

Análisis de mercados financieros mediante PCA, ventanas móviles y métricas de red

Máximo Caprari

Introducción

El análisis busca caracterizar el comportamiento conjunto de activos financieros a través de retornos móviles, reducción de dimensionalidad y construcción de métricas de red. El objetivo es identificar patrones colectivos y detectar periodos de crisis en el mercado.

Tickers y Descarga de Datos

En esta sección se describen los tickers de los activos financieros utilizados en el análisis de componentes principales (PCA) y el método de descarga de datos utilizando la librería `yfinance`.

Los datos financieros de estos tickers se descargan utilizando la librería `yfinance`, que permite acceder a datos históricos de precios de activos. A continuación se muestra un ejemplo de cómo descargar los datos:

```
import yfinance as yf

# Lista de tickers
TICKERS = ['ARGT', 'GGAL', 'BMA', 'BBAR', '^MERV', 'PAM', 'YPF',
            'ILF', 'TGS', 'CRESY', 'EWZ', 'TEO', 'EEM', '^GSPC',
            'IRS', 'MELI', 'CL=F', '^TNX',
            'ZS=F', 'ZC=F', 'ZW=F', 'GC=F', 'ARS=X', 'USDBRL=X']

# Descargar datos de precios para los tickers
data = yf.download(TICKERS, start='2020-01-01', end='2023-01-01')
```

Este código descarga los precios históricos de los activos listados desde el 1 de enero de 2020 hasta el 1 de enero de 2023.

Table 1: Tickers utilizados en el análisis

Ticker	Descripción
ARGT	Global X MSCI Argentina ETF
GGAL	Grupo Financiero Galicia (Argentina)
BMA	Banco Macro S.A. (Argentina)
BBAR	Banco BBVA Argentina S.A.
ÍMERV	Índice Merval (Argentina)
PAM	Pampa Energía S.A. (Argentina)
YPF	YPF S.A. (Argentina)
ILF	iShares Latin America 40 ETF
TGS	Transportadora de Gas del Sur S.A. (Argentina)
CRESY	Cresud S.A.C.I.F. y A. (Argentina)
EWZ	iShares MSCI Brazil ETF
TEO	Telecom Argentina S.A.
EEM	iShares MSCI Emerging Markets ETF
ĜSPC	S&P 500 Index
IRS	IRSA Inversiones y Representaciones S.A.
MELI	MercadoLibre, Inc.
CL=F	Crudo Ligero (WTI)
ĤTNX	Treasury Yield 10 Years
ZS=F	Soya (Futuros)
ZC=F	Maíz (Futuros)
ZW=F	Trigo (Futuros)
GC=F	Oro (Futuros)
ARS=X	Peso Argentino/Dólar Estadounidense
USDBRL=X	Real Brasileño/Dólar Estadounidense

Construcción de retornos

Dada una serie de precios P_t para un activo, se define el retorno logarítmico como

$$r_t = \ln \left(\frac{P_t}{P_{t-1}} \right).$$

Este retorno es aditivo en el tiempo y resulta adecuado para el análisis estadístico.

En código, los retornos se calculan de forma vectorizada:

```
returns = np.log(prices / prices.shift(1))
```

Ventanas móviles

Se emplean ventanas de tamaño $W = 60$ días con un paso de $S = 5$ días. Para cada ventana w se calculan métricas como el retorno medio anualizado

$$\mu_w = \frac{252}{W} \sum_{t \in w} r_t,$$

la volatilidad anualizada

$$\sigma_w = \sqrt{252} \cdot \text{std}(r_t, t \in w),$$

y el ratio de Sharpe

$$S_w = \frac{\mu_w}{\sigma_w}.$$

Métricas de red

Con la matriz de correlaciones ρ_{ij} entre activos se define una distancia

$$d_{ij} = \sqrt{2(1 - \rho_{ij})}.$$

A partir de esta matriz de distancias se construye un grafo ponderado y se extraen métricas globales como:

- Coeficiente de clustering promedio
- Peso total del árbol mínimo de expansión
- Longitud promedio de caminos

PCA y reducción de dimensionalidad

El análisis de componentes principales permite transformar el espacio de N activos en un número reducido de factores k que capturan la mayor parte de la varianza. Sea X la matriz de retornos estandarizados. El PCA descompone

$$X = U\Sigma V^\top,$$

donde las columnas de V son los vectores propios que representan los componentes principales.

