento plotearemos la matriz de correlaciones:



En el gráfico superior podemos ver las correlaciones en forma de *heatmap*, si es azul la correlación será positiva, blanco sin correlación y rojo correlación negativa. Así, de un simple vistazo podemos ver cómo existe una alta correlación entre los bonos a más de 6 meses, lo cual nos confirma lo que hemos deducido anteriormente: Existen dos grupos diferenciados en los bonos, los de largo plazo(más de 1 año) y los de corto plazo(menos de 1 año).

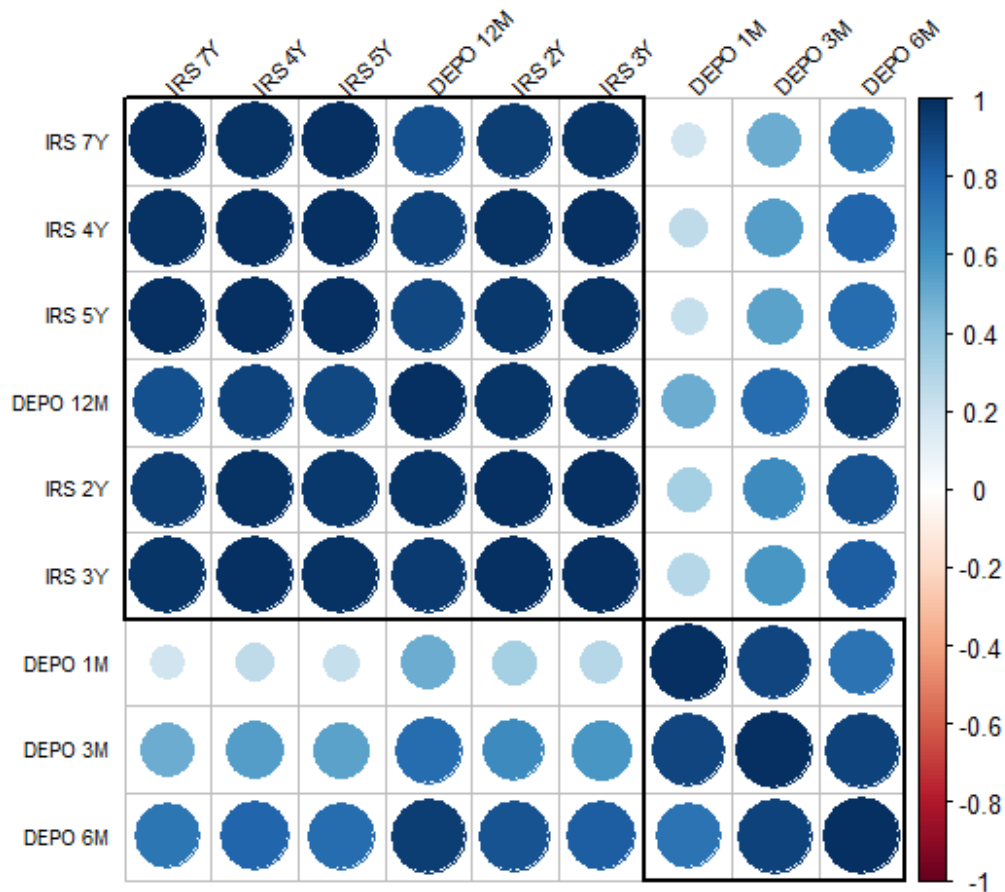
También podemos representar las correlaciones de la siguiente manera:



Está es otra buena manera de observar las correlaciones, y de ver cómo en los depósitos a 1 mes empeora la correlación conforme avanza el plazo, además aquí se ve de manera muy clara como es la correlación perfecta de los bonos a largo plazo, ya que se trata de una línea perfecta, parece ser que el corte entre los grupos podrá estar entre el bono a 12 meses o a 6.

