

Recuadro 2

Caracterización y desempeño de las sociedades fiduciarias: una aproximación mediante técnicas de *machine learning*

Daniela X. Gualtero Briceño
Javier E. Pirateque Niño*

1. Introducción

Al comparar los indicadores financieros de las sociedades fiduciarias (SFD), usualmente se supone que este tipo de entidades pertenecen a un grupo cuyas características son similares. No obstante, en la práctica cada entidad adopta un modelo de negocio con diferentes particularidades desde el punto de vista de la exposición del activo, las fuentes de fondeo, las contrapartes, los productos ofrecidos, el tamaño, entre otros.

Farnè y Vouldis (2017) establecen que, debido a esta heterogeneidad, el análisis financiero se debería hacer con referencia al modelo de negocio y no al sector; entendiendo por modelo de negocio a las decisiones estratégicas relacionadas con el conjunto de actividades realizadas por la organización y la estructura de su hoja de balance.

Acorde con lo anterior, el presente recuadro propone una metodología estadística que combina una técnica para la identificación de *clusters* con una de reducción de dimensionalidad para obtener grupos de entidades con modelos de negocios afines. Para esto, en un primer paso se realizan varios análisis con el fin de definir el número óptimo de *clusters*. Posteriormente, se aplica la metodología propuesta, y la clasificación resultante se valida al comparar algunos indicadores de desempeño entre los *clusters* obtenidos.

* Los autores pertenecen al Departamento de Estabilidad Financiera del Banco de la República. Las opiniones aquí contenidas son responsabilidad exclusiva de los autores y no comprometen al Banco de la República ni a su Junta Directiva.

2. Caracterización de los datos y definición del número óptimo de *clusters*

Para la realización del ejercicio se tomó el balance general a dos dígitos de cada una de las SFD, incluyendo los saldos de los fondos administrados y custodiados separados por tipo¹. Cada variable se dividió entre el activo y el resultado obtenido se promedió para todo el año 2019². Con el fin de reducir la dimensionalidad y facilitar el análisis de la información, se utilizó la técnica de componentes principales (PCA, por su sigla en inglés) para obtener el primer componente principal de las series agrupadas en tres grupos: exposición en el activo (Exp. Activo), estructura y características del fondeo (Fondeo) y composición de los fondos administrados (Fondos)³. De esa manera, la base de datos final contiene tres variables para 28 SFD.

Con el fin de tener un primer acercamiento a las similitudes y diferencias entre las entidades, se analizó las distancias entre ellas. Para esto, se calculó la distancia de los datos entre pares de entidades y luego se agruparon progresivamente para formar un árbol jerárquico.

Las distancias sirven para observar qué tan cercanas (similares) o alejadas (diferentes) se encuentran las entidades una de otra. Existen varios métodos para el cálculo de las distancias, el más común es el de la distancia euclidiana (d_{ij}). Para calcular esta medida entre un par de entidades i y j , se usa la siguiente expresión:

$$d_{ij} = \sqrt{\sum_{w=1}^W (x_{iw} - x_{jw})^2}$$

Donde W es el número total de variables disponibles para calcular la distancia, x_{iw} es el valor de la variable w para la entidad i y de forma análoga x_{jw} es el valor de la misma variable para la entidad j .

El resultado del cálculo de las distancias para cada par de entidades indica que las mayores distancias se dan entre las SFD del número 22 en adelante con las entidades de la 1 a la 22. Así, preliminarmente se pueden identificar dos grupos.

Luego de calcular las distancias entre pares de entidades, mediante el método de agrupamiento jerárquico se cons-

1 Los tipos de fondos usados en el ejercicio fueron: fiducia de inversión, fiducia inmobiliaria, fiducia de administración, fiducia de garantía, fondos de pensiones voluntarias, recursos de la seguridad social, fondos de inversión colectiva, fondos de capital privado, otros recursos de la seguridad social y activos custodiados.

