

**Universidad de Buenos Aires**  
**Facultad de Ciencias Económicas**

Carrera de Grado - Actuario

**Estadística Actuarial**

1° cuatrimestre 2022

**Trabajo Práctico Simulación**

Cátedra: Landro - Del Rosso

Docentes: Landro - Del Rosso

Apellido, Nombre y N° de Registro de los 6 (seis) autores:

- Bocchio, Mateo. Registro: 899166.
- Fernández Petrillo, Camila. Registro: 899043.
- Gorbato, Ezequiel. Registro: 896447.
- Huergo, Ian Matias. Registro: 899961.
- Rojas, Agustin Silvio. Registro: 896355.
- Schneider, Steven. Registro: 900263.

**“Análisis de Series Temporales Univariadas”**

**Modelos *ARIMA* (p,d,q)**



**UBA**  
Universidad de Buenos Aires



**Universidad de  
Buenos Aires**  
Facultad de Ciencias  
Económicas

## Resumen

Se realiza un análisis descriptivo y econométrico de una serie autogenerada y ejecutada en el lenguaje de programación Python, a partir de la aplicación de los test de significatividad individual, de incorrelación, de normalidad de los errores, de presencia de raíz unitaria; y a partir de realizar análisis gráficos sobre el comportamiento de la serie y sus funciones correspondientes - FAC y FACP.

En primer lugar, se ejecuta la serie permitiendo así realizar un análisis gráfico que concluye en la condición de *no estacionariedad* de la serie. A continuación se realiza un breve análisis descriptivo. Luego, se evalúa de forma gráfica el comportamiento de la función de autocorrelación y la función de autocorrelación parcial donde se ratifica la condición de *no estacionariedad* de la serie. A partir de dicho análisis, se realizará un test de Dickey-Fuller Aumentado para confirmar dicha aseveración. Se concluirá que la serie posee al menos una raíz unitaria y se finalizarán los estudios de la condición de estacionariedad de la serie.

En segundo lugar, se realiza la primera diferencia de la serie, en búsqueda de convertir el proceso en estacionario. Se realiza un análisis gráfico sobre la serie en su primera diferencia, y sobre sus funciones de autocorrelación y autocorrelación parcial. Se concluye que la serie observada corresponde a un *proceso estacionario*. A continuación, se realiza el test de Dickey-Fuller para convalidar dicha afirmación, y se obtiene como resultado que la serie no posee raíces unitarias, es decir, que la serie es *estacionaria*.

En tercer lugar, se estiman distintos modelos ARIMA para concluir cuál de ellos se ajusta mejor a la serie *original*. Se concluye, a partir de la observación de distintos criterios de información, que el modelo que mejor ajusta a la serie es el modelo ARIMA (4,1,3). A continuación, se realizan los respectivos test de significatividad individual y los diversos análisis de los residuos, donde se concluye que los supuestos de normalidad e incorrelación *no* se cumplen para la serie trabajada.

Por último, se procede a realizar una estimación de los 10 primeros horizontes de predicción del modelo, de forma de predecir los valores que el modelo utilizado toma para los períodos futuros. Se obtienen dichas estimaciones y se realizan intervalos de confianza a 95%, 97,5% y 99% de confianza. Sin embargo, dicho análisis se encuentra limitado por el no cumplimiento de los supuestos de normalidad e incorrelación de los residuos de la serie en cuestión.

## **Abstract**

A descriptive and econometric analysis of a self-generated series executed in the Python programming language is carried out, based on the application of the tests of individual significance, of non-correlation, of normality of errors, of the presence of a unit root; and from performing graphical analysis on the behavior of the series and its corresponding functions - ACF and PACF.

First of all, the series is executed, thus allowing a graphical analysis to be carried out that concludes in the condition of non-stationarity of the series. A brief descriptive analysis follows. Then, the behavior of the autocorrelation function and the partial autocorrelation function are evaluated graphically, where the non-stationarity condition of the series is ratified. Based on this analysis, an Augmented Dickey-Fuller test will be performed to confirm this assertion. It will be concluded that the series has at least one unit root and the studies of the stationarity condition of the series will be completed.

