

Estimación de retiros netos para la administración de recursos de terceros

Juan Esteban Calderón Gómez



Contenido

| | | |
|------|--------------------------------|---|
| I. | Introducción | 3 |
| II. | Descripción de los datos | 3 |
| III. | Metodología..... | 4 |
| IV. | Resultados..... | 5 |
| V. | Conclusiones..... | 6 |
| VI. | Referencias | 7 |
| VII. | Anexos | 7 |

I. Introducción

El presente reporte tiene como finalidad presentar los resultados obtenidos dentro del desarrollo de un modelo de machine learning el cual busca establecer una técnica de analítica predictiva para series de tiempo, que permita calcular de forma diaria el nivel de retiro neto probable a 1 día de los diferentes portafolios que componen el Fondo Voluntario de Pensiones de Fiduciaria Davivienda.

En línea con lo anterior el estudio aborda tres frentes que apalancan la definición de alcanzables específicos, el primero de ellos se relaciona con la búsqueda de una implementación computacional que sea práctica, y que requiera la menor interacción posible por parte de un usuario. El segundo frente se enmarca en el uso de información disponible públicamente, aunque un desarrollo a futuro puede estar enmarcado en el uso de datos internos de la organización y el último pilar se enfoca en la forma de explicar el modelo como tal, ya que a pesar de que en general la audiencia es especializada en temas de inversiones, no necesariamente lo es en temas de analítica.

Lo mencionado anteriormente está incluido dentro de una necesidad que tiene un equipo de inversiones dentro de la entidad, y es que existe un riesgo inherente al administrar recursos del público, y es el hecho de que ocurra el evento en el que los requerimientos de los clientes para retirar su dinero sean mayores a los recursos que se tengan en cuentas de ahorro en el portafolio. Si bien en el escenario en el que la organización no cuente con los recursos disponibles (por estar invertidos en el mercado de valores) no implica que el cliente pierde su dinero, si ocasiona una serie de consecuencias reputacionales y legales para la misma.

En ese sentido, la necesidad de realizar el presente análisis radica en la generación de una herramienta que permita estimar el nivel probable de retiro de recursos por parte de los clientes, de tal manera que se pueda realizar una adecuada gestión de estos.

A lo largo del presente documento es posible apreciar que el modelo desarrollado permite tener una estimación adecuada de los niveles de retiros netos

por parte de los clientes de los portafolios, en promedio la diferencia entre el valor estimado y el observado se ubica en niveles de 0.03%, de igual manera fue posible ver que existe un indicio de que la industria que ofrece el tipo de productos mencionados previamente se comporta de manera homogénea

II. Descripción de los datos

Buscando responder a los planteamientos presentados previamente, este proyecto recurre en principio a información que se encuentra disponible de forma pública en el sitio web <https://www.datos.gov.co/dataset/Fondos-de-Pensiones-Voluntarias/gpzw-wmxd> en el cual la Superintendencia Financiera de Colombia publica los datos que las entidades que administran fondos voluntarios de pensiones tienen la obligación de reportarle a dicho ente, entre estos datos se encuentra el valor de portafolio, valor de unidad para realizar operaciones de ingreso y salida de clientes y valores de retiros de recursos de los diferentes fondos.

El procesamiento central de los datos se puede diferenciar en 4 partes o etapas las cuales se describen a continuación:

- i. **Historia:** Dado que se cuenta con una limitante computacional, desde la fuente mencionada previamente, únicamente es posible descargar hasta un máximo de 5 años.
- ii. **Totales:** existe un grupo de registros que contiene la agrupación de los demás para cada día. por ejemplo, si una empresa en particular tiene los fondos f_1 y f_2 existe un F tal que $f_1 \subset F$ y $f_2 \subset F$. Por ello todos los F son excluidos ya que el interés del estudio recae sobre los fondos en particular y no sobre el total.
- iii. **Nulos:** aquellas entidades cuyo número de faltantes sea mayor al 1% en un periodo de tiempo diario de la venta de tiempo cargada (5 años) son desestimadas del análisis.
- iv. **Incompletos:** de la masa de fondos resultantes, todos aquellos que no se encuentren dentro de la totalidad de días de la ventana de tiempo, con un margen de tolerancia sobre el número máximo de

ocurrencias de la compañía, son igualmente excluidos.

