

TÉCNICAS DE AGRUPACIÓN Y REDUCCIÓN DE LA DIMENSIÓN

CURSO 2018/19

ESTRUCTURA TEMPORAL DE LOS TIPOS DE INTERÉS

Práctica 2

JORGE CASAN VÁZQUEZ

1. **EXECUTIVE SUMMARY**
2. **INTRODUCCIÓN**
3. **¿TIENE SENTIDO REALIZAR UN ANÁLISIS DE COMPONENTES PRINCIPALES?**
4. **¿CUÁNTOS COMPONENTES PERMITIRÍAN EXPLICAR LA ESTRUCTURA SUBYACENTE DE LOS TIPOS DE INTERÉS?**
5. **¿TIENE SENTIDO LLEVAR A CABO UNA ROTACIÓN DE LAS VARIABLES SUBYACENTES?**
6. **PREDICCIÓN DEL BONO A 10 AÑOS**
7. **CONCLUSIONES**
8. **BIBLIOGRAFÍA**

EXECUTIVE SUMMARY

A través de este informe lo que se pretende es resaltar la importancia del análisis de los componentes principales, pero para poder realizar el análisis pertinente deberemos de calcular la prueba de esfericidad de Barlett y ver si el resultado arrojado expresado en forma de p-valor es lo suficientemente significativo para rechazar la hipótesis nula y a su vez comprobar si podemos factorizar las variables originales a través de la medida de adecuación muestral o también llamada KMO. A su vez el MSA arroja un resultado alto, siendo un 0,88 indicando un buen nivel de multicolinealidad entre las variables. Viendo que tales resultados son positivos, el análisis de componentes principales resulta fundamental a la hora de explicar el porcentaje de la varianza explicada de cada una de las dimensiones, siendo la Dimensión 1 la que mejor explica la varianza en un 80,73% y en segundo lugar la Dimensión 2, en un 17,54%, todo aquello que no recoge la varianza explicada de la anterior dimensión. Además, a través del gráfico de sedimentación ayuda a determinar la capacidad explicativa de cada componente a medida que se va incorporando al modelo, siendo el componente 1 y 2 los que más capacidad explicativa aportan, a partir de aquí la pendiente va perdiendo fuerza hasta convertirse en una línea con la dimensión 9. También realizaremos una rotación ortogonal de los ejes factoriales a través del análisis Varimax con el objetivo de maximizar la varianza explicada por cada uno de ellos, equilibrando las diferencias de los autovalores, siendo el depósito a 3 meses la variable más destacada. Los resultados arrojados después de la rotación por observaciones no varían. Sin embargo, realizando la rotación por variables vemos que la Dimensión 1 pierde capacidad explicativa y la Dimensión 2 aumenta en un 20% aproximadamente el porcentaje de la varianza explicada. Finalmente, podremos predecir el valor del bono a 10 años, y para ello emplearemos las observaciones suplementarias, en donde se comprobará que el error entre los datos reales y nuestro análisis de predicción es relativamente pequeño.

INTRODUCCIÓN

Hemos rellenado a partir de la observación 785 de la columna depósitos a 1 mes los datos del LIBOR USD serie completa para poder trabajar. A partir de aquí eliminamos la tabla referente a fecha y el bono a 10 años, puesto que lo que pretendemos a través de este análisis es predecir el comportamiento del bono a 10 años.

Antes de contestar directamente a la pregunta deberemos preguntarnos si la correlación entre las variables es lo bastante significativa como para realizar la prueba de Barlett que no es más que la factorización de la matriz de los coeficientes de correlación.

¿TIENE SENTIDO REALIZAR UN ANÁLISIS DE COMPONENTES PRINCIPALES?

Nos cargamos las librerías FactoMiner y Factoextra, así como la librería tidyverse y dplyr. Vamos a ver cómo quedan estructurados los datos a través del comando head().

	X	DEPO.1M	DEPO.3M	DEPO.6M	DEPO.12M	IRS.2Y	IRS.3Y	IRS.4Y	IRS.5Y	IRS.7Y	IRS.10Y
1	02/01/1995	6.000	6.500	7.000	7.750	8.170	8.24	8.25	8.22	8.24	8.25
2	03/01/1995	5.938	6.500	7.000	7.813	8.220	8.28	8.29	8.29	8.30	8.31
3	04/01/1995	5.938	6.500	7.000	7.813	8.130	8.20	8.23	8.24	8.26	8.28
4	05/01/1995	5.898	6.438	6.938	7.688	8.145	8.22	8.26	8.27	8.29	8.30
5	06/01/1995	5.875	6.438	6.938	7.688	8.070	8.20	8.25	8.27	8.29	8.29
6	09/01/1995	5.875	6.375	6.875	7.688	8.135	8.25	8.29	8.31	8.33	8.34

Observamos que a mayor plazo de vencimiento mayor tipo de interés, cuestión que tiene todo el sentido por el coste de oportunidad que genera el dinero, vemos que a partir de 5 años en adelante se estabiliza, de los bonos del año 95.

También podemos ver la estructura temporal del vencimiento de los bonos en el año 98, siendo este el último de nuestra tabla.