En código:

```
from sklearn.decomposition import PCA

pca = PCA(n_components=2)
X_pca = pca.fit_transform(returns.fillna(0))
```

Clusterización

A partir de los primeros componentes se aplica KMeans con $k = 3$ para identificar grupos de activos con comportamiento similar:

```
from sklearn.cluster import KMeans

kmeans = KMeans(n_clusters=3, random_state=42)
clusters = kmeans.fit_predict(X_pca)
```

Proyección temporal en componentes principales

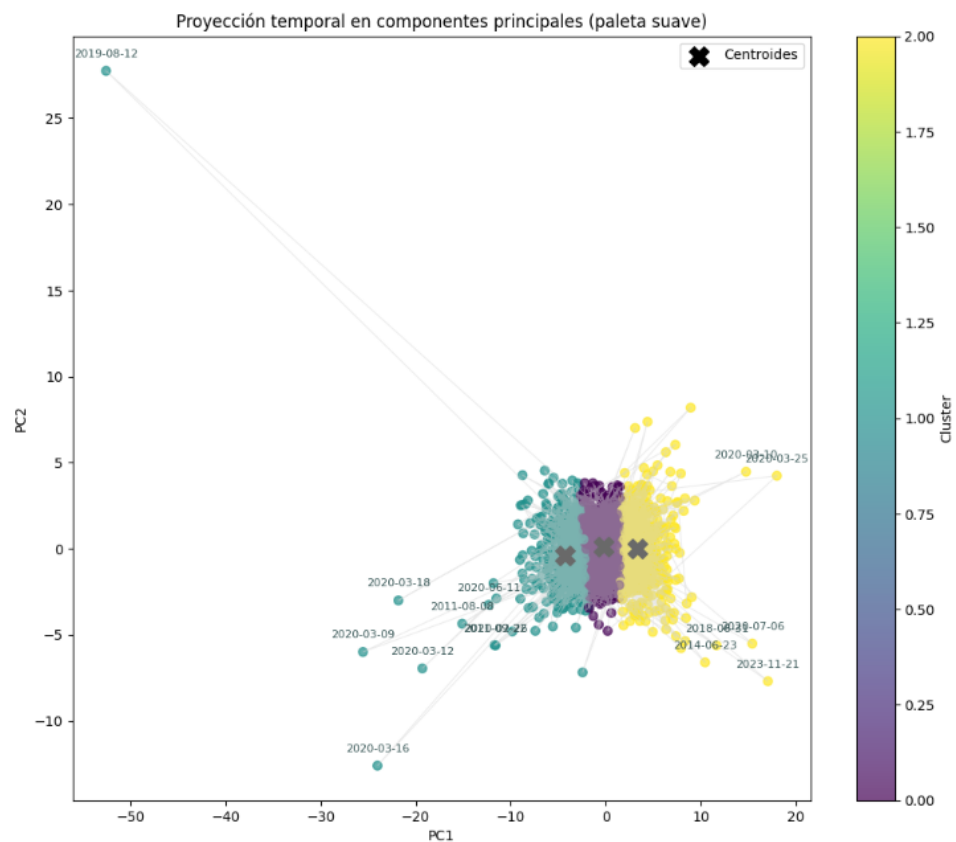


Figure 1: Proyección de ventanas móviles en el espacio definido por los dos primeros componentes principales (PC1 y PC2). Los colores representan los clústeres y las cruces negras señalan los centroides de cada grupo.

En el gráfico se observa cómo cada punto corresponde a una **ventana de 60 días**, desplazada cada 5 días en el tiempo. Estas ventanas fueron proyectadas sobre los dos primeros componentes principales, que explican una fracción relevante de la varianza total.