Los pasos previos configuran el proceso general que se realizó para el alistamiento de la información previo a la construcción del modelo.

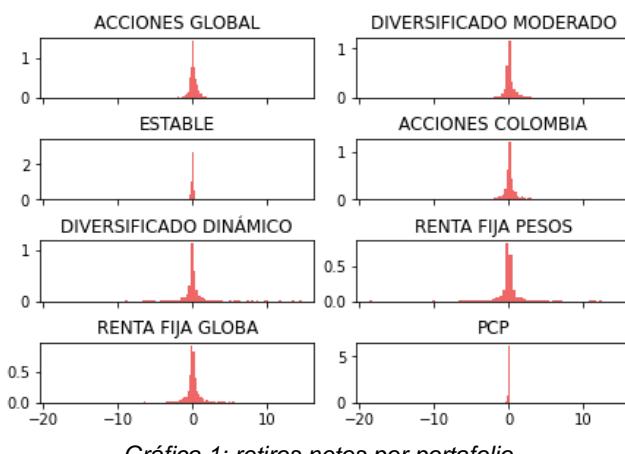
III. Metodología

Tal como se mencionó previamente, el propósito central del presente documento se encuentra basado en la construcción del análisis prospectivo del comportamiento de una serie de tiempo que representa el nivel de retiros de recursos que un determinado cliente o grupo de clientes puede llegar a realizar en un horizonte de tiempo de 1 día.

Es importante entonces, definir la forma en la cual se construye un retiro neto:

$$\text{Retiro Neto} = \text{Aportes} - \text{Retiros}$$

En ese sentido, en una primera instancia al observar la distribución existente de los datos, es posible ver unas distribuciones leptocúrticas con una inclinación a presentar valores alrededor del 0%, que como se observa en la gráfica 1 la cual contiene algunos portafolios específicos, en general las magnitudes altas (superior al 5% diario) son casos puntuales que se presentan en ciertos momentos del tiempo.



Gráfica 1: retiros netos por portafolio

En ese orden de ideas la selección de la población a estudiar y en el caso del presente estudio se encuentra centrado en los Fondos voluntarios de pensiones de fiduciaria Davivienda, los cuales contienen información de forma diaria continua para los últimos 5 años. La razón para tomar un periodo de tiempo completo radica en la necesidad de incluir la mayor cantidad de

ciclos económicos posibles, dado que en los momentos de crisis de los mercados financieros es donde se presentan los mayores retiros de los clientes, vale la pena tener una serie de tiempo que incluya la crisis del 2020 y alguno de los periodos de turbulencias importante como lo fue 2016 en el caso del mercado local y 2015 en dólar.

El modelo base usado para realizar la estimación del retiro probable por parte de los clientes es ARIMA (autorregresive integrated moving average) que de acuerdo a Gujarati & Porter (2010) A diferencia de los modelos de regresión, en los cuales la variable a explicar Y_t se explica por las k variables regresoras X_1, X_2, \dots, X_k , en los modelos de series de tiempo Box-Jenkins (ARIMA) Y_t se explica por valores pasados o rezagados de si misma. En ese sentido los modelos ARIMA trabajan con una única serie de tiempo, la cual es aquella variable de interés que se quiere trabajar. En la sección de Anexos se encuentra el detalle de la expresión matemática del modelo.

El proceso de aprendizaje se inicia con la partición de los datos en un grupo de 80% entrenamiento y 20% testeo, la razón para realizar esta selección se encuentra basada en el hecho de que computacionalmente resulta muy costoso en términos de tiempo el proceso de selección del modelo más optimo usando una partición de datos más grande. Sin embargo, el proceso de aprendizaje se realiza con ventanas deslizantes, que van incrementando por cada día transcurrido de la base de testeo.

Una vez se ha realizado la partición de los datos disponibles, el algoritmo aplicado para determinar el mejor modelo en un momento del tiempo parte del grupo de entrenamiento el cual se puede denotar por el conjunto de días $\{t_0; t_n\}$ sobre el cual se realizaría una primera calibración, a partir del dato observado t_{n+1} hasta el dato más reciente disponible se valida la estimación realizada, contrastando si el modelo realiza una proyección adecuada o no, luego cada dato de la muestra de testeo es incluido en la muestra de entrenamiento para así repetir nuevamente el proceso hasta llegar al dato más reciente.