	X	DEPO.1M	DEPO.3M	DEPO.6M	DEPO.12M	IRS.2Y	IRS.3Y	IRS.4Y	IRS.5Y	IRS.7Y	IRS.10Y
973	23/09/1998	5.59375	5.500	5.406	5.254	5.105	5.145	5.215	5.295	5.405	5.555
974	24/09/1998	5.52734	5.438	5.344	5.156	5.025	5.075	5.155	5.245	5.355	5.495
975	25/09/1998	5.38672	5.313	5.250	5.063	4.935	4.975	5.065	5.145	5.245	5.385
976	28/09/1998	5.37500	5.313	5.250	5.063	4.925	4.965	5.045	5.125	5.235	5.375
977	29/09/1998	5.34375	5.313	5.250	5.063	4.945	4.985	5.075	5.155	5.255	5.395
978	30/09/1998	5.37500	5.313	5.246	5.063	4.815	4.855	4.925	5.005	5.115	5.245

La idea es eliminar la columna correspondiente al bono a q0 años puesto que es la columna objeto de predicción, la tabla nos quedaría de la siguiente manera:

	X	DEPO. 1M	DEPO. 3M	DEPO. 6M	DEPO. 12M	IRS. 2Y	IRS. 3Y	IRS. 4Y	IRS. 5Y
1	02/01/1995	6.000	6.500	7.000	7.750	8.170	8.24	8.25	8.22
2	03/01/1995	5.938	6.500	7.000	7.813	8.220	8.28	8.29	8.29
3	04/01/1995	5.938	6.500	7.000	7.813	8.130	8.20	8.23	8.24
4	05/01/1995	5.898	6.438	6.938	7.688	8.145	8.22	8.26	8.27
5	06/01/1995	5.875	6.438	6.938	7.688	8.070	8.20	8.25	8.27
6	09/01/1995	5.875	6.375	6.875	7.688	8.135	8.25	8.29	8.31

Hay que tratar la columna de fechas para realizar un análisis sobre los datos. A continuación la idea es mostrar a través de un dataframe la estructura que presentan los datos:

La prueba de la esfericidad de Barlett contrasta la hipótesis nula de que la matriz de correlaciones sea una matriz identidad, en cuyo caso no existiría correlación alguna entre las variables y el modelo de análisis factorial no sería pertinente.

Barlett calcula un estadístico basado en el determinante de la matriz de correlaciones

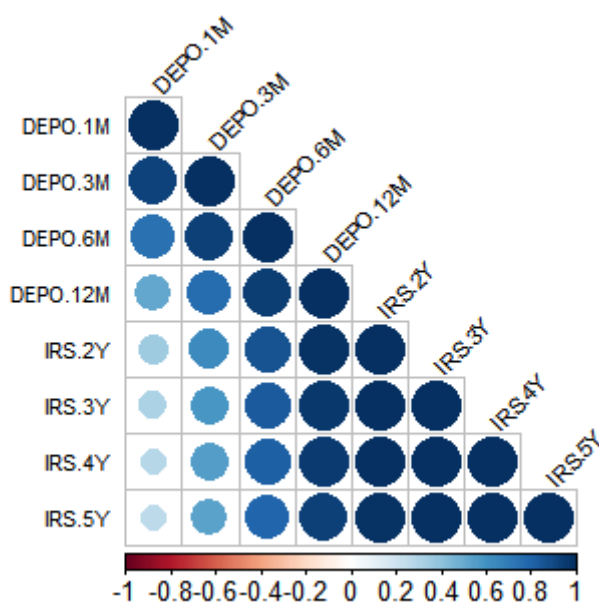
$$-[n-1-(2k+5)/6] \ln |\mathbf{R}| \sim \chi^2_{(k^2-k)/2}$$

En donde “n” es el tamaño de la muestra, 949 observaciones, “k” el número de variables de la matriz, 10 variables que muestran los diferentes tipos de interés así como el vencimiento de ellos clasificados por su plazo y “R” la matriz de correlaciones, en nuestro caso vendría expresada a través del siguiente gráfico:

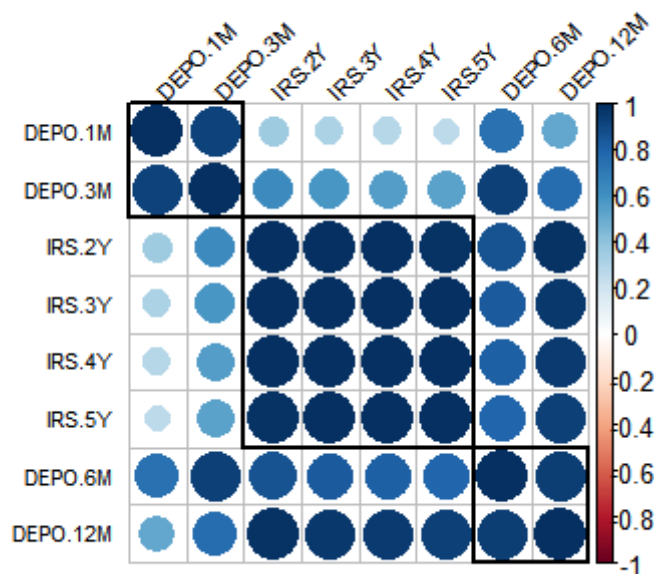
```
> correlaciondatabonos
```

	DEPO. 3M	DEPO. 6M	DEPO. 12M	IRS. 2Y	IRS. 3Y	IRS. 4Y	IRS. 5Y	IRS. 7Y	IRS. 10Y
DEPO. 3M	1.0000000	0.9279445	0.7643855	0.6333915	0.5895566	0.5591534	0.5341612	0.4954058	0.4625498
DEPO. 6M	0.9279445	1.0000000	0.9440732	0.8607230	0.8245985	0.7953775	0.7693432	0.7281305	0.6842335
DEPO. 12M	0.7643855	0.9440732	1.0000000	0.9731443	0.9515135	0.9292639	0.9074514	0.8714994	0.8255048
IRS. 2Y	0.6333915	0.8607230	0.9731443	1.0000000	0.9948671	0.9836059	0.9699235	0.9443331	0.9063116
IRS. 3Y	0.5895566	0.8245985	0.9515135	0.9948671	1.0000000	0.9966304	0.9891730	0.9719242	0.9425898
IRS. 4Y	0.5591534	0.7953775	0.9292639	0.9836059	0.9966304	1.0000000	0.9977443	0.9875916	0.9660694
IRS. 5Y	0.5341612	0.7693432	0.9074514	0.9699235	0.9891730	0.9977443	1.0000000	0.9956517	0.9806293
IRS. 7Y	0.4954058	0.7281305	0.8714994	0.9443331	0.9719242	0.9875916	0.9956517	1.0000000	0.9941705
IRS. 10Y	0.4625498	0.6842335	0.8255048	0.9063116	0.9425898	0.9660694	0.9806293	0.9941705	1.0000000

Viéndolo desde un punto de vista gráfico podemos verlo de la siguiente forma:

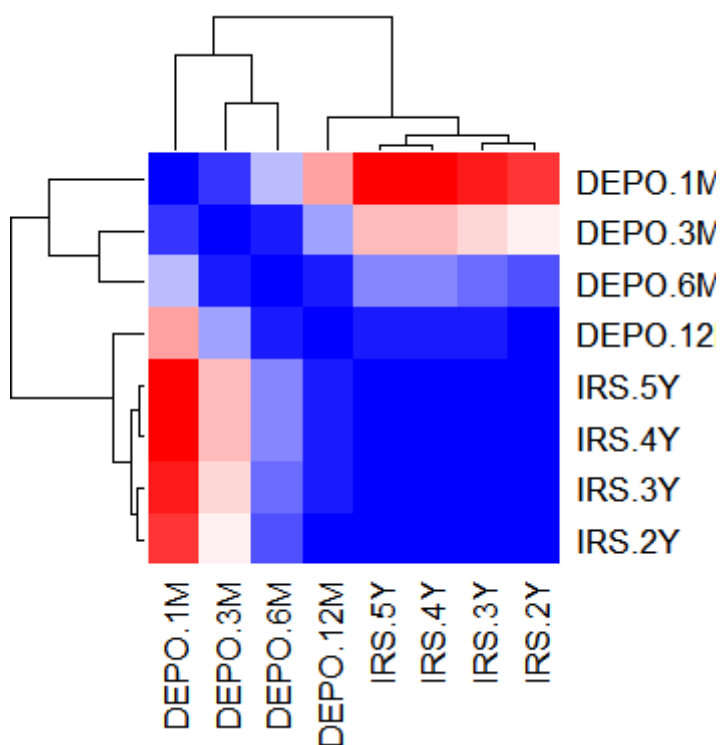
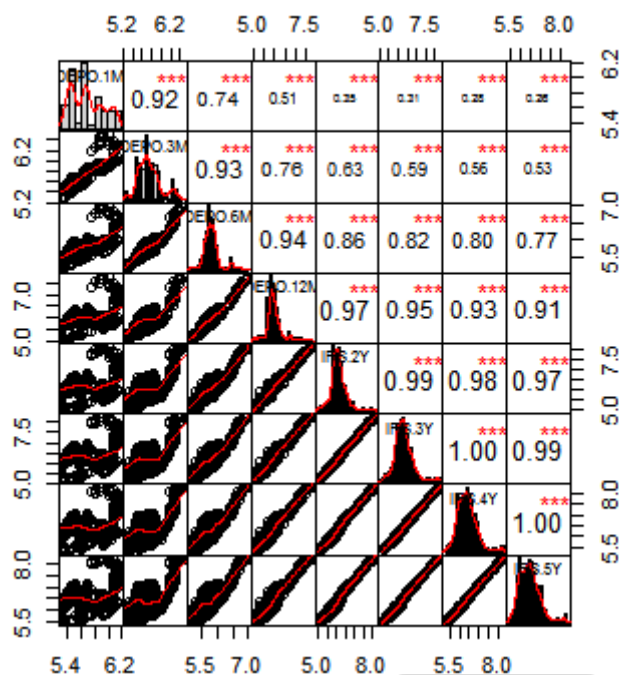


También podemos añadir tres rectángulos de color a las agrupaciones generadas con `order= "hclust"`, que sustituye el orden original.