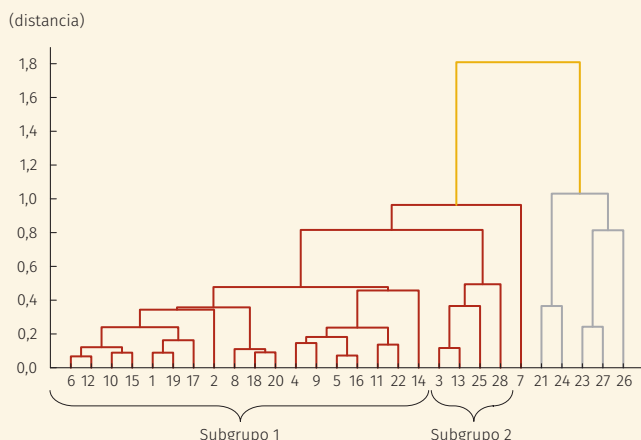
2 No se tomaron datos de antes de 2019 debido a los cambios que hubo a comienzo de ese año en el registro de los negocios de las SFD (Circular Externa 034 de 2018 de la Superintendencia Financiera de Colombia), los cuales hicieron que la información de los fondos administrados de antes y después de esa fecha no fueran comparables.

3 El primer componente principal de cada grupo de variables explicó el 72,6%, el 76,8% y el 60,7% de la variabilidad para Exp. Activo, Fondeo y Fondos, en su orden.

truye un árbol jerárquico o dendrograma que aporta información sobre el número de grupos que se pueden identificar en los datos y resume la información de la matriz de distancias. El método que se usó para construir este dendrograma consiste en: 1) agrupar pares de entidades cuya distancia sea la menor; 2) calcular la distancia entre los grupos obtenidos en el paso anterior; 3) asociar grupos cuya distancia sea bastante corta⁴. El último paso se repitió hasta que se obtuvo un solo grupo.

En el Gráfico R2.1 se presenta el dendrograma correspondiente a los datos de las SFD. La altura de cada rama horizontal muestra la distancia entre los grupos que se desprenden de la rama. Así, se identifican dos grandes grupos los cuales se resaltan en colores rojo y gris y concuerdan con lo encontrado al calcular las distancias entre pares de entidades. Dentro del grupo rojo se pueden identificar dos subgrupos adicionales.

Gráfico R2.1
Dendrograma SFD



Fuente: Superintendencia Financiera de Colombia; cálculos de los autores.

3. Metodología

Para la creación de los *clusters* se utilizó la metodología de *k-medias*. Esta es una metodología de *machine learning* de clasificación no supervisada, la cual es usada para encontrar patrones o grupos en los datos. La pertenencia a un grupo se determina con base en la distancia de cada punto al centro del *cluster*. La implementación de la metodología se realiza usando un algoritmo que minimiza la suma de cuadrados dentro de los grupos (*SCDG*) para todas las variables. Matemáticamente, esta se encuentra dada por:

$$SCDG = \sum_{g=1}^G \sum_{j=1}^P \sum_{i=1}^{n_g} (x_{ijg} - \bar{x}_{jg})^2$$

Donde G es el número total de grupos; P es el número total de variables; n_g es el número total de elementos en el grupo

4 Para calcular la distancia entre grupos existen varios métodos, los principales se describen en León (2017).

g ; x_{ijg} es el valor de la variable en el elemento i del grupo g , y \bar{x}_{jg} es la media de esta variable j en el grupo. Bajo esta metodología el número de grupos es prefijado (Peña, 2002).

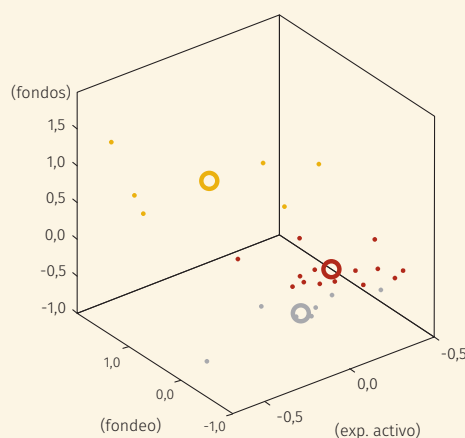
4. Resultados

Con base en la información del dendrograma, se procedió a aplicar la metodología de PCA y *k-medias* para las SFD con un número de *clusters* igual a tres⁵. El resultado se presenta en el Gráfico R2.2, donde cada eje representa cada una de las tres variables de la base de datos, cada punto una SFD y los aros la media de cada *cluster*. El *cluster* uno (representado mediante color gris) se compone por siete entidades, el *cluster* dos (representado por el color rojo) se forma por catorce entidades y, finalmente, el *cluster* tres (representado por el color amarillo) contiene seis entidades.