Dado el objetivo planteado inicialmente para el proyecto que en términos generales busca estimar los retiros probables por parte de los clientes de la entidad por tanto la medida de desempeño se calcula como un score que conjuga tres frentes, por un lado un criterio numérico que pesa el 20% dentro del Score y es el RMSE, por el otro lado una variable que trata de validar

el número de estimaciones con signo positivo cuando el valor observado tiene signo negativo (60%) y un 20% restante proveniente de un rango de desviación bajo una perspectiva de negocio:

- **RMSE:** con una ponderación del 20% busca construir un ranking de la dispersión de la estimación entre los diferentes modelos, bajo la siguiente fórmula:

$$RMSE = \sqrt{\frac{(\hat{y}_t - y_t)^2}{n}}$$

Donde, \hat{y}_t es el dato estimado por el modelo
 y_t es el dato real observado
 n es el número de observaciones

- **Falsos Negativos:** dada la necesidad de construir un modelo que estime los retiros de los clientes, es un caso de mayor implicación tener una estimación que refleje un aporte de clientes cuando en realidad se presentó un retiro, es decir que la matriz de confusión se hace necesaria por cuanto permite estimar el número de eventos que presentan la situación ya descrita, por ello un 60% del indicador de desempeño, se encuentra reflejado en el ranking que ocupa el modelo como consecuencia de evaluar el porcentaje de falsos negativos dentro del total de estimaciones, según lo presentado en la tabla 1.

| | | Observación | |
|------------|--------|----------------|----------------|
| | | Retiro | Aporte |
| Estimación | Retiro | True Positive | False Positive |
| | Aporte | False Negative | True Negative |

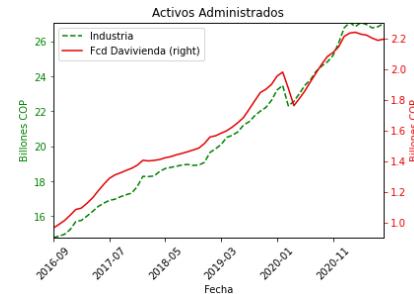
Tabla 1: zonas de estimación

- **Rango de estimación:** a través de un criterio de negocio es posible ver que cuando hay una diferencia en la estimación, superior al 2.5% hay un efecto significativo en la gestión de los portafolios, por tanto, el 20% restante en la medida de desempeño, está enmarcado por el ranking que ocupa el modelo en relación con el número de veces que la estimación estuvo en un rango por fuera de +2.5% o -2.5% frente al dato real de retiro.

IV. Resultados

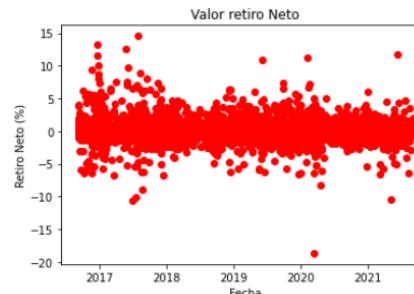
Los fondos voluntarios de pensiones se ofrecen por diferentes entidades dentro del sector financiero sin embargo es posible observar en la gráfica 2 que los movimientos de los activos administrados por la industria son similares a aquellos que ha presentado históricamente el fondo voluntario de pensiones de Fiduciaria Davivienda, con lo cual es posible evidenciar inicialmente que podría verse que se trata de un grupo de entidades que ofrecen un producto que se comporta de forma homogénea.

De acuerdo con lo planteado en la definición del problema, la variable de interés está centrada en los retiros por parte de los clientes de los diferentes fondos administrados, que de acuerdo con lo definido previamente un retiro neto se define como la suma de todos los aportes menos todos los retiros realizados por parte de los clientes.



Gráfica 2: Evolución Activos administrados

En el gráfico 3 es posible evidenciar el comportamiento de los retiros netos para todos los fondos, que a lo largo del tiempo mantienen valores cercanos a cero con eventos esporádicos que se caracterizan por tener valores sustancialmente mas altos de lo que normalmente se observa.



Gráfica 3: retiros históricos

En línea con lo anterior es pertinente pensar que si cuando un cliente realiza un retiro de alguno de los

Estimación de retiros netos para la administración de recursos de terceros

fondos administrados, lo hace porque ha observado rendimientos negativos en su inversión, sin embargo, en la tabla 2 es posible ver que, para la mayoría de los fondos administrados por Fiduciaria Davivienda que cuentan con todas las observaciones en el periodo de estudio, no pareciera existir una relación lineal entre el porcentaje de retiro neto y el porcentaje de retorno del fondo diario.