CLUSTERS

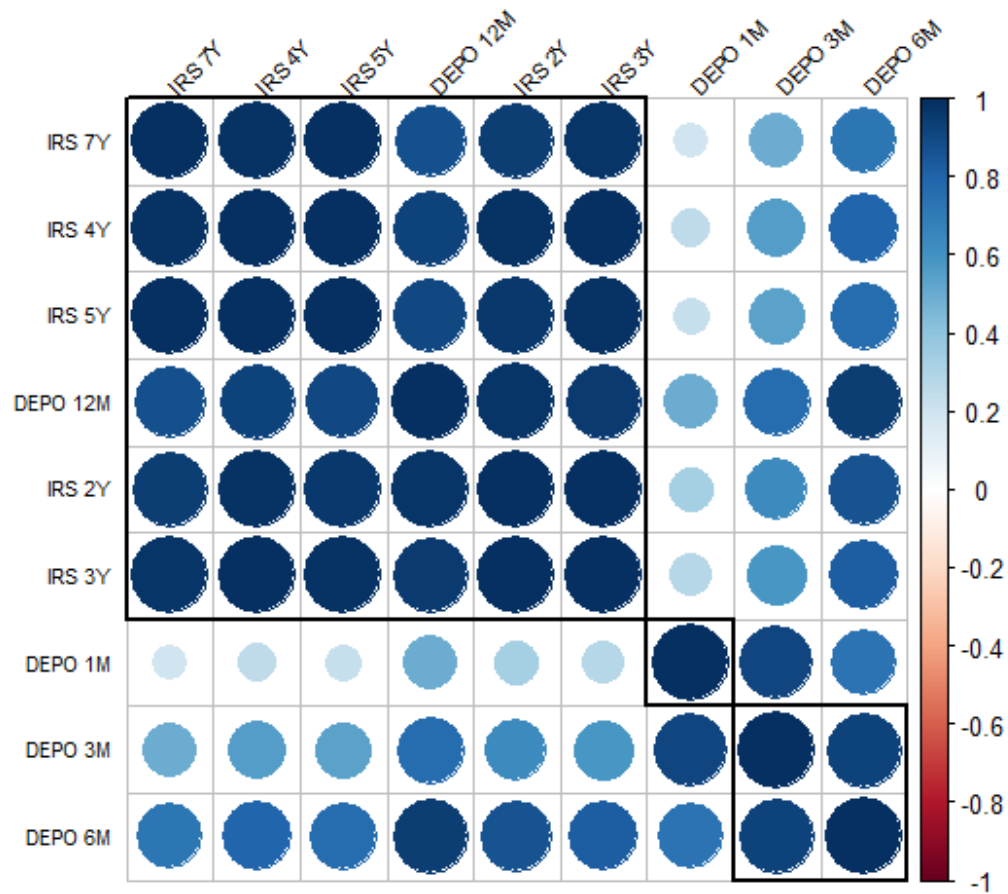
Lo último que haremos en este apartado es realizar un mapa de calor para definir los posibles grupos o clusters de manera visual:



Si establecemos dos clusters, podemos observar que existen dos tipos de plazos:

- A corto plazo: Bonos con una duración de 1-6 meses.
- A largo plazo: Bonos con una duración por encima de 12 meses.

Aún así, parece que encontramos una división en los bonos a 3 meses, ya que su correlación respecto al bono a 1 mes es menor. Vamos a probar con 3 clases:



Ahora podemos observar como tenemos 3 clases, con una mejor diferenciación:

- A muy corto plazo: Bonos con una duración de 1 mes.
- A corto-medio plazo: Bonos con una duración de 3-6 meses.
- A largo plazo: Bonos con una duración mayor a 1 año.

ANÁLISIS DE COMPONENTES PRINCIPALES

Una vez hemos comprendido mejor el dataset y lo tenemos listo para trabajar con el análisis de componentes principales.

Comenzaremos el ACP intentando resolver a la siguiente cuestión:

¿ TIENE SENTIDO LLEVAR A CABO, EN ESTE CASO, UN ANÁLISIS DE COMPONENTES PRINCIPALES?

Para resolver a esta cuestión, tendremos que analizar las correlaciones existentes entre las distintas varibales, analizar el *determinante de la matriz de correlaciones*, el *test de esfericidad de Barlett* y el *KMO*.