Secondly, the first difference of the series is performed, in search of turning the process into stationary. A graphical analysis is performed on the series in its first difference, and on its autocorrelation and partial autocorrelation functions. It is concluded that the observed series corresponds to a stationary process. Next, the Dickey-Fuller test is carried out to validate this statement, and the result obtained shows that the series does not have unit roots, that is, that the series is stationary.

In third place, different ARIMA models are estimated to conclude which of them best fits the original series. It is concluded, based on the observation of different information criteria, that the model that best fits the series is the ARIMA model (4,1,3).

Finally, an estimation is done for the first 10 prediction horizons of the model, in order to predict the values that the model used will take for future periods. Those estimations are obtained and confidence intervals are done at 95%, 97.5% and 99% confidence. However, this analysis is limited by the fact that the assumptions of normality and uncorrelatedness of the residuals of the series in question are not met.

## **Índice:**

<b>Introducción</b>	<b>5</b>
<b>Marco Teórico</b>	<b>6</b>
Modelos	6
Test de Ljung Box	8
Test de Jarque-Bera	8
<b>Análisis de resultados</b>	<b>10</b>
Serie Original	10
Análisis descriptivo	10
Análisis de estacionariedad	13
Primera diferencia de la serie	14
Modelos ARIMA	15
Análisis de los residuos	16
Predicción	17
Intervalos de confianza	17
<b>Conclusiones</b>	<b>18</b>
<b>Futuras investigaciones y Limitaciones</b>	<b>19</b>
<b>Bibliografía</b>	<b>20</b>
<b>Anexos:</b>	<b>20</b>

# Introducción

En el presente trabajo, se generó una serie de tiempo a partir de una semilla creada en base a los números de registro de los 6 participantes.

Las series temporales muestran una secuencia de observaciones que se registran en intervalos de tiempo regulares. Sirven para estudiar la relación causal entre variables que presentan cambios en el tiempo, y tienen influencia entre sí.

Una serie puede ser entendida a través de 4 componentes inobservables: ciclicidad, tendencia, estacionalidad y ruido.

Lo cual se puede expresar como:

$$Y_t = C_t + T_t + S_t + \epsilon_t^1$$

A partir de métodos de descomposición, podremos aislar dichos elementos:

La ciclicidad hace referencia a patrones que se identifican en ciertos intervalos de tiempo; la tendencia es el movimiento hacia arriba o hacia abajo de los datos; cuando se habla de estacionalidad se hace referencia al patrón que muestran los datos en intervalos regulares de tiempo (principalmente referido a las estaciones del año); y finalmente el ruido son irregularidades aleatorias, que presentan una distribución con media igual a cero. Este último componente es el único componente aleatorio de las series.

El objetivo de este trabajo será aplicar los conocimientos adquiridos en clase. En primera instancia se realizará un análisis descriptivo de la serie. Se calculará media, mediana, desvío, varianza, percentiles, asimetría y kurtosis. Posteriormente se desarrollará un análisis econométrico de la serie. Se analizará la estacionariedad de la serie poniendo a prueba la presencia de raíces unitarias mediante el test de Dickey Fuller. Además, este análisis será complementado con la visualización gráfica de la serie, de su función de autocorrelación, función de autocovarianzas y función de autocorrelación parcial.

Debido a que la serie original resulta ser no estacionaria, se estimarán modelos ARIMA para la serie de tiempo con el objetivo de encontrar el que mejor se adapte a la serie original teniendo en cuenta los criterios de información y en base a el modelo seleccionado se realizarán predicciones y se calcularán intervalos de confianza.

La totalidad de los gráficos, tests, estimaciones y demás fueron realizados a través del lenguaje de programación Python. Los resultados serán plasmados e interpretados en este archivo, y además se podrá acceder a los códigos utilizados en el apartado "Anexo".