A raíz de lo anterior, hay un posible indicio de que no es conveniente usar el retorno del portafolio para estimar el valor probable de retiro de recursos, es por ello que el presente proyecto opta por usar un modelo que dependa únicamente de la serie misma a estimar y puntualmente a través de un modelo ARIMA de acuerdo a lo mencionado previamente.

| Portafolio | Coefficiente de Correlación |
|------------------------|-----------------------------|
| RENTA FIJA PESOS | 42% |
| ESTABLE | 21% |
| PCP | 5% |
| DIVERSIFICADO DINÁMICO | 3% |
| DIVERSIFICADO MODERADO | 4% |
| ACCIONES GLOBAL | 2% |
| ACCIONES COLOMBIA | -3% |
| RENTA FIJA PESOS | -3% |

Tabla 2 : Coeficiente de correlación de Pearson entre retorno diario y valor de retiro

En ese sentido, el proceso de estimación de retiros netos consistió en la generación de un grupo de modelos factibles que permitiesen estimar el retiro neto probable a 1 día, a cada uno de esos modelos, se calcula el indicador de desempeño expuesto previamente, en la tabla 3 se listan los modelos seleccionados por cada portafolio, donde se observa que en general el modelo mejor ranqueado es el 1, este se caracteriza por tener un componente autorregresivo de un nivel, diferenciación para reducir estacionariedad y promedio móvil también a un nivel.

| Portafolio | Modelo | Ranking | | | Score |
|------------------------|--------|---------|--------------------|---------|-------|
| | | RMSE | % Falsos Negativos | % Rango | |
| ACCIONES COLOMBIA | 3 | 2 | 1 | 2 | 0.096 |
| ACCIONES GLOBAL | 1 | 3 | 1 | 1 | 0.072 |
| DIVERSIFICADO DINÁMICO | 2 | 2 | 1 | 2 | 0.096 |
| DIVERSIFICADO MODERADO | 1 | 1 | 1 | 1 | 0.024 |
| ESTABLE | 1 | 1 | 2 | 1 | 0.048 |
| PCP | 1 | 2 | 1 | 1 | 0.048 |
| RENTA FIJA GLOBA | 1 | 1 | 1 | 1 | 0.024 |
| RENTA FIJA PESOS | 3 | 1 | 1 | 3 | 0.072 |

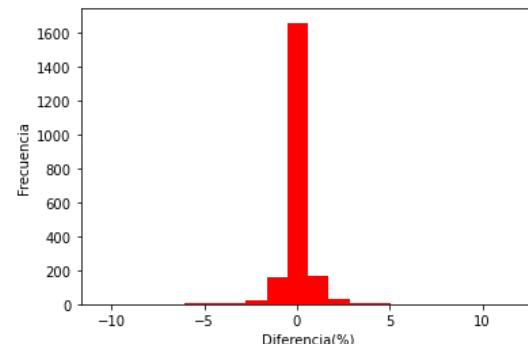
Tabla 3: selección del mejor modelo

Los modelos 3 y 2, guardan los mismos parámetros frente al modelo 1, sin embargo, se diferencian en el

número de rezagos seleccionados, en el caso del modelo 3 se caracteriza por tener 3 rezagos y en el caso del modelo 2 por tener 2 rezagos.

El score presentado en la tabla 3, se calcula de acuerdo con la definición presentada previamente sobre la medida de desempeño, seleccionando así aquel modelo que presentara el menor score, es decir el mejor ranking posible entre las 3 estimaciones, en la sección de anexos es posible ver estos resultados para todos los modelos.

En línea con lo presentado previamente al analizar los niveles porcentuales de diferencias entre las estimaciones y el valor real observado de los modelos seleccionados por cada portafolio, es posible ver que se han presentado niveles de hasta 10%, contando con una media de 0.03% y un rango intercuartil que va entre el -0.0014% y el 0.002%. En línea con esto en el gráfico 4 es posible apreciar la distribución de las diferencias para todos los portafolios en conjunto.



Gráfica 4: Distribución diferencias entre estimación y observado

V. Conclusiones

Dentro del análisis de los movimientos de los activos administrados en fiduciaria Davivienda para los fondos voluntarios de pensiones es posible evidenciar que existe cierta homogeneidad frente a los movimientos observados en las entidades financieras que ofrecen productos de características similares.