Los clústeres muestran tres regímenes diferenciados:

- Un clúster compacto cercano al origen, que refleja periodos de estabilidad de mercado, con correlaciones moderadas y variabilidad baja.
- Un clúster desplazado hacia la derecha (PC1 positivo), que agrupa ventanas donde el mercado mostró dinámicas expansivas o de recuperación.
- Un clúster disperso hacia valores extremos de PC1 negativos, donde aparecen eventos puntuales de estrés, tales como caídas abruptas o shocks externos.

Es interesante notar que fechas particulares como marzo de 2020 o agosto de 2019 aparecen alejadas del centro, indicando **ventanas de crisis**. La distancia de estos puntos respecto al centro refleja el grado de anomalía respecto a la dinámica típica de los activos.

La posición de los centroides sintetiza el comportamiento promedio de cada régimen. Matemáticamente, el centroide c_k del clúster k se define como:

$$c_k = \frac{1}{n_k} \sum_{i \in C_k} x_i,$$

donde C_k es el conjunto de ventanas asignadas al clúster k y x_i la proyección PCA de la ventana i .

En términos prácticos, esta representación permite mapear **transiciones de régimen de mercado** y localizar temporalmente los periodos que se apartan de la dinámica habitual.

Biplot de los componentes principales

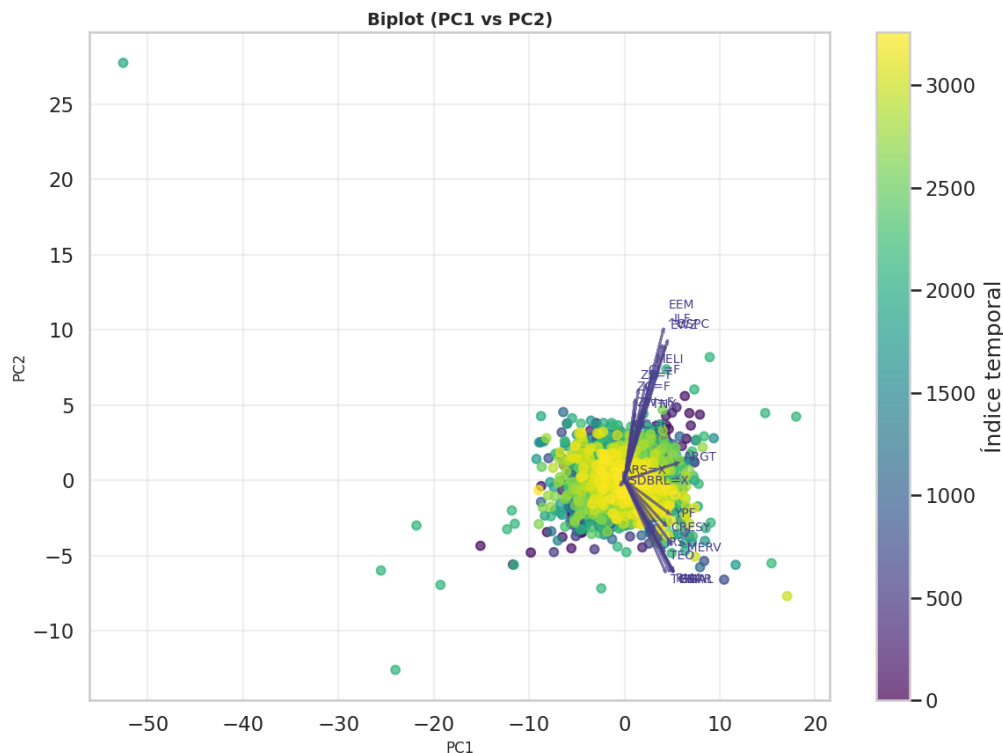


Figure 2: Biplot de las ventanas móviles en el espacio definido por los dos primeros componentes principales. Los colores indican el índice temporal de cada ventana, mientras que los vectores muestran la contribución de las variables originales.

