

# **Predicciones de volatilidad en sectores de la Bolsa Mexicana de Valores.**

Equipo 20:

César García Gutiérrez  
Juan Rosendo González Fera  
Emilio Linarez Meneses  
Set Jafet Renedo Ortega

9 de marzo de 2021

## Postwork 1.

### Identificación de un problema.

El poder predecir a partir de un modelo estadístico de una serie de tiempo se ha vuelto toda una rama de investigación en estadística pues permite hacer previsiones sobre los fenómenos que se estudian. Estos fenómenos van desde la temperatura diaria en una región, las personas que viajan en avión cada mes, hasta los precios históricos de las acciones de una empresa en el mercado financiero.

En este caso estamos interesados en predecir valores futuros de precios de acciones en distintas empresas. Dichas empresas fueron seleccionadas de acuerdo a los sectores existentes, los cuales son:

- Industrial.
- Materiales.
- Productos de consumo.
- Salud.
- Telecomunicaciones.
- Financieros.
- Productos de consumo no básicos.

La lista completa de sectores se puede consultar en [3].

Para lograr esto, primero es necesario saber de donde obtener la información y cómo bajarla para poder ajustar un modelo de serie de tiempo. Estos pasos serán respondidos en las siguientes secciones.

### Investigación.

Para entender este tema un poco más a fondo es necesario preguntarnos ¿Qué es un activo financiero?, ¿Cuáles son sus características?, ¿Cuáles son los factores que influyen?.

Un activo financiero es un instrumento financiero que otorga a su comprador el derecho a recibir ingresos futuros por parte del vendedor. Es decir, es un derecho sobre los activos reales del emisor y el efectivo que generen.

Al contrario que los activos tangibles (un coche o una casa por ejemplo), los activos financieros no suelen tener un valor físico. El comprador de un activo financiero posee un derecho (un activo) y el vendedor una obligación (un pasivo). Los activos financieros pueden ser emitidos por cualquier unidad económica (empresa, Gobierno, etc).

Un activo financiero obtiene su valor de ese derecho contractual. Gracias a estos instrumentos las entidades, que poseen deuda se pueden financiar y, a su vez, las personas que quieren invertir sus ahorros consiguen una rentabilidad invirtiendo en esa deuda.

### Las características de los activos financieros.

Los activos financieros tienen tres características fundamentales:

- Liquidez: capacidad de convertir el activo en dinero sin sufrir pérdidas.
- Rentabilidad: cuanto más interés aporta el activo mayor es su rentabilidad.
- Riesgo: probabilidad de que el emisor no cumpla sus compromisos. Cuanto mayor sea el riesgo, mayor será la rentabilidad.

Cada una de ellas puede variar según el tipo de activo financiero. Además, existe una fuerte relación entre rentabilidad, riesgo y liquidez. Según la magnitud de una afectará a la otras. Por ejemplo, un activo financiero menos líquido tendrá más riesgo y por tanto exigirá una mayor rentabilidad.

### Modelos de series de tiempo.

Dado que la investigación en series de tiempo de activos financieros es muy amplia, actualmente se cuentan con bastantes modelos para poder hacer predicciones, entre los cuales vale la pena destacar los famosos modelos *ARCH* y *GARCH* [2], estos modelos de series de tiempo nos ayudan a predecir la volatilidad futura de los precios de una empresa, no el valor futuro de los precios en sí. Sin embargo esta información es bastante útil pues nos da una idea de que tan susceptible en el futuro serán los precios de las acciones, ya que una volatilidad baja se traduce a una baja variación de los precios futuros, mientras que una alta volatilidad está asociada a una variación grande de los precios.

### Búsqueda de soluciones anteriores.

Dado que el análisis de series de tiempo es un tema bastante famoso en estadística y en economía se tiene demasiada literatura e investigaciones al respecto, solo por mencionar algunas fuentes se tiene que para los modelos *ARCH* y *GARCH* se pueden consultar [1] y [2].