Adicionalmente los procesos de retiros de clientes de los fondos voluntarios de pensiones tienen un comportamiento que resulta especial, en cuanto a que contrario a lo que resulta intuitivo, existe indicio de que no presentan una correlación inversa con los retornos

obtenidos por los portafolios, es decir que en el momento en el que se presentan retiros de clientes no se presenta necesariamente un retorno, las evidencias muestran que efectivamente en un periodo de largo plazo no hay una correlación negativa entre estas variables, sin embargo en un futuro desarrollo al presente modelo puede incluir el retorno del portafolio como una variable explicativa.

En línea con lo anterior al validar el comportamiento de la variable de retiro neto para todos los fondos administrados, fue posible observar que hay una clara concentración alrededor del cero con ciertos momentos del tiempo donde se presentan retiros netos significativamente negativos, pero también positivos.

Finalmente la medida de desempeño permite evidenciar que los resultados obtenidos en general presentan un nivel de ajuste aceptable, en el caso del porcentaje de fallas para falsos negativos (estimado positivo frente a observado negativo) presenta niveles entre el 15% y el 22% para los diferentes modelos usados lo largo de la base de testeo la cual representa un año de estimaciones, por su parte el nivel de días donde la diferencia entre el valor estimado y el valor observado es mayor a 2.5% se encuentra en un rango razonable entre el 1.5% y el 13% de las veces, sin embargo para los modelos seleccionados está alrededor del 1.5%.

VI. Referencias

- Box, G., Jenkins, G., Reinsel, G., & Ljung, G. (2016). *Time Series Analysis*. United States: 2016.
- Gujarati, D., & Porter, D. (2010). *Econometría*. México D.F.: The McGraw-Hill.
- Montenegro García, Á. (2011). *Análisis de Series de Tiempo*. Bogotá: Pontificia Universidad Javeriana.
- Tsay, R. (2010). *Analysis of Financial Time Series*. Estados Unidos: Wiley.

VII. Anexos

Indicador de desempeño por cada modelo

A continuación, se presenta el resumen del resultado para cada modelo y portafolio seleccionado, tomando como referencia el score del indicador de desempeño el cual como se mencionó previamente es una forma de agrupar las tres perspectivas establecidas en la definición del indicador de desempeño.

| Portafolio | Modelo | RMSE | Ranking | | Score |
|------------------------|--------|------|--------------------|---------|-------|
| | | | % Falsos Negativos | % Rango | |
| ACCIONES COLOMBIA | 3 | 2 | 1 | 2 | 0.10 |
| ACCIONES COLOMBIA | 1 | 1 | 2 | 3 | 0.14 |
| ACCIONES COLOMBIA | 2 | 3 | 3 | 1 | 0.22 |
| ACCIONES GLOBAL | 1 | 3 | 1 | 1 | 0.07 |
| ACCIONES GLOBAL | 2 | 2 | 2 | 2 | 0.19 |
| ACCIONES GLOBAL | 3 | 1 | 3 | 3 | 0.22 |
| DIVERSIFICADO DINÁMICO | 2 | 2 | 1 | 2 | 0.10 |
| DIVERSIFICADO DINÁMICO | 3 | 1 | 2 | 3 | 0.14 |
| DIVERSIFICADO DINÁMICO | 1 | 3 | 3 | 1 | 0.22 |
| DIVERSIFICADO MODERADO | 1 | 1 | 1 | 1 | 0.02 |
| DIVERSIFICADO MODERADO | 2 | 2 | 2 | 2 | 0.19 |
| DIVERSIFICADO MODERADO | 3 | 3 | 3 | 3 | 0.65 |
| ESTABLE | 1 | 1 | 2 | 1 | 0.05 |
| ESTABLE | 3 | 3 | 1 | 3 | 0.22 |
| ESTABLE | 2 | 2 | 3 | 2 | 0.29 |
| PCP | 1 | 2 | 1 | 1 | 0.05 |
| PCP | 3 | 1 | 3 | 3 | 0.22 |
| PCP | 2 | 3 | 2 | 2 | 0.29 |
| RENTA FIJA GLOBA | 1 | 1 | 1 | 1 | 0.02 |
| RENTA FIJA GLOBA | 2 | 2 | 3 | 2 | 0.29 |
| RENTA FIJA GLOBA | 3 | 3 | 2 | 3 | 0.43 |
| RENTA FIJA PESOS | 3 | 1 | 1 | 3 | 0.07 |
| RENTA FIJA PESOS | 1 | 2 | 2 | 1 | 0.10 |
| RENTA FIJA PESOS | 2 | 3 | 3 | 2 | 0.43 |