---

<sup>1</sup> Descomposición de la serie  $Y_t$  en su forma aditiva. Se debe recordar que la misma también puede ser descompuesta en su forma multiplicativa, tal que:  $Y_t = C_t * T_t * S_t * \epsilon_t$

# Marco Teórico

## Modelos

Los modelos ARIMA(p,d,q) están compuestos por una parte autoregresiva que se representa con un AR(p), una parte de media móvil o término de error MA(q) y por último el orden de integración del modelo I(d) que se refiere a la cantidad de veces que se debe diferenciar la serie para convertirla en estacionaria. Pero para llegar a los modelos ARIMA, primero se deben analizar los modelos por separado:

**Modelos AR(p):** son modelos que miden la dependencia lineal de las variables con sus observaciones pasadas. En los modelos autorregresivos los resultados anteriores del proceso afectan al comportamiento futuro. Se representan de la siguiente manera:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t$$

Utilizando la notación de operadores:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) Y_t = \varepsilon_t$$

**Modelos MA(q):** son modelos estrictamente estacionarios, de memoria muy corta. El valor actual de un modelo MA depende de las  $q$  últimas innovaciones. Se puede representar de la siguiente forma:

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

Utilizando la notación de operadores:

$$Y_t = \varepsilon_t (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$$

**Modelos ARMA(p,q):** es un modelo mixto que combina características de los modelos puros AR y MA. Se pueden expresar de la siguiente forma:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

Utilizando la notación de operadores:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) Y_t = \varepsilon_t (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$$

Con la notación de operadores se pueden calcular las raíces de los coeficientes, ya sea de la parte AR o de la parte MA. A través de los coeficientes de la parte autorregresiva

del modelo se puede analizar si la serie es estacionaria. Que sea estacionaria implica que los momentos de orden  $s$  sean invariantes a través del tiempo y que la varianza sea constante y finita. Por lo tanto para que el modelo sea estacionario, o dicho de otra manera integrado de orden 0, se pide que las raíces de la parte autoregresiva sean en módulo mayores que la unidad. Cuando la serie temporal tiene al menos una raíz unitaria, el proceso no es estacionario, se trata de un proceso integrado  $I(d)$ . La presencia de raíces unitarias genera dependencia temporal.

Un proceso puede ser no estacionario en la media, en la varianza, en las autocorrelaciones o en otras características de la distribución de las variables. Cuando el nivel de la serie no es estable en el tiempo, pudiendo tener tendencia creciente o decreciente, la serie no es estacionaria en la media. Cuando la variabilidad o las autocorrelaciones se modifican con el tiempo, diremos que la serie es no estacionaria en la varianza o en las autocorrelaciones. Los procesos no estacionarios más importantes son los procesos integrados que tienen la propiedad fundamental que al diferenciarlos se convierten en procesos estacionarios. De aquí surgen los modelos ARIMA.

**Modelos ARIMA(p,d,q):** Tres factores definen a los modelos ARIMA (p,d,q). En esta notación  $p$  es el orden de la parte autoregresiva,  $d$  es el número de raíces unitarias (orden de integración del proceso) y  $q$  es el orden de la parte media móvil.

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) (1 - B)^d Y_t = \varepsilon_t (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$$

Al diferenciar estos procesos integrados, desaparece la dependencia temporal. La dependencia temporal la podemos observar en la función de autocorrelación (FAC) la cual disminuye linealmente con el tiempo, es por eso que estos procesos son de memoria larga.

Es posible realizar una gran variedad de pruebas estadísticas sobre los modelos mencionados previamente. A continuación se detallarán tres pruebas que permiten analizar la significatividad individual de los coeficientes, normalidad e incorrelación.

## Test de significatividad individual

Se plantea como hipótesis nula que el coeficiente es igual a cero, por lo cual se considera no significativo. En cambio, la hipótesis alternativa indica que el coeficiente es significativo. Es decir:

$$H0 : \phi_i = 0$$

$$H1 : \phi_i \neq 0$$

El  $t$  empírico utilizado viene dado por la siguiente fórmula:

$$T^e = \frac{\hat{\phi}_i - \phi_i}{\sigma(\hat{\phi}_i)} \sim t_{n-p-1}$$

Y la regla de decisión indica que en aquellos casos donde el t empírico en módulo sea mayor a dos, se rechaza  $H_0$ , y se concluye en que el parámetro es significativo.