1. DETERMINANTE DE LA MATRIZ DE CORRELACIONES

Cuanto más bajo sea, mayor asociación tendrán las variables entre sí, de forma que será adecuado llevar a cabo el ANFAC.

El problema que tiene es que si existe una alta correlación entre las variables puede fallar, por lo que tendremos que realizar otros test.

Tras el cálculo del determinante de la matriz de correlaciones obtenemos un resultado de 2.64891e-18, lo que nos indicaría una alta colinealidad, aún así realizaremos el test de esfericidad de Barlett y el KMO para comprobar.

2. TEST DE ESFERICIDAD DE BARLETT

Se emplea para contratar la hipótesis de que la matriz de correlaciones es una matriz identidad, I . Se efectúa una prueba χ^2 a partir de una transformación del determinante de la matriz de correlaciones, $|R|$. Considerando que $|R| = 1$ si, y sólo si, $R = I$, podemos establecer la hipótesis nula indistintamente así:

$$H_0: R = I$$

Si se acepta la hipótesis, las variables no estarán correlacionadas, por lo que la nube de puntos en el espacio adoptaría la forma de una esfera.

Tenemos que recordar que el test de esfericidad no es adecuado para más de 100 observaciones.

Tras el cálculo del test obtenemos un p-value de 2.2e-16, por lo tanto rechazamos la hipótesis nula lo que indicará presencia de asociación entre las variables, estando en consecuencia plenamente justificado el empleo del ANFAC.

3. ÍNDICE DE KMO DE KAISER-MEYER-OLKIN.

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 (por debajo de 0.7) desaconsejan el empleo del ANFAC (esto ocurrirá 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.

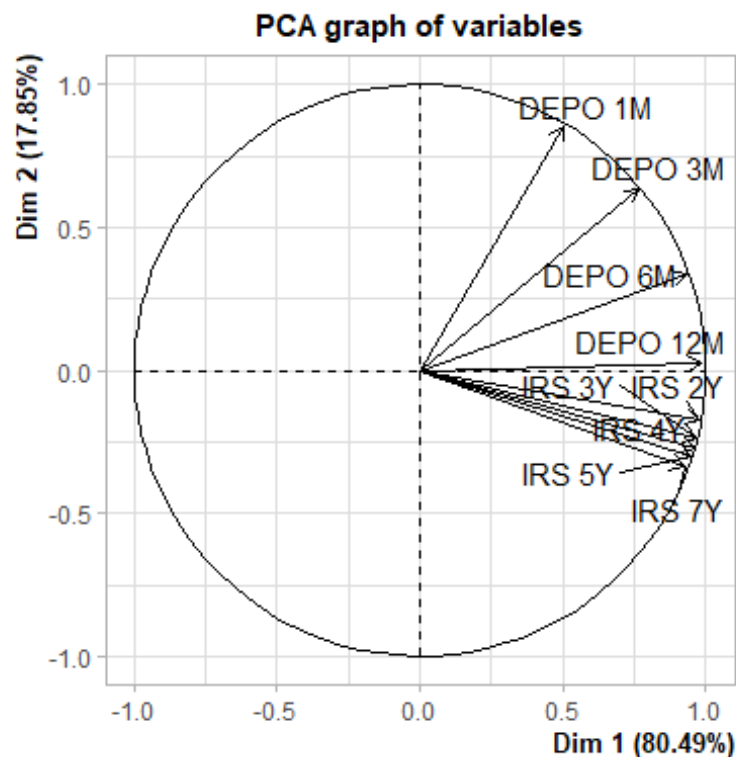
Tras realizar el cálculo del índice de KMO obtenemos una puntuación general de 0.87, por lo que se aconseja el uso del ACP.

CONCLUSIÓN:

Tras realizar el cálculo del determinante de la matriz de correlaciones, el test de esfericidad y el índice de KMO determinamos que es adecuado realizar el Análisis de Componentes Principales.

¿CUÁNTOS COMPONENTES PERMITIRÍAN EXPLICAR, ADECUADAMENTE, LA ESTRUCTURA SUBYacente DE LOS TIPOS DE INTERÉS AQUÍ ANALIZADOS?