Formulación Modelos ARIMA

Sea Y_t una variable de interés que se busca explicar a través de la generación de rezagos que pueden ser entendidos como el uso de observaciones pasadas para la estimación de la variable a explicar. En ese sentido se puede formular un modelo AR de orden 2 bajo la siguiente formulación:

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + u_t$$

Donde, Y_t variable a explicar en el momento t
 α coeficiente a estimar
 u_t ruido blanco

De igual manera es posible para Y_t determinar un modelo explicativo a través de promedios móviles, el cual puede ser definido como sigue a continuación para un modelo de orden 2:

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2}$$

Donde, Y_t variable a explicar en el momento t

α coeficiente a estimar
 u_t ruido blanco

En ese sentido un modelo ARIMA se configura con la construcción de una mezcla de los modelos presentados anteriormente, sin embargo, se adiciona un proceso de integración de la serie de tal forma que se cuente con un proceso estacionario en el tiempo.

$$Y_t = -(\Delta^d Y_t - Y_{t-1}) + \sum_{i=1}^p \alpha_i * Y_{t-i} - \sum_{i=1}^q \beta_i * u_{t-i}$$

Donde, Δ^d es la integración de la serie Y_t

Test de Estacionariedad

De acuerdo con Box y otros (2016) una forma de determinar si una serie de tiempo presenta estacionariedad es a través de la determinación de la existencia o no de una raíz unitaria en la serie. En ese sentido una herramienta es el test de Dickey – Fuller, quienes propusieron una evaluación basada en el estimador de mínimos cuadrados para un proceso autorregresivo, en ese sentido en términos prácticos las hipótesis pueden ser establecidas como siguen:

H_0 : Hay raiz unitaria (No estacionaria)

H_1 : No hay raiz unitaria (estacionaria)

Con un P-valor menor a 0.05 se rechaza H_0 y por tanto es necesario diferenciar la serie para que sea no estacionaria.

Por tanto, a continuación, se presenta el resultado del test Dickey – Fuller, donde es posible concluir

que las series de todos los portafolios requieren ser diferenciados para convertirse en series estacionarias:

| Portafolio | Test F | P-Value |
|------------------------|--------|----------|
| ACCIONES COLOMBIA | -8.51 | 1.17E-13 |
| ACCIONES GLOBAL | -6.83 | 1.90E-09 |
| DIVERSIFICADO DINÁMICO | -6.69 | 4.12E-09 |
| DIVERSIFICADO MODERADO | -6.92 | 1.15E-09 |
| ESTABLE | -9.58 | 2.20E-16 |
| PCP | -4.81 | 5.22E-05 |
| RENTA FIJA GLOBA | -11.34 | 1.04E-20 |
| RENTA FIJA PESOS | -9.27 | 1.32E-15 |

Estimación modelo seleccionado por portafolio

A continuación, se presenta el nivel de aciertos que ha presentado el mejor modelo seleccionado para cada portafolio, adicionalmente se presenta la suma de los datos estimados como True positive y True negativa que representarían el mejor escenario ya que el signo de la estimación es el mismo del dato observado

| Portafolio | False Negative | False Positive | True Positive | True Negative | % Efectividad |
|------------------------|----------------|----------------|---------------|---------------|---------------|
| ACCIONES GLOBAL | 15% | 19% | 15% | 52% | 67% |
| ESTABLE | 19% | 20% | 45% | 15% | 61% |
| DIVERSIFICADO DINÁMICO | 21% | 22% | 24% | 33% | 57% |
| RENTA FIJA GLOBA | 22% | 24% | 18% | 36% | 54% |
| DIVERSIFICADO MODERADO | 22% | 20% | 29% | 29% | 58% |
| ACCIONES COLOMBIA | 21% | 25% | 15% | 39% | 54% |
| RENTA FIJA PESOS | 19% | 20% | 31% | 30% | 61% |
| PCP | 17% | 38% | 15% | 31% | 46% |