## Test de Ljung Box

Es una prueba estadística que permite analizar la incorrelación. La hipótesis nula y la hipótesis alternativa se plantean de la siguiente manera:

$$H_0: \rho_0(Y) = \rho_1(Y) = \dots = 0$$

$$H_1: \text{existe al menos un } \rho_i(Y) \neq 0$$

El t empírico se calcula mediante la siguiente fórmula:

$$Q_j^{LB} = n(n+2) \frac{\sum \rho_i^2(\hat{\varepsilon})}{n-i} \sim \chi^2_{j-(p+q+1)}$$

Y la regla de decisión indica que en aquellos casos donde el p valor sea mayor a alpha, se decide no rechazar  $H_0$ , es decir que los errores están incorrelacionados.

## Test de Jarque-Bera

Es una prueba estadística que permite testear normalidad. La hipótesis nula y la hipótesis alternativa se plantean de la siguiente manera:

$$H_0: \varepsilon_t \sim N$$

$$H_1: \text{no especifica}$$

El t empírico se calcula mediante la siguiente fórmula:

$$JB = \frac{n}{6} [\widehat{AS^2}(\varepsilon) + \frac{1}{4} (\widehat{k}(\varepsilon) - 3)^2] \sim \chi^2_2$$

Y la regla de decisión indica que si el p valor es mayor a alpha, se decide NO rechazar  $H_0$ , es decir que los errores siguen una distribución normal.

En aquellos casos en los que se desee decidir entre varios modelos cuál es mejor, se pueden utilizar los criterios de información. Es posible mencionar al menos tres criterios. Uno de ellos es el Akaike, el cual se considera un criterio eficiente, y también es posible mencionar otros dos, que son criterios consistentes: Schwarz y Hannan-Quinn. Se considera que el modelo más adecuado será aquel con criterio de información de menor valor.

Para testear la presencia de raíces unitarias se utiliza el test de Dickey Fuller. En aquellos casos en los que haya presencia de tendencia se utiliza el test de Dickey Fuller Aumentado.



El procedimiento del test de Dickey Fuller Aumentado es el siguiente:

Partiendo de un AR(P) con término independiente y termino lineal:

$$Y_t = a_0 + a_1 t + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \epsilon_t$$

Se resta  $Y_{t-1}$  miembro a miembro, y agregando términos restando y sumando se obtiene la serie diferenciada:

$$\nabla Y_t = a_0 + a_1 t + (\phi_1^* - 1)Y_{t-1} - \sum_{j=1}^{p-1} \phi_{j+1}^* \nabla Y_{t-j} + \epsilon_t$$

$$\text{Con } \phi_j^* = \sum_{i=j}^p \phi_i$$

Esta última ecuación es la Ecuación de Dickey Fuller Aumentada.

Una vez que se llega a esta ecuación, se plantean las hipótesis:

$$H_0: \phi_1^* - 1 = 0 \quad I(1)$$

$$H_1: \phi_1^* - 1 < 0 \quad I(0)$$

En el Test de Dickey Fuller, se plantea como hipótesis nula que la serie es integrada de orden 1, mientras que la hipótesis alternativa indica que la serie es estacionaria.

**Hipótesis:**

$H_0$ : La serie es integrada de orden 1, es decir que posee una raíz unitaria, por lo tanto no es estacionaria.

$H_1$ : La serie es integrada de orden 0, es decir que no posee una raíz unitaria y que es estacionaria.

Una opción para tomar una decisión es mediante el estadístico:

$$\tau^e = \frac{\phi_1^* - 1}{\sigma(\phi_1^*)}$$

Si este valor es menor al  $\tau^c$ , se decide rechazar la hipótesis nula, de lo contrario se decide no rechazarla.