Este biplot combina dos tipos de información:

- La proyección de las **ventanas temporales** sobre el plano ($PC1, PC2$), donde cada punto está coloreado según el índice temporal. Esto permite ver la evolución dinámica del sistema a lo largo del tiempo.
- La representación de las **cargas de las variables originales** mediante vectores. La dirección y magnitud de cada vector indica el aporte de cada activo o variable al espacio de los componentes principales.

En términos matemáticos, si X es la matriz de datos estandarizados y V la matriz de autovectores, la proyección de los datos sobre los primeros dos componentes se obtiene como:

$$Z = X \cdot V_{(2)},$$

donde $V_{(2)}$ contiene las dos primeras columnas de V .

El ángulo entre vectores refleja el grado de correlación entre las variables originales. Por ejemplo, activos financieros que apuntan en la misma dirección tienden a estar correlacionados, mientras que aquellos que forman ángulos cercanos a 180° presentan correlaciones negativas.

El patrón temporal muestra que la mayor parte de las ventanas se concentran en torno al centro, pero algunas se proyectan de forma extrema sobre PC1 negativo, lo que coincide con periodos de crisis.

El biplot es particularmente útil porque no solo permite visualizar las **relaciones entre las observaciones**, sino también la **estructura de dependencia entre las variables**, proporcionando un mapa conjunto de dinámicas de mercado y factores latentes.

[colback=gray!5, colframe=black!40, title=Lecturas clave del biplot]

- **Estructura de factores:** Bloque global (EEM, \hat{G} SPC, EWZ) define un eje común de riesgo global; commodities (CL=F, ZS=F, ZC=F) muestran un ciclo propio aunque correlacionado; activos argentinos (GGAL, BMA, YPF, PAM, ARGTE, \hat{M} ERV) agrupan en un factor local.
- **Monedas y tasas:** \hat{T} NX se ubica en posición intermedia (a veces se mueve con el equity global, a veces en contra); ARS=X y USDBRL=X aportan poco a PC1/PC2 pero ayudan a interpretar cobertura frente a riesgo local.
- **Relaciones entre grupos:** Acciones argentinas forman un bloque separado del eje global, reflejando un factor soberano/político propio; materias primas aportan diversificación parcial frente al riesgo global.
- **Implicancias:** Con PC1 y PC2 se obtienen dos factores relevantes: *global* y *local-Argentina*; útiles para construir portfolios de factores, pensar coberturas y detectar cambios de régimen.