Dados los resultados del agrupamiento mediante *k-medias*, se procede a analizar las características que identifican y diferencian a cada *cluster* de los demás.

En general, se observa que las entidades del *cluster* uno son las de mayor tamaño y apalancamiento⁶, seguidas por las entidades de los *cluster* dos y tres. El activo promedio de las entidades del *cluster* uno es dos y quince veces mayor que el de las entidades pertenecientes a los *cluster* dos y tres, respectivamente. Por su parte, las entidades del *cluster* dos son las que poseen el portafolio de inversiones

Gráfico R2.2
Clusters usando *k-medias* y PCA para las SFD



Fuente: Superintendencia Financiera de Colombia; cálculos de los autores.

5 Para corroborar el número óptimo de *clusters* se hizo el ejercicio de *k-medias* con 1,2 y hasta 20 *clusters* y se calculó la SCDG de cada entidad al centro más cercano. A medida que aumenta el número de *clusters*, disminuye la SCDG. El número óptimo de *clusters* es aquel donde la SCDG empieza a disminuir marginalmente. En este caso dicho cambio marginal empezó a observarse a partir del tercer *cluster*.

6 El apalancamiento se define como la razón entre activo y patrimonio. El pasivo de las SFD está conformado mayormente por cuentas por pagar, instrumentos financieros, obligaciones laborales y provisiones.

más diversificado (Gráfico R2.3). En cuanto a ingresos, estos se componen principalmente de comisiones para los tres *clusters*. A continuación, se caracteriza en mayor detalle cada uno:

Cluster uno: su activo se compone principalmente de inversiones. Estas inversiones se encuentran poco diversificadas, siendo en su mayoría inversiones en renta variable. Las entidades de este *cluster* se dedican principalmente a la fiducia de administración e inmobiliaria (Gráfico R2.4).

Cluster dos: son SFD que se caracterizan por un activo compuesto en su mayoría por inversiones. Estas últimas se encuentran más diversificadas que las del *cluster* uno, teniendo una participación tanto en renta variable como en deuda privada y deuda pública. En cuanto a los fondos administrados, estas entidades en promedio se caracterizan por tener una participación mayor que las demás en la administración de recursos de la seguridad social.

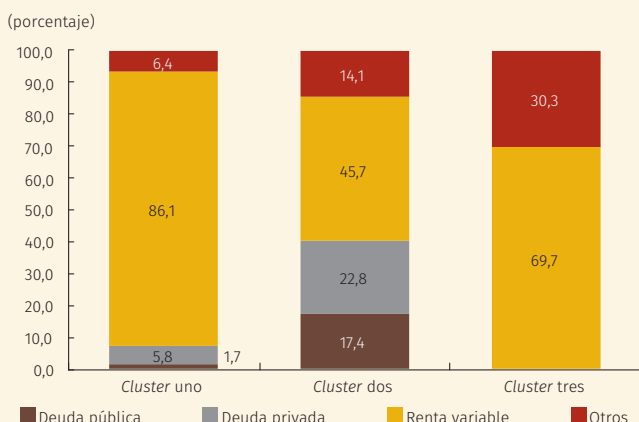
Cluster tres: el activo de estas SFD, a diferencia de los otros dos *clusters*, se compone en mayor medida de efectivo. Además de comisiones, perciben ingresos por concepto de operaciones en el mercado monetario. En promedio se dedican mayoritariamente a la fiducia de inversión, seguido de fiducia de administración y de administración de fondos de pensiones voluntarias. Este *cluster* también se caracteriza por contener entidades relativamente nuevas, con menos de cinco años de existencia.

Después de haber caracterizado a cada *cluster*, se procede a analizar y comparar cuatro indicadores de desempeño financiero y de eficiencia, estos son: quebranto patrimonial (QP), ROA, margen operativo⁷ y razón de cobertura⁸. Para hacer este análisis: 1) se obtuvo el promedio de cada indicador por entidad para los datos mensuales desde enero de 2015, y 2) los promedios obtenidos en el paso anterior se clasificaron por *cluster*, para luego plasmarlos en los Gráficos R2.5 a R2.8.