Otra opción es mediante el cálculo del p-value. Las dos posibilidades son:

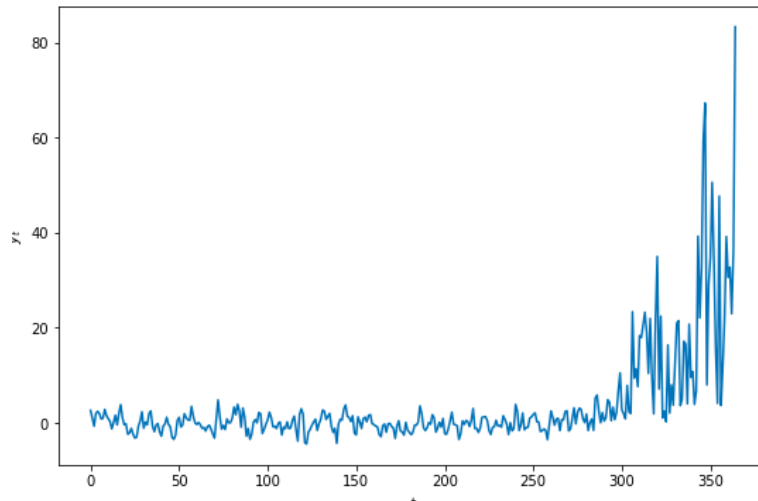
**p-value > 0.05:** no se rechaza  $H_0$ , por lo que se puede sostener la hipótesis de presencia de raíz unitaria.

**p-value <= 0.05:** se rechaza  $H_0$ ; por lo que se puede sostener que no hay ninguna raíz unitaria.

## Análisis de resultados

### Serie Original

Dada la semilla autogenerada se obtuvo una serie de tiempo que se comporta de la siguiente manera:



**Figura 1:** Serie de tiempo autogenerada

A simple vista, se observa que la serie es **no estacionaria en la media**, ya que la esperanza no se mantiene constante a través del tiempo. Se observa que al final de la serie, el comportamiento cambia y que hay cierta tendencia.

Si se decide analizar la serie dividida en intervalos, hasta  $t=300$  aproximadamente la serie pareciera comportarse de forma estacionaria, pero cuando se analiza toda la serie en su conjunto, no lo es. Dicha situación se puede solucionar utilizando el poderoso test de Zivot & Andrews.

El test mencionado se aplica en series que poseen cambios estructurales, en las cuales se presume que lo más conveniente es cortar la serie, separarla en dos y así analizarla. Aún así, en ésta investigación se analizará toda la serie en su conjunto. Este problema podrá ser abordado en futuras investigaciones.

Antes de continuar con los temas de estacionariedad y modelos ARIMA, se dedicará la primera parte del análisis de resultados al análisis descriptivo de la serie calculada previamente en Python.

### Análisis descriptivo

Media	Mediana	Varianza	Desvío	1° decil	1° cuartil	3° cuartil	Asimetría	Kurtosis
3,3950	0,4818	103,958	10,196	-1,9814	-0,9077	2,3360	3,9376	19,2097

**Tabla 1:** Análisis descriptivo de la serie.

Recordemos que dichas medidas numéricas se emplean para describir conjuntos de datos.

Existen tres principales conjuntos de medidas de interés para cualquier conjunto de datos: la localización de su centro<sup>2</sup>, o *medidas de tendencia central*; su variabilidad<sup>3</sup> y su forma.

Este informe propone abordar dos medidas de tendencia central: la media y la mediana; tres medidas de variabilidad: la varianza, el desvío estándar, y los cuartiles; y dos medidas de forma: la asimetría; y la kurtosis.

La *media* de las observaciones  $X_1, X_2, \dots, X_n$  es el promedio aritmético de éstas, y se denota por:

$$\bar{X} = \sum_{i=1}^n \frac{X_i}{n}$$

La *mediana* de un conjunto de observaciones es el valor para el cual, cuando todas las observaciones se ordenan de manera creciente, la mitad de éstas es menor que este valor, y la otra mitad mayor. Es decir, la mediana es el valor de la observación que se encuentra a la mitad del conjunto ordenado.