Para saber cuántos componentes permiten explicar la estructura subyacente de los tipos de interés analizados, recordemos que son desde los bonos a 1 mes hasta bonos a 7 años.



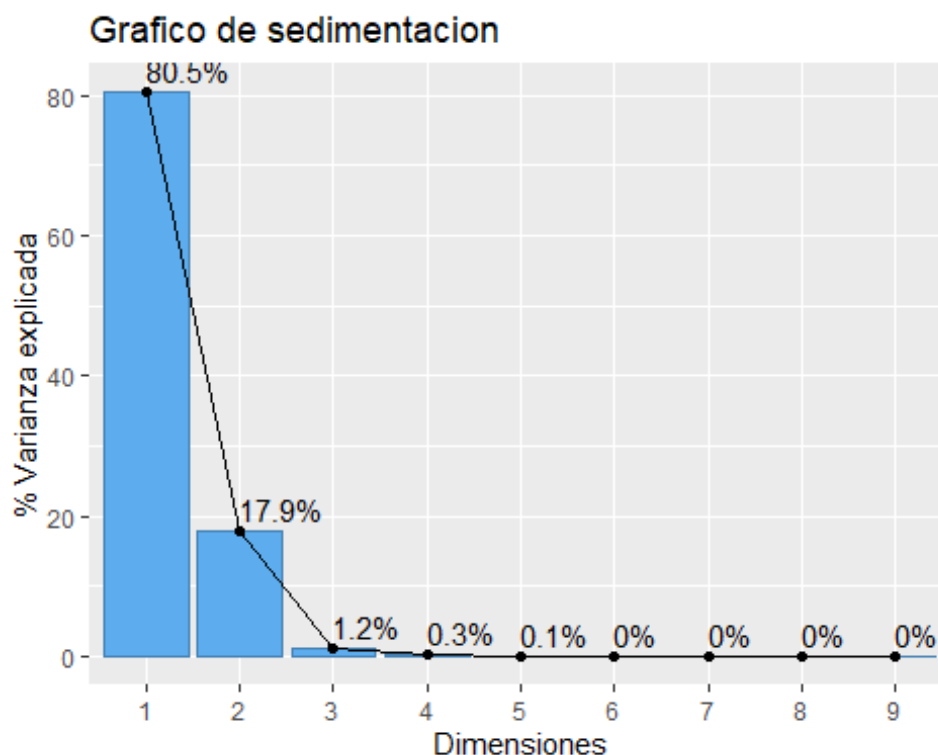
Podemos observar cómo a partir de dos dimensiones nuevas podemos explicar una gran parte de los datos, en concreto la Dim 1 explica el 80.49% de la varianza mientras que la Dim 2 explica el 17.85%, por lo que con dos dimensiones somos capaces de explicar el 98.34%.

Podemos observar cómo los depósitos a corto plazo son más explicados por la segunda dimensión, sobre todo el bono a 1 mes, mientras que los que son más largos tienen más explicación de la Dim 1.

##		EIGENVALUE	PERCENTAGE OF VARIANCE	CUMULATIVE PERCENTAGE OF VARIANCE
##	COMP 1	7.244544E+00	8.049493E+01	80.49493
##	COMP 2	1.606518E+00	1.785020E+01	98.34513
##	COMP 3	1.119973E-01	1.244415E+00	99.58954
##	COMP 4	2.712761E-02	3.014179E-01	99.89096
##	COMP 5	6.081229E-03	6.756922E-02	99.95853
##	COMP 6	3.133400E-03	3.481555E-02	99.99335
##	COMP 7	3.926834E-04	4.363149E-03	99.99771
##	COMP 8	1.279276E-04	1.421418E-03	99.99913
##	COMP 9	7.825952E-05	8.695502E-04	100.00000

Si obtenemos los autovalores de ACP podemos observar lo que comentábamos anterioremente, como con 2 componentes (Dim 1 y Dim 2) somos capaces de explicar el 98.34513% de la varianza acumulada, y cómo añadiendo componente a componente somos capaces de explicar el 100%. Aún así, como esos saltos no son muy grandes, no interesará añadir estos componentes, ya que compliran más el modelo sin proporcionarnos una gran explicación.