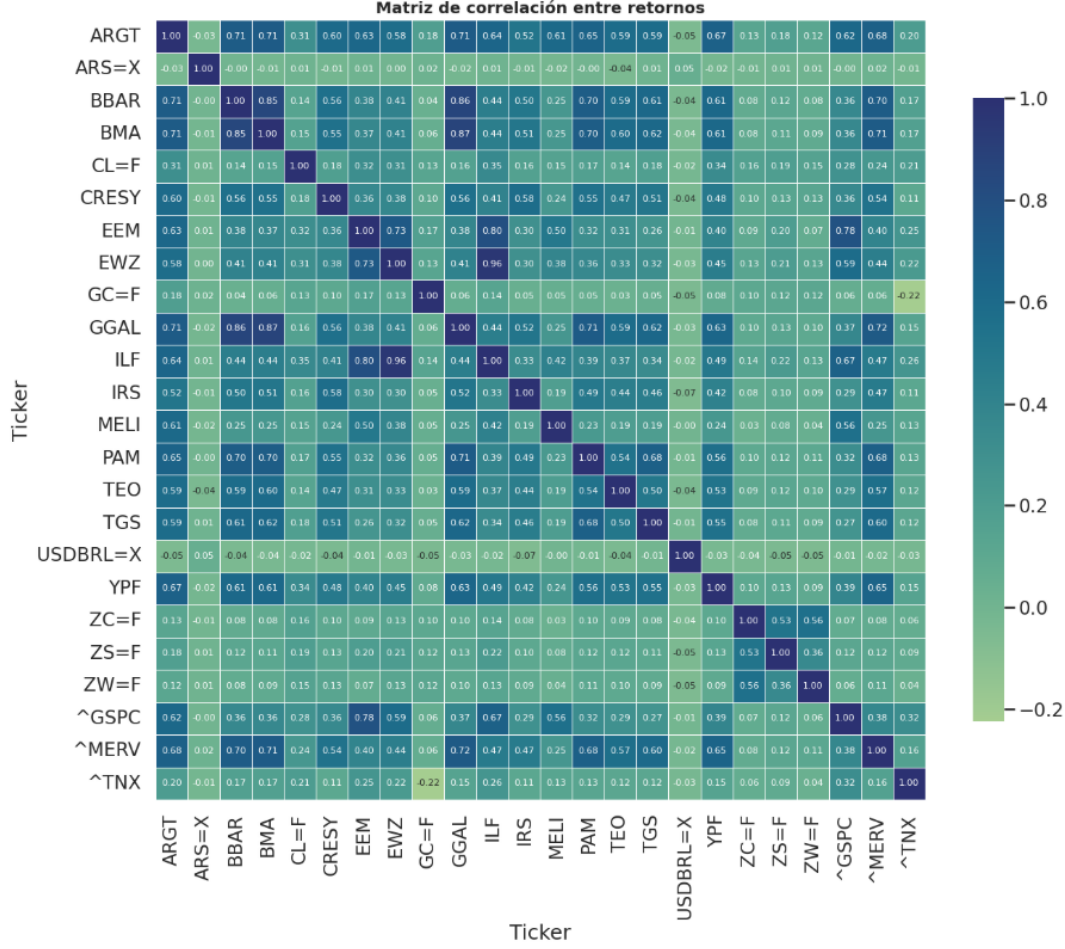


Figure 3: Matriz de correlación entre retornos de activos financieros seleccionados.

La Figura 3 muestra la matriz de correlaciones de Pearson entre los retornos de distintos activos, incluyendo acciones argentinas (GGAL, BMA, YPF), índices regionales (ARGT, ILF, EWZ), commodities (CL=F, ZS=F, ZC=F, GC=F), y referencias globales como el S&P500 (^GSPC) y el índice MSCI emergentes (EEM).

Se observa un bloque de alta correlación entre las acciones financieras argentinas (GGAL, BMA, BBAR), consistente con su exposición común al riesgo local. A su vez, estos activos muestran una correlación positiva significativa con el índice Merval (^MERV) y con el ETF ARGT, lo que refleja la naturaleza de estos últimos como proxies del mercado accionario argentino.

Por otro lado, índices globales como EEM y EWZ muestran correlaciones moderadas, capturando la sensibilidad compartida a shocks externos. En contraste, los commodities agrícolas (ZS=F, ZC=F, ZW=F) presentan correlaciones más bajas con el mercado accionario, lo que confirma su rol como activos diversificadores.

Formalmente, la correlación entre dos series r_i y r_j se define como:

$$\rho_{ij} = \frac{\text{Cov}(r_i, r_j)}{\sigma_{r_i} \sigma_{r_j}},$$

donde Cov denota la covarianza y σ la desviación estándar de cada serie. Valores cercanos a 1 reflejan movimientos conjuntos fuertes, mientras que valores cercanos a 0 sugieren independencia en los retornos.