En este gráfico observamos la distribución de cada variable en la diagonal. Por debajo, los diagramas de dispersión por pares con línea de ajuste. Por encima, el valor del

coeficiente de correlación con los nds como estrellas. También podemos comprobar el mapa del calor:



Fuente: Elaboración propia

Para la prueba de la esfericidad de Barlett he obtenido el siguiente resultado:

Si el p-value es < al nivel de significación, que generalmente suele ser del 0,05, en tal caso rechazamos la hipótesis nula, en caso contrario no se podría realizar el análisis factorial. Como en este caso el p-valor es muy bajo, rechazamos la hipótesis nula.

El índice de Kaiser-Meyer-Olkin o medida de adecuación muestral KMO tiene el mismo objetivo que la prueba de esfericidad de Barlett en donde determina si podemos factorizar las variables originales, para ello debemos de tener presente la matriz de correlaciones entre las 9 variables, las cuales he denominado como “databonos”. Este análisis estadístico compara los valores de las correlaciones entre las variables y sus correlaciones parciales. Si el KMO está próximo a 1 el análisis de componentes principales será relevante para el análisis estadístico, en caso contrario, no.

1 >= KMO >= 0.9 muy bueno
 0.9 >= KMO >= 0.8 meritorio
 0.8 >= KMO >= 0.7 mediano
 0.7 >= KMO >= 0.6 mediocre
 0.6 >= KMO > 0.5 bajo
 KMO <= 0.5 inaceptable

Como se puede observar todas las variables son mayores a 0,70; esto quiero decir que la relación que existe entre las variables es alta, y por ende, como he dicho antes el análisis de componentes principales se podrá hacer.

La correlación anti-imagen o también llamada MSA es 0,88, que no es más que el coeficiente de correlación parcial, la fórmula es la siguiente:

$$MSA_i = \frac{\sum_{j \neq i} r_{ij}^2}{\sum_{j \neq i} r_{ij}^2 + \sum_{j \neq i} a_{ij}^2}$$

r_{ij} = correlación simple.
 a_{ij} = correlación parcial.

Tomado todo ello en su conjunto podemos afirmar que tiene sentido realizar el análisis de componentes principales ya que permite tratar el conjunto de las 10 variables.

Con esta técnica podemos sacar conclusiones estadísticamente significativas y saber las correlaciones que existen entre las variables.

¿CUÁNTOS COMPONENTES PERMITIRÍAN EXPLICAR LA ESTRUCTURA SUBYACENTE DE LOS TIPOS DE INTERÉS?

Para realizar el análisis de componentes principales tendremos que determinar el porcentaje de varianza explicada de cada uno de los componentes. Como se puede observar, la dimensión 1 explica el 80,73% de la varianza, y la Dimensión 2 el resto de la información de la varianza explicada que la Dimensión 1 no recoge, por lo tanto, sumadas las dos en su conjunto podemos decir que su varianza acumulada explica en más de un 90%, siendo el porcentaje de la varianza explicada, más precisamente el 98,27%, por lo que es bastante significativo, podemos desprendernos del resto de las dimensiones, que tomadas en su conjunto explican menos de un 2% la varianza total explicada.

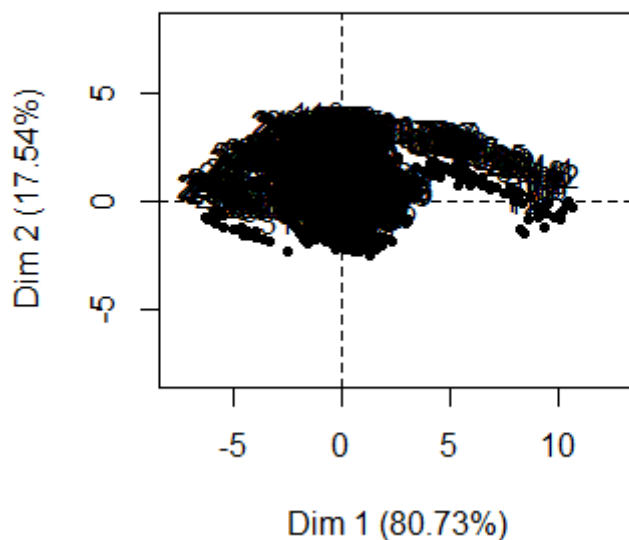
Lo podemos ver de la siguiente forma:

Realizando un análisis individualizado de las 10 primeras observaciones observamos que la dimensión 1 explica gran parte de la varianza explicada.

A su vez, realizando un análisis de cada una de las variables vemos que las dimensiones 1 y 2 son las estadísticamente más significativas.

Desde un punto de vista gráfico, se resume la tabla del análisis individualizado de cada una de las observaciones.