En general, se observa un mejor comportamiento de los cuatro indicadores para las entidades del *cluster* uno; el ROA y el margen operativo son similares para el *cluster* uno y dos. Por su parte, el *cluster* 3 se caracteriza por presentar bajos desempeños para los cuatro indicadores. En específico se encuentra lo siguiente:

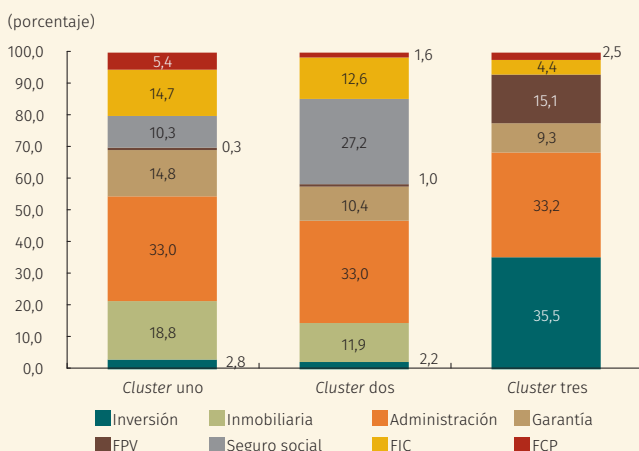
QP: la mediana del *cluster* uno es mayor que la de los demás *clusters*, lo siguen el *cluster* dos y el tres. Esto sugiere que, en promedio, las SFD del *cluster* uno tienen una mayor capacidad para crear valor a partir de su capital social, y sugiere también un mayor nivel de vulnerabilidad financiera para las entidades del *cluster* tres⁹.

Gráfico R2.3
Composición de las inversiones
(promedio por *cluster* desde 2015)



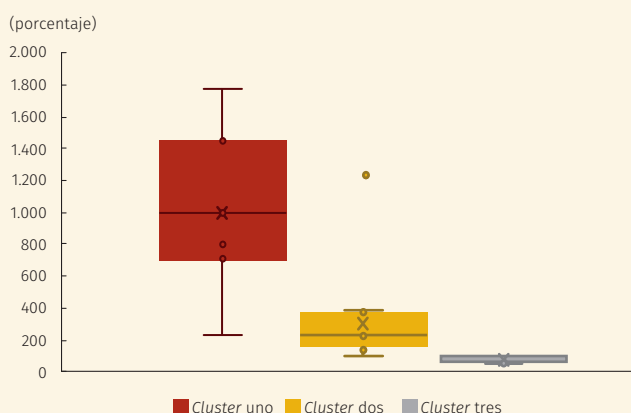
Fuente: Superintendencia Financiera de Colombia; cálculos de los autores.

Gráfico R2.4
Estructura de los fondos administrados por grupos
(promedio 2019)



Fuente: Superintendencia Financiera de Colombia; cálculos de los autores.

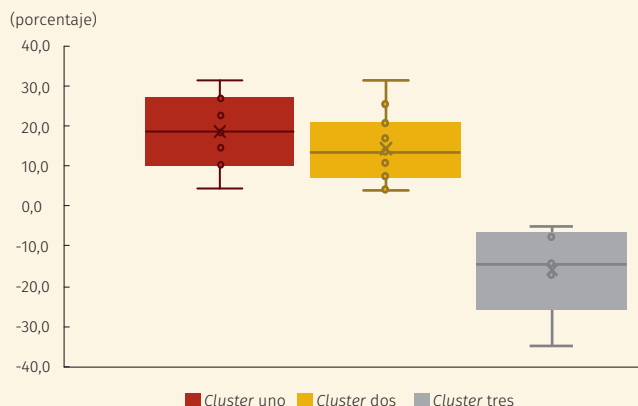
Gráfico R2.5
Indicador quebranto patrimonial
(promedio por entidad desde enero de 2015)



Fuente: Superintendencia Financiera de Colombia; cálculos de los autores.