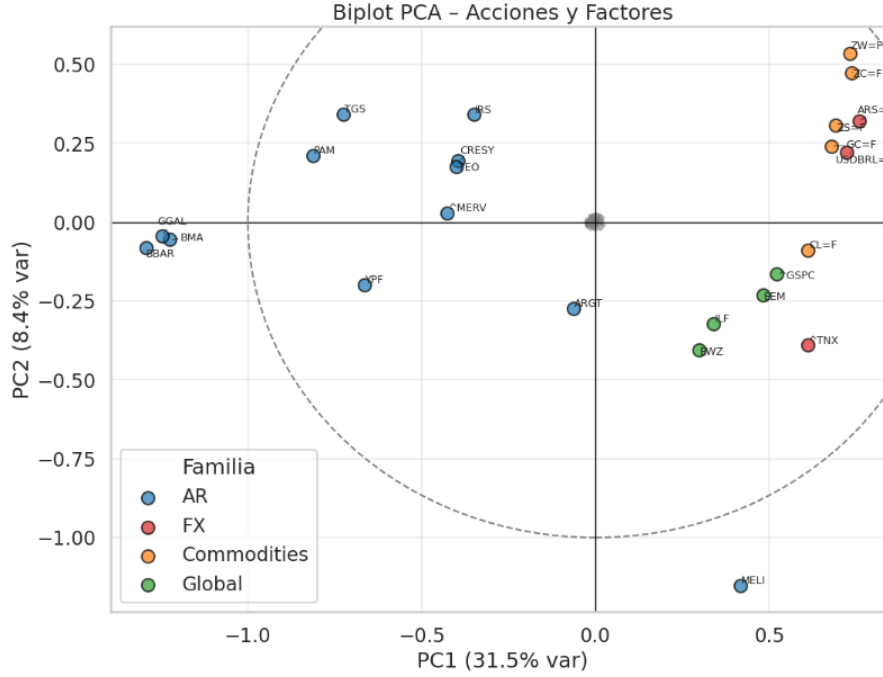


Figure 4: Biplot de PCA para acciones argentinas, factores globales, tipos de cambio y commodities.

La Figura 4 muestra el resultado del Análisis de Componentes Principales (PCA) aplicado a la matriz de retornos estandarizados. El primer componente principal (PC1) explica aproximadamente un 31.5% de la varianza total, mientras que el segundo (PC2) captura un 8.4%. El gráfico organiza los activos en función de sus cargas principales, diferenciados por familias: *acciones argentinas* (azul), *tipos de cambio* (naranja), *commodities* (verde) y *factores globales* (rojo).

Se observa que los activos argentinos (GGAL, BMA, BBAR, YPF) se agrupan en la zona izquierda del plano, revelando su dinámica común impulsada por el riesgo local. En contraste, las monedas (ARS=X, USDBRL=X) y los commodities (CL=F, ZW=F, ZS=F) se proyectan hacia la derecha, mostrando una diferenciación clara en la exposición a factores externos.

El PCA se fundamenta en la descomposición espectral de la matriz de covarianzas de los retornos:

$$\Sigma = V\Lambda V^T,$$

donde V son los vectores propios (direcciones principales) y Λ los valores propios (varianza explicada). La proyección de cada activo x sobre el k -ésimo componente se calcula como:

$$z_{ik} = x_i \cdot v_k,$$

lo que permite representar en dos dimensiones gran parte de la estructura multivariada del sistema financiero bajo estudio.