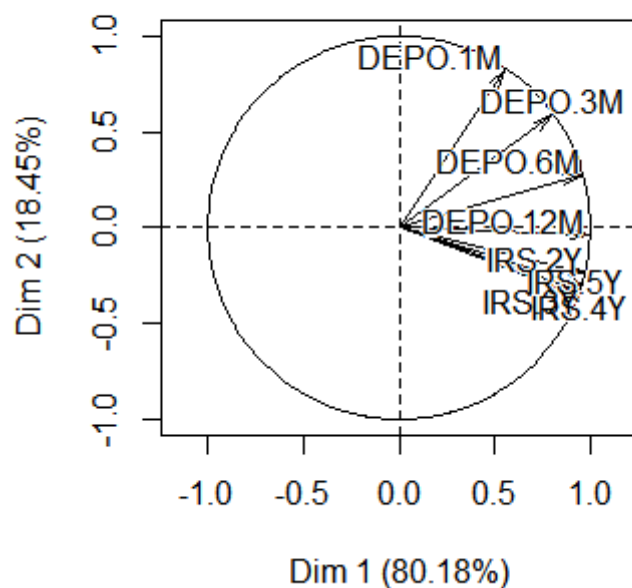
Individuals factor map (PCA)

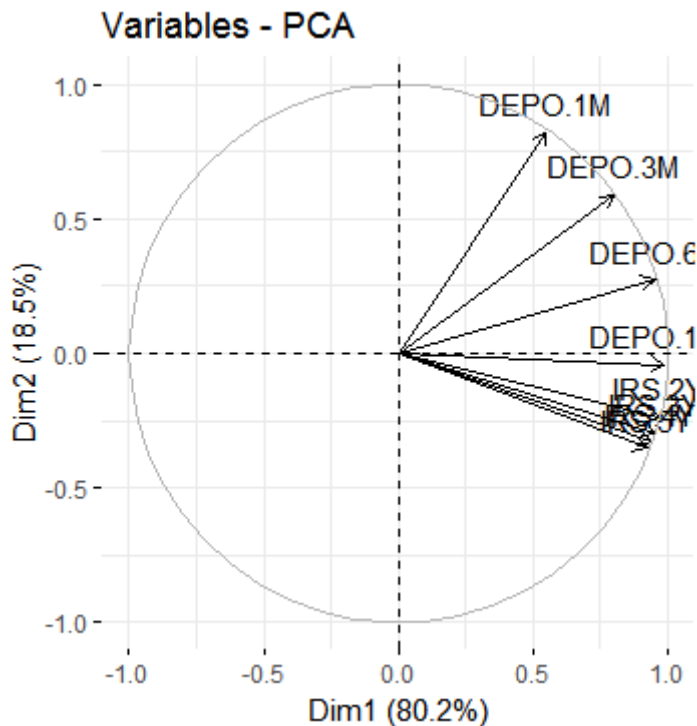


Fuente: Elaboración propia

También podemos ver desde un punto de vista gráfico el porcentaje de la varianza explicada de la dimensión 1 y 2 por variables.

Variables factor map (PCA)

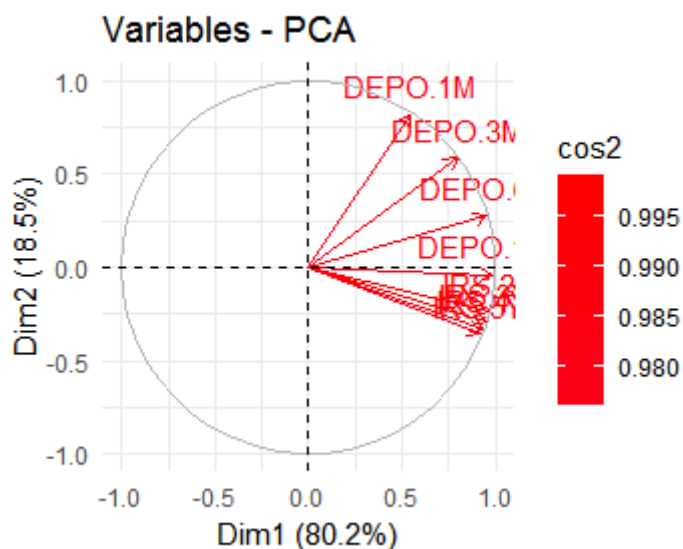




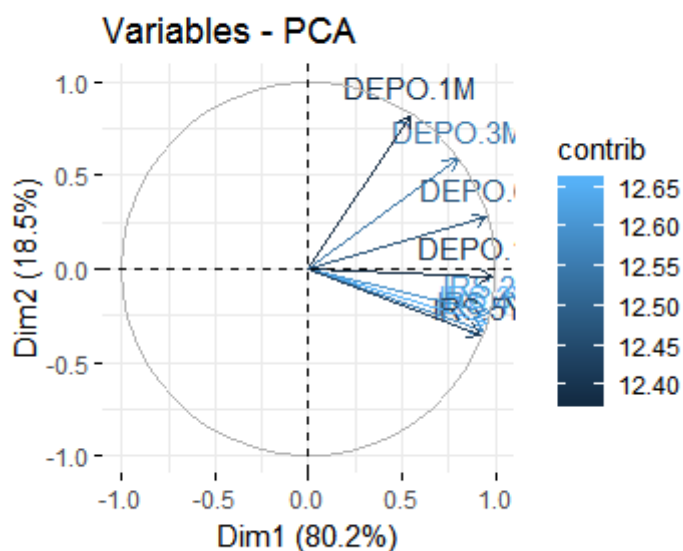
Fuente: Elaboración propia

Para ver el porcentaje de la varianza acumulada elegiremos tantos autovectores como autovalores sean mayores de 1 de la matriz de correlaciones de las variables. Vemos que el porcentaje de la varianza acumulada de la dimensión 1 y dimensión 2 explican en un 98,7% la varianza explicativa.