Esto lo podemos representar gráficamente:

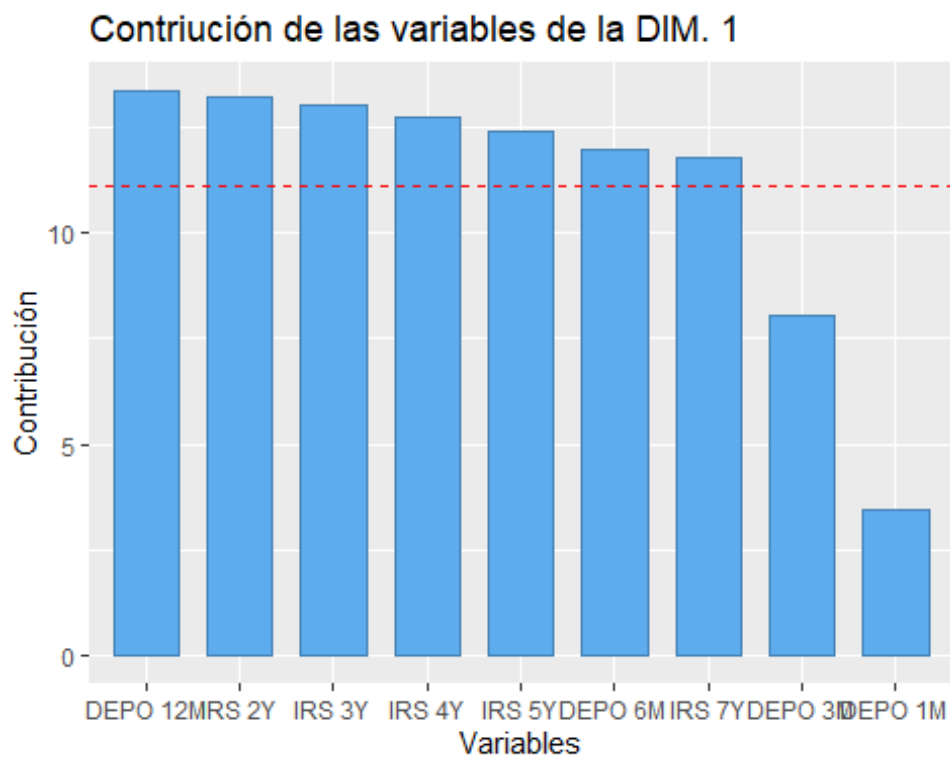


De esta manera, podemos visualizar mejor lo que comentábamos anteriormente. Así, el componente 1 es el más importante, seguido del 2 y los demás no aportan casi información. Existen dos reglas para la selección de componentes: la primera se basa en seleccionar aquellos componentes que superen $1/n$, es decir, tenemos 9 variables, por lo que cualquier dimensión que explique más de $1/9$ de la varianza la rentendremos, en este caso, el umbral si sitúa en 11,11%, por lo tanto, seleccionaríamos las dos primeras dimensiones, pero no la tercera. La otra es la regla del codo, que establece la selección dónde se produzca un mayor cambio de ángulo, en este caso podríamos estaría entre las dimensiones 2-3 o 3-4.