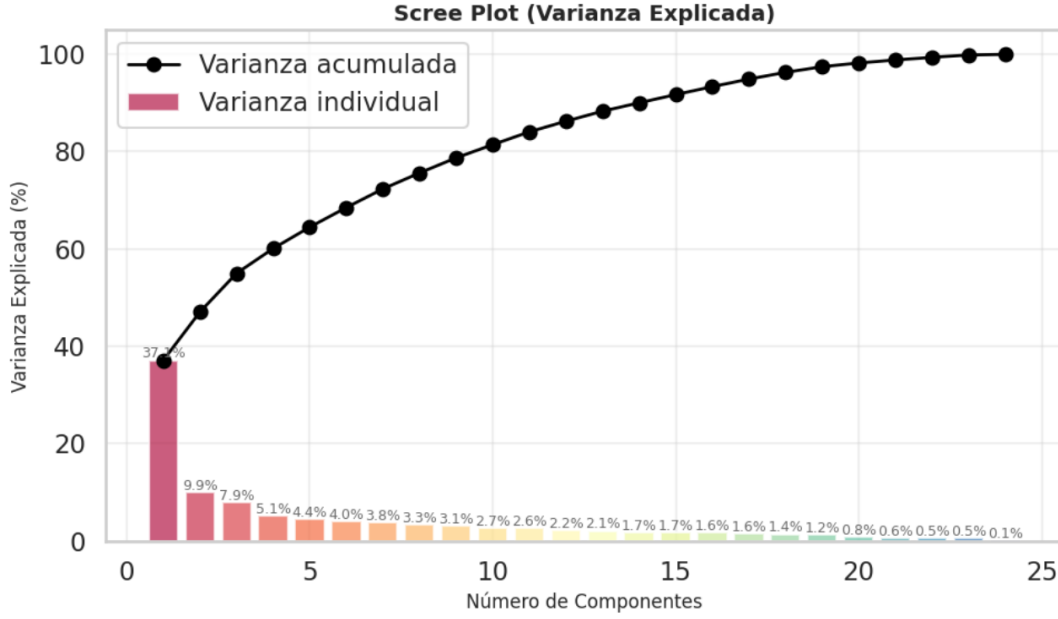


Figure 5: Scree Plot de varianza explicada por componentes principales.

La Figura 5 presenta el Scree Plot utilizado para evaluar la cantidad óptima de componentes principales en el análisis. La barra inicial muestra que el primer componente ($PC1$) explica aproximadamente un 37% de la varianza total. El segundo componente aporta cerca de un 10%, y a partir del tercero los aportes individuales descienden progresivamente por debajo del 10%.

El gráfico de línea negra representa la **varianza acumulada**. Se observa que con los primeros cinco componentes se captura más del 60% de la variabilidad, mientras que con alrededor de diez componentes se supera el 80%. A partir de allí, los incrementos son marginales, indicando un punto de rendimientos decrecientes.

En términos formales, si denotamos la varianza explicada del componente i como λ_i , entonces:

$$\text{Varianza explicada individual} = \frac{\lambda_i}{\sum_j \lambda_j}, \quad \text{Varianza acumulada}(k) = \frac{\sum_{i=1}^k \lambda_i}{\sum_j \lambda_j}$$

Este comportamiento sugiere que un modelo parsimonioso puede construirse seleccionando únicamente las primeras componentes, ya que concentran la mayor parte de la estructura informativa de los datos. El “codo” del gráfico se identifica entre la segunda y cuarta componente, lo cual apoya una reducción dimensional conservadora en ese rango.

Table 2: Cargas principales de los activos en PC1 y PC2

Ticker	PC1	PC2
ARGT	0.301	0.040
GGAL	0.280	-0.214
BMA	0.277	-0.215
BBAR	0.276	-0.214
ÊMERV	0.268	-0.149
PAM	0.257	-0.213
YPF	0.253	-0.077
ILF	0.247	0.329
TGS	0.234	-0.215
CRESY	0.233	-0.105
EWZ	0.231	0.316
TEO	0.227	-0.165
EEM	0.224	0.357
ÊSPC	0.214	0.318
IRS	0.208	-0.138
MELI	0.155	0.244
CL=F	0.118	0.223
ÊTNX	0.089	0.150
ZS=F	0.079	0.211
ZC=F	0.060	0.186
ZW=F	0.059	0.154
GC=F	0.044	0.108
ARS=X	-0.003	0.012
USDBRL=X	-0.016	-0.008

El **primer componente principal (PC1)** concentra principalmente las acciones argentinas (ARGT, GGAL, BMA, BBAR, ÊMERV, PAM, YPF, TGS). Sus coeficientes son elevados y positivos, lo cual indica que este factor resume de manera conjunta la dinámica local del mercado accionario argentino.

El **segundo componente principal (PC2)** está dominado por variables globales y de mercados emergentes, como EEM, ILF, EWZ y ÊSPC, junto con algunas materias primas (CL=F, ZS=F, ZC=F). Esto refleja un factor asociado a la exposición internacional y a commodities, distinto del factor puramente local.