Cuando, para representar una variable, son necesarios más de 2 componentes principales, la variable está alejada del borde del círculo, no siendo el caso; las variables que en los dos componentes principales que definen el mapa quedan cercanas al centro tienen poca posibilidad de ser correctamente representadas por ellos. A través del siguiente gráfico se resalta la contribución de los factores extraídos en la explicación de cada variable, es decir, su comunalidad.



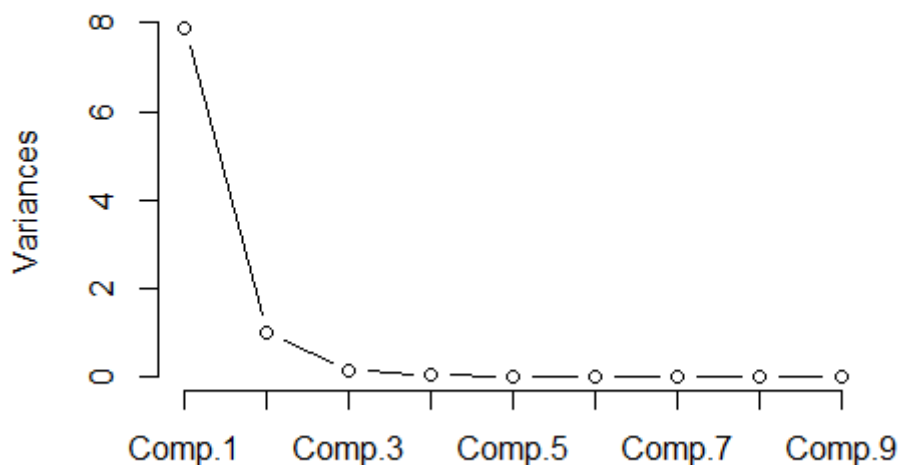
En cuanto a la contribución de cada variable a la explicación de los ejes principales , podemos verlo en escala de color:



Por otra parte, a través del gráfico de sedimentación, el cual se suele utilizar como contraste gráfico, se observa el número de componentes principales a retener. Según este

criterio se retiene todas las componentes que estén situadas en la zona de sedimentación, entendiendo por esta la parte del gráfico en la que los componentes empiezan a no presentar pendientes fuertes.

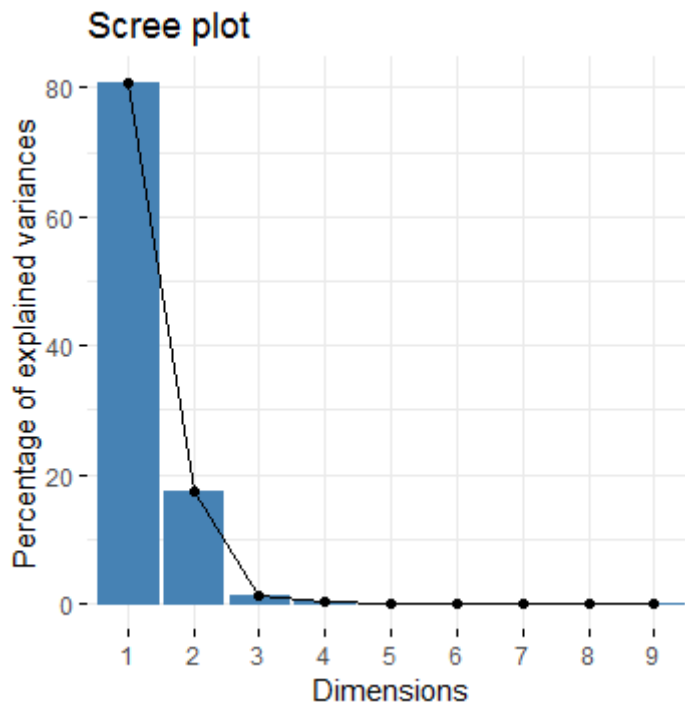
Gráfico_de_Sedimentacion



Fuente: Elaboración propia

Según el gráfico de sedimentación de las componentes se retienen todas aquellas variables que se encuentren en la zona de sedimentación, siendo la parte del gráfico en la que los componentes empiezan a no presentar fuertes pendientes, que como podemos observar se puede ver a partir de la componente 4 en adelante, esto es, hasta la componente 9.

También podemos representarlo a través del porcentaje de la varianza explicada. La dimensión 1 explica el 80,73% de la varianza explicada y todo aquello no explicado previamente queda explicado por la segunda dimensión y así sucesivamente.