La *varianza* de las observaciones  $X_1, X_2, \dots, X_n$  es el promedio del cuadrado de las distancias entre cada observación y la media del conjunto de observaciones, y se denota por:

$$S^2 = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n - 1}$$

El *desvío estándar* es la raíz cuadrada positiva de la varianza, y se denota por:

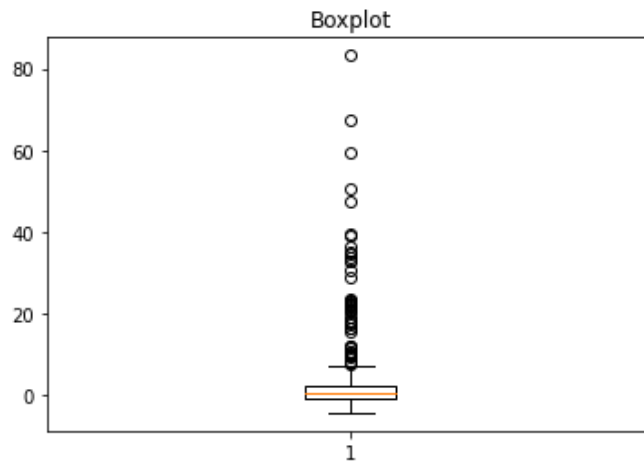
$$S = \sqrt{\sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n - 1}}$$

En la serie de datos presentada, el desvío es muy elevado, lo cual genera una gran dispersión de los datos respecto a su media. Lo podemos visualizar mejor con el siguiente diagrama de cajas donde hay muchos valores atípicos o muy alejados respecto a su media:

---

<sup>2</sup> La tendencia central de un conjunto de datos es la disposición de estos para agruparse ya sea alrededor del centro o de ciertos valores numéricos (Canavos, 1988, p. 12).

<sup>3</sup> La variabilidad de un conjunto de datos es la dispersión de las observaciones en el conjunto (Canavos, 1988, p. 12).



**Figura 2:** Diagrama de cajas de la serie

Los *percentiles*  $P_k$  son ciertos valores que dividen la sucesión de datos ordenados en cien partes porcentualmente iguales, y se calculan mediante la siguiente fórmula:

$$P_k = L_k + \frac{k \left( \frac{n}{100} \right) - F_k}{f_k} * c \quad k = 1, 2, 3, \dots, 99$$

La *asimetría* es un valor que nos proporciona información sobre si los datos se distribuyen en forma simétrica o sesgada hacia los lados, y se denota por:

$$AS(x) = \frac{mc_3}{\sigma^3}$$

La *kurtosis* es un valor cuya magnitud nos indica si los datos se distribuyen simétricamente de forma normal (curva mesocúrtica), más empinados que la curva normal (curva leptocúrtica), o más aplanados que la curva normal (curva platicúrtica); y se denota por:

$$k = \frac{\sum_{i=1}^n \frac{f_i (X_i - \bar{X})^2}{n}}{s^4}$$

4

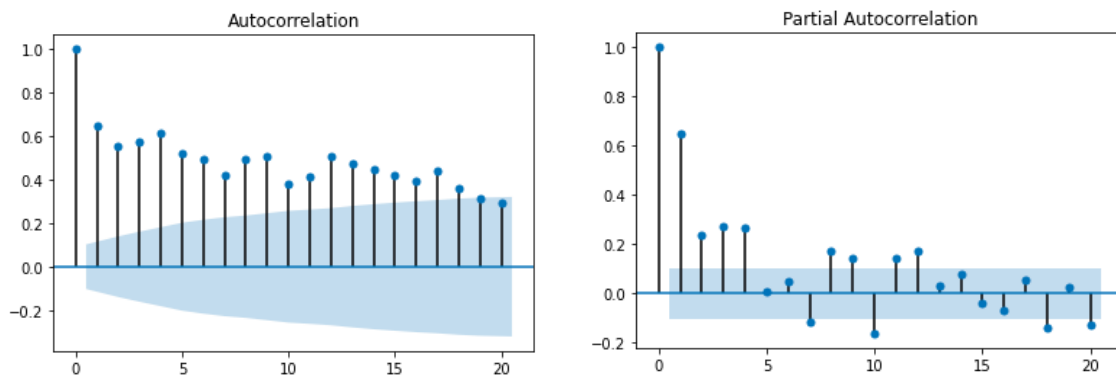
## Análisis de estacionariedad

A continuación, se trabajará la estacionariedad de la serie, analizando su función de autocorrelación y la función de autocorrelación parcial.