En términos formales, cada componente puede escribirse como combinación lineal de los activos:

$$PC1 = \sum_i w_i^{(1)} \cdot X_i, \quad PC2 = \sum_i w_i^{(2)} \cdot X_i$$

donde $w_i^{(1)}$ y $w_i^{(2)}$ son las cargas listadas en la Tabla 2.

Conclusión:

- PC1 captura la variación conjunta del sistema financiero argentino.

- PC2 representa la influencia de factores globales (acciones internacionales y commodities).
- El tipo de cambio (ARS=X, USDBRL=X) no tiene peso relevante en estos dos primeros componentes, lo que sugiere que su dinámica se explica en dimensiones posteriores.

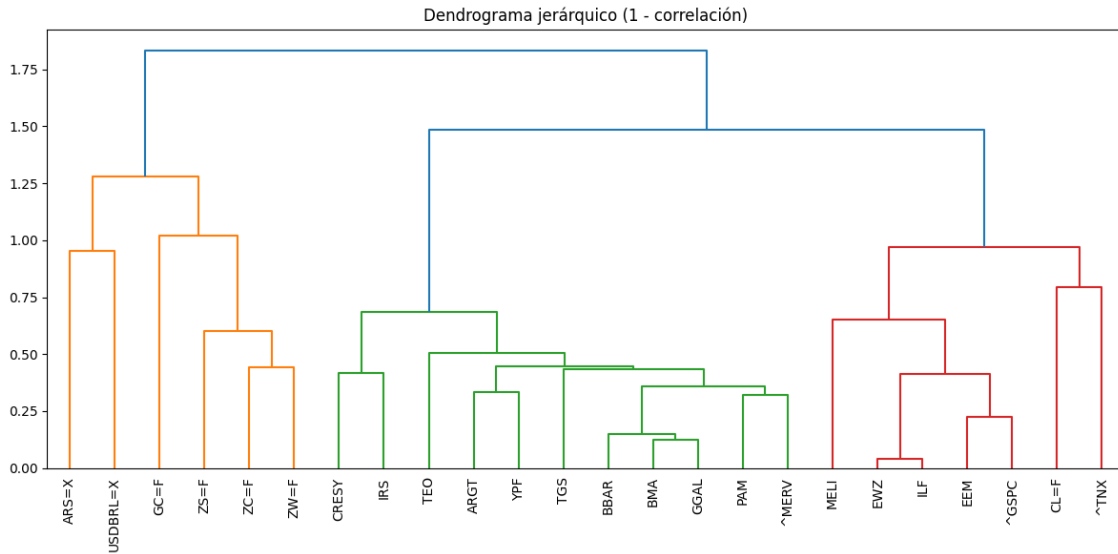


Figure 6: Dendrograma jerárquico de activos (1 - correlación).

Lecturas principales del dendrograma

Cluster 1 (izquierda, naranja): FX y commodities agro/metales. ARS=X y USDBRL=X se agrupan mostrando dinámica común de monedas emergentes. Se suman commodities agrícolas y metales (oro, soja, maíz, trigo). Comparten un factor ligado a shocks de riesgo global y flujos hacia emergentes.

Cluster 2 (centro, verde): Acciones argentinas / riesgo local. ADRs y acciones locales argentinas aparecen muy próximas junto con el índice $\hat{M}ERV$ y el ETF ARGTE. Comparten un fuerte factor de riesgo local. MELI queda algo más separado por tener ingresos más globales.

Cluster 3 (derecha, rojo): Global equities + petróleo + tasa US. Se observa un bloque compacto formado por ILF, EWZ, $\hat{G}SPC$, EEM. A este grupo se suman petróleo (CL=F) y la tasa larga de EEUU ($\hat{T}NX$), reflejando su cercanía con el riesgo global.

Conexión entre clusters. Los tres bloques son relativamente independientes: (1) FX + commodities, (2) Argentina, (3) global equities + petróleo y tasas. El FX se alinea más con commodities que con equities locales, lo que confirma factores diferenciados.