Fuente: Elaboración propia

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

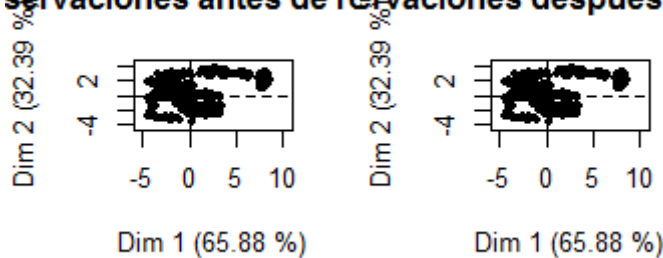
La rotación Varimax facilita la interpretación del significado de los factores seleccionados. Fue desarrollado por Kaiser (1958) en donde efectúa una rotación ortogonal de los ejes factoriales en donde se minimiza el número de variables que tiene saturaciones altas en cada factor. Simplifica la interpretación de las variables observadas optimizando la interpretación por filas. El principal objetivo de la rotación Varimax es conseguir que la correlación de cada una de las variables sea lo más próxima a 1 con sólo uno de los factores y próxima a 0 para el resto.

Con las observaciones antes y después de la rotación, tanto por el número de observaciones como de variables, hemos obtenido 4 gráficos, en donde se muestra el proceso de rotación. Cada gráfico contiene las saturaciones de las variables en los factores.

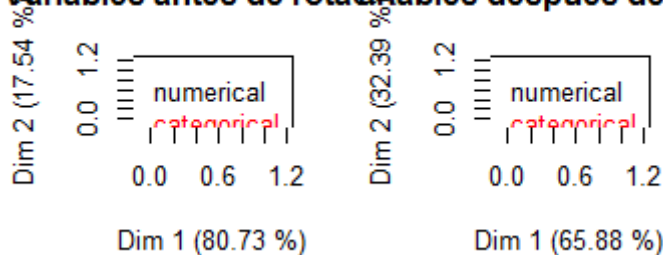
Realizando la rotación por número total de observaciones no vemos grandes diferencias, es decir, el porcentaje total de la varianza explicada por cada una de las dos dimensiones no varía antes y después de la rotación. Sin embargo, si efectuamos la rotación por variables los resultados son diferentes. Como hemos dicho al inicio del informe, antes de la rotación la dimensión 1 explica el 80,2% de la varianza explicada y la dimensión 2 el 18,5%. Después de la rotación vemos que la Dimensión 1 pierde capacidad explicativa perdiendo en más de un 20%, pero por otro lado la Dimensión 2 aumenta en un 20% de capacidad explicativa de la varianza.

Tanto antes como después de la rotación, las dimensiones 1 y 2, tomadas en su conjunto, explican un 98,7% la varianza total explicada.