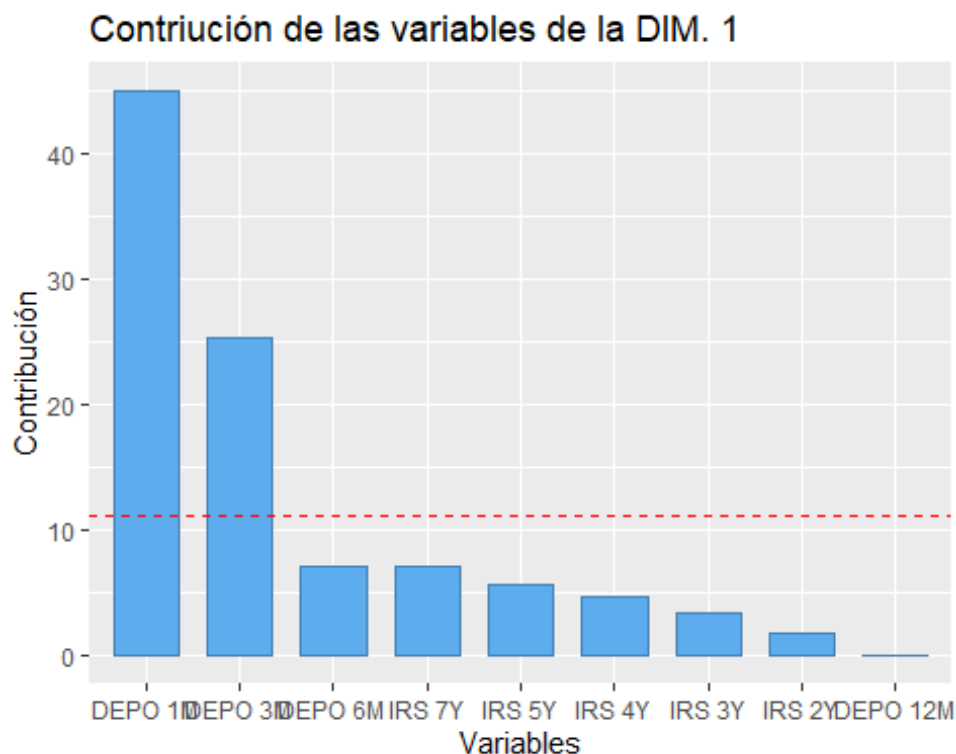
También podemos anaizar lo que aporta cada bono a las dimensiones:

##			DIM.1	DIM.2	DIM.3	DIM.4	DIM.5
##	DEPO	1M	3.439381	44.93511385	18.04746168	31.7745696	1.800575
##	DEPO	3M	8.042391	25.26417324	0.08898925	35.8725087	23.679973
##	DEPO	6M	11.961328	7.16490654	12.59499609	6.7102070	16.838025
##	DEPO	12M	13.363208	0.04886567	25.26420280	1.8090391	15.913029
##	IRS	2Y	13.231030	1.81111739	6.90389278	11.8004777	20.496582
##	IRS	3Y	13.019990	3.40552172	0.21917322	4.3371674	8.641580
##	IRS	4Y	12.738287	4.63991124	2.06794001	0.3178405	1.173193
##	IRS	5Y	12.411082	5.62809691	9.08658277	0.4737811	0.329674
##	IRS	7Y	11.793304	7.10229344	25.72676140	6.9044090	11.127370

Cómo podemos obserar, 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. Esto confirma, la hipótesis inicial de los dos grupos o clases de bonos. Para un mejor entendimiento lo graficaremos:



Como hemos comentado anteriormente, la Dim 1 está fuertemente explicada por los bonos por encima de 6 meses, mientras que los bonos a corto plazo (1-3 meses) no influyen en esta dimensión, sin embargo en la dimensión 2, ocurre lo contrario, está explicada por los bonos a corto plazo.



En ambos casos la línea roja representa la explicación media, por lo tanto, aquellas barras que la superen explicarán más que la media.

CONCLUSIÓN:

Podemos concluir que tras el cálculo de Análisis de Componentes principales podemos afirmar que con dos componentes somos capaces de explicar el 98,34% de la varianza, lo que significa, explicar de manera adecuada la estructura de los tipos de interés analizados. Además, hemos comprobado cómo la duración de estos bonos afecta a la contribución de dichas compentes.

¿TIENE SENTIDO LLEVAR A CABO UNA ROTACIÓN DE LAS VARIABLES SUBYACENTES?

La rotación factorial pretende seleccionar la solución más sencilla e interpretable, siempre siguiendo el criterio de parsimonia. En síntesis, consiste en hacer girar los ejes de coordenadas, que representan a los factores, hasta conseguir que se aproximen al máximo a las variables en que están saturados.

Método Varimax: Introducido por Kaiser en 1958, se basa en la determinación de la *simplicidad* de un factor, medida por la varianza de los cuadrados de sus saturaciones en las variables observables.