## Postwork 2.

Una vez que nos hemos ubicado en el problema vale la pena describir los datos a usar y las preguntas a responder.

En este proyecto se desean obtener los datos de los precios diarios de las acciones de algunas empresas que cotizan en la Bolsa Mexicana de Valores y que pertenecen a los distintos sectores. Como se puede observar en la página de la BMV, cuentan con 10 sectores industriales [3] sin embargo se seleccionaron 7 sectores, los cuales son:

- Del sector industrial se seleccionó la empresa **Grupo Aeroportuario del Pacífico**.
- Del sector de materiales se seleccionó la empresa **Cemex**.
- Del sector de productos de consumo se seleccionó la empresa **Walmart**.
- Del sector de salud se seleccionó la empresa **Corporativo Fragua**.
- Del sector financiero se seleccionó la empresa **Grupo Financiero Banorte**.
- Del sector de telecomunicaciones se seleccionó la empresa **América Móvil**.
- Del sector de productos de consumo no básicos se seleccionó la empresa **Grupo Elektra**.

Una vez obtenidos los datos de valores de dichas empresas se planean responder las siguientes preguntas:

- ¿El valor de una empresa depende exclusivamente de ésta o existe alguna relación entre sectores?
- ¿Es posible hacer pronósticos de los valores de dichas empresas?
- Ya ajustados los modelos correspondientes, ¿cómo se selecciona el mejor modelo?
- A partir de los modelos realizados se pueden responder preguntas bastante interesantes como:

- ¿Cual empresa disminuirá su valor en el futuro?
- ¿Cual empresa aumentará su valor en el futuro?
- ¿Existen empresas de distintos sectores que se encuentren fuertemente relacionadas?

## Postwork 3.

Es de suma importancia conocer de donde recolectar los datos de interés, para esto se realizaron algunas búsquedas en páginas dedicadas a recolección de datos de valores de empresas de diferentes sectores, siendo el sitio de *Yahoo Finance* [4] el más adecuado para nuestros propósitos ya que cuenta con una API que nos permite descargar los valores históricos de distintas empresas.

Uno de los puntos a favor para usar esta API es que no se requiere un registro previo en ningún sitio y además el número de peticiones es adecuado para nuestros propósitos. Cabe mencionar también que es muy raro que la información acerca de los precios de las empresas presenten datos nulos o algún tipo de dato incorrecto.

## Postwork 4.

Ya seleccionadas las empresas veamos los datos solicitados a la API transformados en un dataframe. En este caso veamos la información del Grupo Aeroportuario del Pacífico:

1	df							
		Open	High	Low	Close	Volume	Dividends	Stock Splits
	Date							
	2006-02-27	14.581499	15.060471	14.566685	14.912335	18404600.0	0.0	0
	2006-02-28	14.813574	15.011088	14.319788	14.828388	9699400.0	0.0	0
	2006-03-01	14.828386	15.653008	14.828386	15.391301	3748300.0	0.0	0
	2006-03-02	15.544375	15.702386	15.430804	15.653008	3410700.0	0.0	0
	2006-03-03	15.554250	15.707324	15.302420	15.391301	2939400.0	0.0	0
	...	...	...	...	...	...	...	...
	2021-03-01	214.580002	215.919998	212.229996	213.960007	1132413.0	0.0	0
	2021-03-02	213.169998	220.610001	210.800003	219.789993	1659351.0	0.0	0
	2021-03-03	220.000000	226.919998	219.979996	226.470001	1286717.0	0.0	0
	2021-03-04	227.350006	231.169998	222.520004	223.399994	999319.0	0.0	0
	2021-03-05	225.649994	229.210007	223.919998	228.889999	964457.0	0.0	0

3771 rows × 7 columns

Figura 1: Dataframe obtenido a partir de la API de *Yahoo Finance*. En este caso los datos corresponden a la empresa Grupo Aeroportuario del Pacífico.