Observaciones antes de rotación Observaciones después de rotación



Variables antes de rotación Variables después de rotación

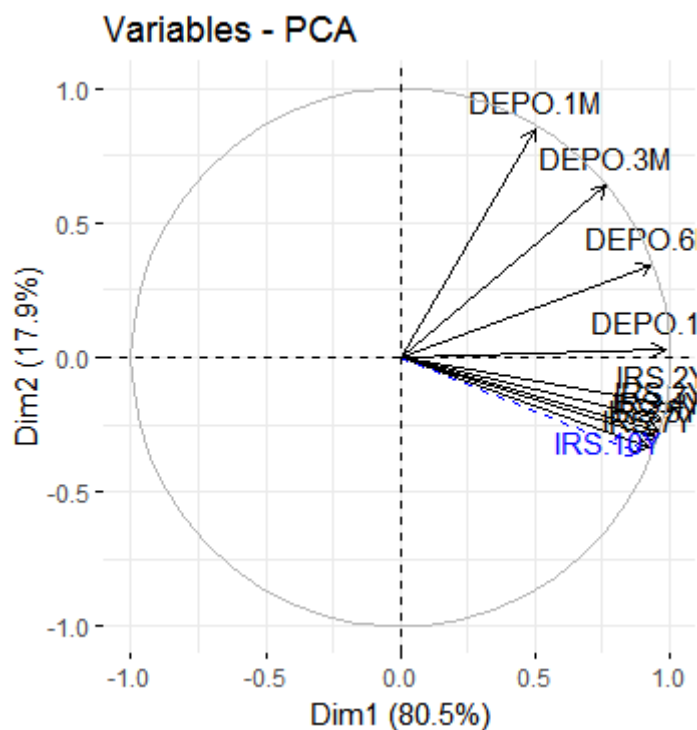


Fuente: Elaboración propia

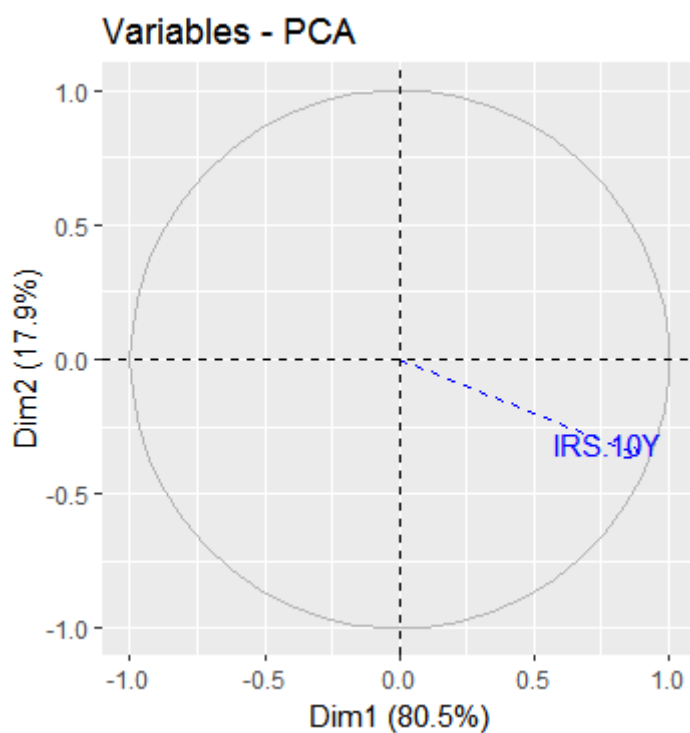
PREDICCIÓN DEL BONO A 10 AÑOS

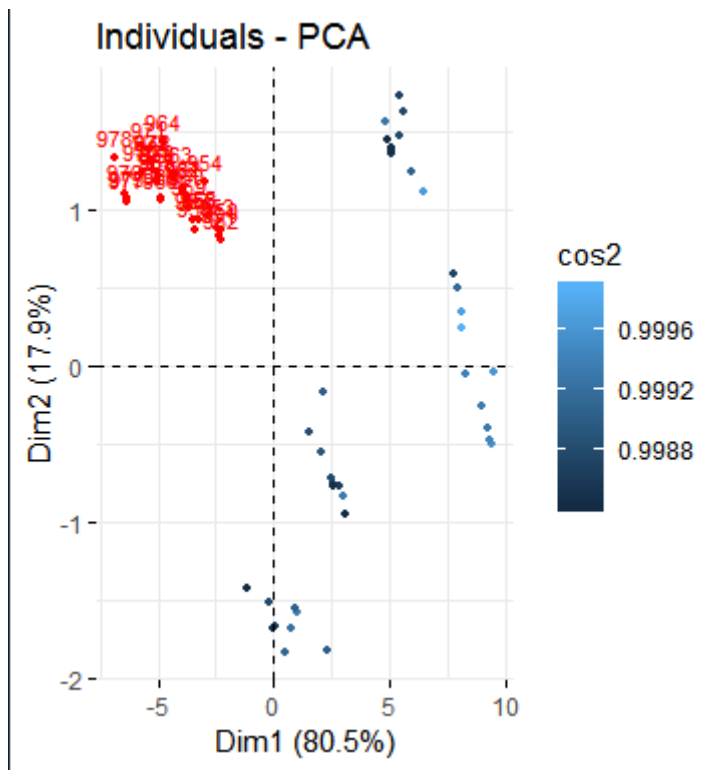
Para predecir el bono a 10 años tendremos que trabajar con las observaciones suplementarias, es decir, desde la observación 950 a la 978. Tendremos que realizar un modelo lineal en R partiendo del resto de las variables.

Para visualizarla junto con el resto de las variables activas:



Si quisiéramos visualizar solamente las suplementarias:





Podemos extraer la conclusión que nuestro modelo y el análisis que ha habido por delante son relevantes a la hora de efectuar la predicción del bono a 10 años.

CONCLUSIONES

Tomado todo ello en su conjunto hemos comprobado como el análisis de componentes principales nos permite descubrir y priorizar las características con las que debe de contar el vencimiento así como el interés de los bonos a diferentes meses y años reduciendo la información estadísticamente no significativa para poder saber qué componentes son importantes para explicar el modelo.

BIBLIOGRAFÍA

“The Information in Long-Maturity Forward Rates, The American Economic Review, Vol. 77, No. 4”. Autores: Fama y Bliss de 1987

“Common factors affecting bond returns, Journal of fixed income, 1991”. Autores: Litterman y Scheinkman

“Modeling Bond Yields in Finance and Macroeconomics, American Economic Review 95.2, 2005”. Autores: Diebold, Piazzesi y Rudebusch

“Forecasting the term structure of government bond yields, Journal of Econometrics, 130.2, 2006”. Autores: Diebold y Li

Paper de 2014 de Moody’s Analytics Principal Component Analysis for Yield Curve Modelling

Paper de 2009 de Novosyolov y Satchkov, Global Term Structure Modeling Using Principal Component Analysis.

“Análisis de componentes principales y análisis factorial” Guía de trabajo en R. Profesor: Giorgio Boccardo

“Análisis de componentes principales”. Autor: Juan Manuel López Zafra

“El análisis factorial”. Autor: Juan Manuel López Zafra

Enlaces de interés:

<http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/MULTIVARIANTE/ACP/ACP.pdf>