Podemos observar que si realizamos una rotación de las variables, las varianzas explicadas modifican, la dimensión 1 pasaría a explicar el 65.4% por lo que reduciría 15 puntos básicos, mientras que la Dim 2 aumentaría hasta explicar el 31.8%. En este caso no nos interesaría realizar la rotación ya que explicaríamos menos varianza, recordemos que sin la rotación eramos capaces de explicar el 98.34513% con dos dimensiones, mientras que ahora explicamos el 97.2% de la varianza con las mismas dimensiones.

CONCLUSIÓN:

Como no experimentamos ningún crecimiento de la varianza explicada, en este caso, no nos interesa realizar la rotación de las variables subyacentes.

PREDICCIÓN DEL BONO A 10 AÑOS

El objetivo es predecir el bono a 10 años estadounidense, para ello elaboraremos un modelo de regresión de componentes principales, en nuestro caso, con 2 componentes, tal y como hemos comprobado anteriormente.

[1,] 6.139403 6.065	[11,] 5.860989 5.755	[21,] 5.610234 5.565
[2,] 6.144255 6.085	[12,] 5.827750 5.745	[22,] 5.531313 5.485
[3,] 6.161167 6.105	[13,] 5.816541 5.745	[23,] 5.574945 5.525
[4,] 6.115523 6.055	[14,] 5.746359 5.705	[24,] 5.568450 5.555
[5,] 5.961408 5.935	[15,] 5.611643 5.575	[25,] 5.515425 5.495
[6,] 5.979951 5.915	[16,] 5.593727 5.565	[26,] 5.432076 5.385
[7,] 5.982538 5.905	[17,] 5.697358 5.665	[27,] 5.417169 5.375
[8,] 5.936134 5.875	[18,] 5.636313 5.635	[28,] 5.442492 5.395
[9,] 5.967560 5.885	[19,] 5.688756 5.675	[29,] 5.295800 5.245
[10,] 5.893248 5.825	[20,] 5.631329 5.605	

Hemos incluido las predicciones de los 29 datos de los bonos norteamericanos a 10 años, apareciendo a la izquierda el valor predicho y a la derecha el real, como vemos tienen muy poca diferencia. Tras el cálculo del MSE (Error Cuadrático Medio) obtenemos un resultado de 0.003004387.

CONCLSUIONES FINALES

Para concluir este informe podemos confirmar que puede establecerse una estructura subyacente que sintetice y agrupe los distintos plazos en virtud de sus características comunes, de manera que podemos con predecir con gran exactitud el valor futuro de los bonos norteamericanos a 10 años con tan solo 2 componentes principales, lo que facilita su trabajo.

Respondiendo a las cuestiones principales, recordemos que son:

1. ¿Tiene sentido llevar a cabo, en este caso, un análisis de componentes principales? Para justificarlo, deberá llevar a cabo las pruebas que estime oportunas, como, por ejemplo el análisis de la matriz de correlaciones, el del determinante de dicha matriz, la prueba de esfericidad de Bartlett, el KMO o el MSA;

Definitivamente tiene sentido, ya que tras realizar los respectivos test como el determinante de la matriz, la prueba de esfericidad de Barlett o el índice KMO Podemos concluir que se puede realizar un análisis de componentes principales de manera adecuada.

2. ¿Cuántos componentes permitirían explicar, adecuadamente, la estructura subyacente de los tipos de interés aquí analizados? Justifique su respuesta empleando, por ejemplo, las pruebas de la varianza explicada o del gráfico de sedimentación;

Como hemos visto durante el trabajo, trabajando con tan solo dos componentes somos capaces de explicar el más del 98% de la varianza de los datos.

3. Finalmente, ¿tiene sentido llevar a cabo una rotación de las variables subyacentes? Para responder, lleva a cabo una rotación Varimax, por ejemplo.

En nuestro caso, la rotación de las variables subyacentes no aporta ninguna mejora a la varianza explicada, por lo tanto, desaconsejamos la rotación.

REFERENCIAS

López Zafra, J.M. EL ANÁLISIS DE COMPONENTES PRINCIPALES. *Técnicas de Agrupación y Reducción de la Dimensión*.

López Zafra, J.M. EL ANÁLISIS FACTORIAL. *Técnicas de Agrupación y Reducción de la Dimensión*.