De la Figura 1 podemos apreciar algunas cosas:

- El índice corresponde a una fecha, lo cual es de esperarse en datos de series de tiempo.
- Se cuentan con 7 columnas:
  - Open.
  - High.

- Low.
- Close.
- Volume.
- Dividends
- Stock Splits.

Para nuestros fines solamente basta concentrarse en los precios de apertura, máximo, mínimo y de cierre, los cuales corresponden a las primeras 4 columnas. También vale la pena verificar que el tipo de dato sea el adecuado y que no haya datos nulos:

```
1 df.info()

<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 3771 entries, 2006-02-27 to 2021-03-05
Data columns (total 7 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Open            3770 non-null   float64
1   High            3770 non-null   float64
2   Low             3770 non-null   float64
3   Close           3770 non-null   float64
4   Volume          3770 non-null   float64
5   Dividends       3771 non-null   float64
6   Stock Splits    3771 non-null   int64
dtypes: float64(6), int64(1)
memory usage: 235.7 KB
```

Figura 2: Información sobre el dataframe de la empresa Grupo Aeroportuario del Pacífico.

Los tipos de datos son adecuados y no hay datos faltantes, por lo que se puede empezar el análisis.

Para el resto de empresas los dataframes tienen la misma estructura (índices y columnas) y los mismos tipos de datos difiriendo solamente en los valores de las acciones. Para todos los dataframes que se trabajaron el número de observaciones fue superior a 1500, lo cual es un número adecuado de observaciones para realizar los análisis.

## Postwork 5.

Como se mencionó en la sección anterior los datos con los que se trabajó no contienen datos nulos (NaN's) y los tipos de datos son correctos para el análisis posterior.

## Postwork 6.

Como ya se mencionó, la API de *Yahoo Finance* se puede usar libremente sin necesidad de crear una cuenta además de que el número de peticiones es bastante alto pues durante este trabajo nunca se tuvieron problemas a la hora de importar los datos.

Esta API se puede usar por medio de la librería `yfinance` la cual se puede importar en Python. La sintaxis que se usó en este trabajo fue bastante sencilla.

1. Primero se pedían los datos a la API por medio del comando `yf.Ticker()`
2. Posteriormente se convertían los datos históricos de los precios de las empresas por medio del método `.history()`

De esta forma sencilla se obtuvieron los dataframes de los precios de cada empresa.

## Postwork 7.

Aunque a priori los datos presentan una buena estructura en el sentido de que el tipo de dato es correcto ocurre que al visualizar los datos pueden aparecer ciertos imprevistos no contemplados o anomalías que corresponden a la naturaleza intrínseca de los datos. En este trabajo nos ocurrieron algunas las cuales vale la pena mencionar:

- Para la empresa **Cemex** apareció un outlier en tres de las cuatro columnas de los precios, por lo que se hizo un análisis cuidadoso de ese punto para decidir como imputarlo.
- Las empresas **Corporativo Fragua** y **América Móvil** presentaron zonas donde los precios permanecieron constantes y después dieron un salto abrupto. Dado que este comportamiento es más debido al manejo de los valores de la empresa se decidió filtrar la base para evitar esas zonas de saltos abruptos.
- Para la empresa **Walmart** se encontró una zona bastante rara en comparación al resto de la serie. Al investigar no se encontró razón por la que dicha zona se encontrara ahí, por lo que se decidió filtrar la base para evitar estas anomalías.

Todos los detalles sobre estas anomalías encontradas en las bases se detallan mejor en el Jupyter notebook.

## Postwork 8.

Finalmente todo lo anterior expuesto en palabras es implementado en código en un Jupyter notebook el cual se puede encontrar en el link del repositorio de Github:

<https://github.com/setjaf/ProyectoBEDUPython20>

## Referencias

- [1] Engle, Robert F., Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*, 1982.
- [2] Tsay, Ruey S., *An Introduction to Analysis of Financial Data with R*
- [3] <https://www.bmv.com.mx/es/empresas-listadas>
- [4] <https://pypi.org/project/yfinance/>