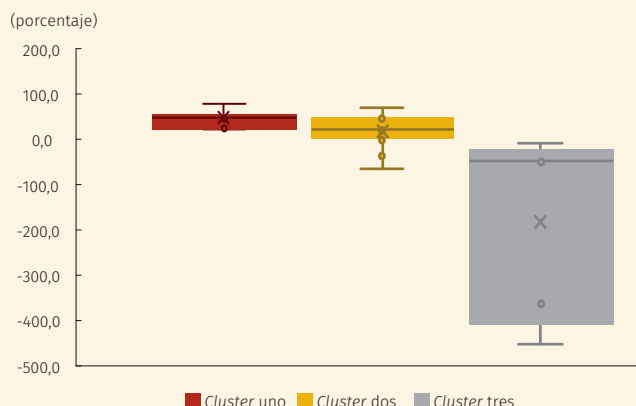
- El margen operativo se define como: $(\text{ingresos operativos} - \text{gastos operativos}) / (\text{ingresos operativos})$
- La razón de cobertura se define como: $\text{comisiones} / \text{gastos de personal}$
- De acuerdo con el Código del Comercio, un nivel inferior al 50% de este indicador es causal de disolución de las sociedades.

Gráfico R2.6
Indicador ROA
(promedio por entidad desde diciembre de 2015)



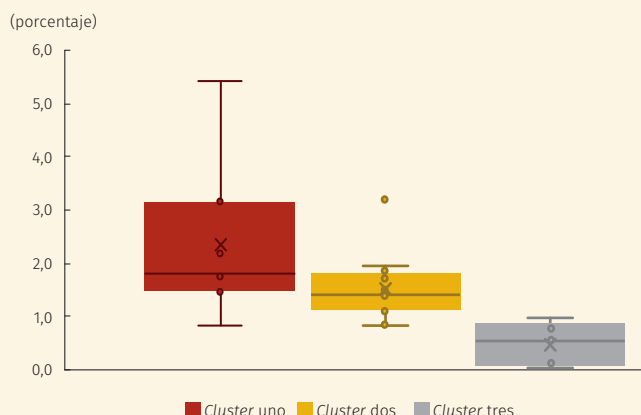
Fuente: Superintendencia Financiera de Colombia; cálculos de los autores.

Gráfico R2.7
Indicador margen operativo
(promedio por entidad desde diciembre de 2015)



Fuente: Superintendencia Financiera de Colombia; cálculos de los autores.

Gráfico R2.8
Indicador razón de cobertura
(promedio por entidad desde diciembre de 2015)



Fuente: Superintendencia Financiera de Colombia; cálculos de los autores.

ROA: se observa un ROA promedio similar entre las entidades de los *clusters* uno y dos. Por otro lado, la mediana y el promedio de este indicador para las entidades del *cluster* tres son negativos.

Margen operativo: las medianas de este indicador son parecidas entre los *clusters* uno y dos, habiendo algunas entidades del *cluster* dos cuyo indicador es negativo. Todas las entidades pertenecientes al *cluster* tres poseen un indicador de margen operativo promedio negativo.

Razón de cobertura: la mediana del *cluster* tres es menor a la de los demás y sus valores son menores a 1 para cada una de las SFD pertenecientes a este. Lo anterior indica que para las entidades pertenecientes a este *cluster* los gastos de personal tienden a ser mayores que las comisiones que reciben.

Con el fin de encontrar evidencia empírica que indique si las diferencias que existen entre los *clusters* encontrados son significativas, se realizó la prueba de Kruskal-Wallis. Esta es una prueba no paramétrica para probar si n muestras pertenecen a una misma población, donde la hipótesis nula es que las n muestras tienen la misma distribución. Con un nivel de significancia del 1% se rechazó para todas las variables de desempeño analizadas. Lo anterior indica que existe una diferencia estadísticamente significativa para cada una de las variables de desempeño entre al menos dos de los tres *clusters*.

5. Conclusión

Usando técnicas estadísticas que minimizan la influencia de los criterios del investigador, se encontró que existen tres grandes *clusters* dentro del sector de las SFD asociados a tres modelos de negocio diferentes. Dicha clasificación resultó pertinente para explicar el desempeño financiero y la eficiencia de las entidades. Lo anterior contribuye al entendimiento de la dinámica de este sector y los riesgos a los que individualmente se enfrenta cada entidad.

Referencias

Farnè, M.; Vouldis, A. (2017). "Business Models of the Banks in the Euro Area", ECB Working Paper, núm. 2070, Banco Central Europeo.
 León, C. (2017). "Banks in Colombia: How Homogeneous Are They?", Borradores de Economía, núm. 1022, Banco de la República Colombia.
 Peña, D. (2013). *Análisis de datos multivariantes*, Madrid: McGraw-Hill España.