En los procesos estacionarios, la función de autocorrelación tiene una importancia fundamental. Si el proceso es estacionario de segundo orden, las características fundamentales del proceso estarán resumidas en su función de autocorrelación.

---

<sup>4</sup> donde  $k = 3$  indica una curva mesocúrtica.



**Figuras 3 y 4:** Funciones de autocorrelación y autocorrelación parcial respectivamente

La FAC es la memoria de la serie, en este caso decrece suavemente a lo largo del tiempo, tiene un comportamiento de decrecimiento lento. Esto es así debido a que el proceso tiene memoria de largo plazo. Al analizar la función de autocorrelación es posible concluir que la serie de tiempo no es estacionaria, en caso de que lo fuera se observaría un decrecimiento exponencial de la FAC. Entonces, se está trabajando con un proceso integrado, por lo que sería necesario diferenciar la serie para convertirla en estacionaria y eliminar la tendencia estocástica.

Aun así, la observación del gráfico de la FAC puede ser complementada con otro análisis. Se analizará el Test de Dickey Fuller Aumentado para comprobar la presencia de raíces unitarias. La presencia de raíces unitarias indicaría no estacionariedad, como ya fue explicado previamente en la sección “Marco teórico”.

En el caso bajo análisis, se realizó el test de Dickey-Fuller aumentado y el resultado del test, que fue realizado a través de Python, fue el siguiente:

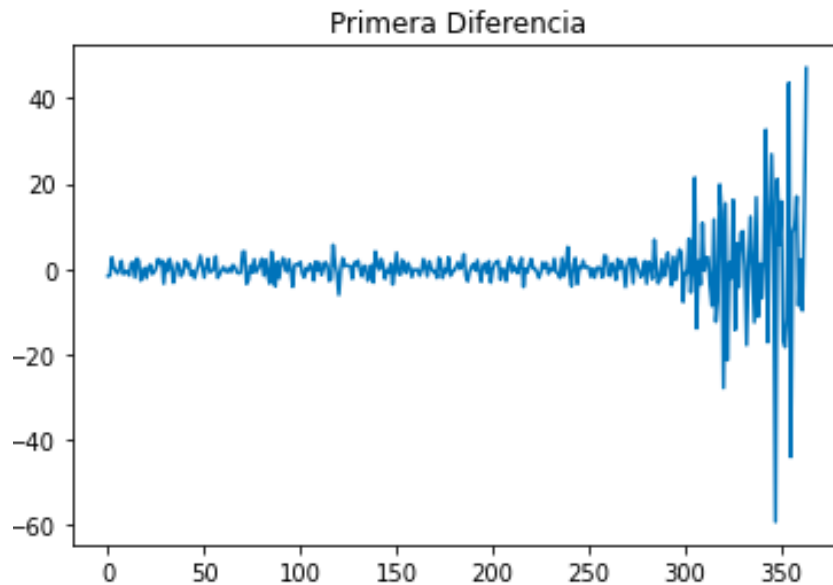
```
Valor del test: 2.983917287272448
p-value: 1.0
El p-value (1.0) es mayor que el alpha (0.05).
NO rechaza H0.
(2.983917287272448, 1.0)
```

Como consecuencia de los resultados, no se rechaza la hipótesis nula. La evidencia empírica disponible permite concluir en que la serie posee al menos una raíz unitaria y por lo tanto se trata de un proceso integrado de orden 1.

El test permite verificar el comportamiento analizado anteriormente a través del gráfico de la FAC. Entonces se confirma la necesidad de diferenciar la serie para convertirla en estacionaria.

## Primera diferencia de la serie

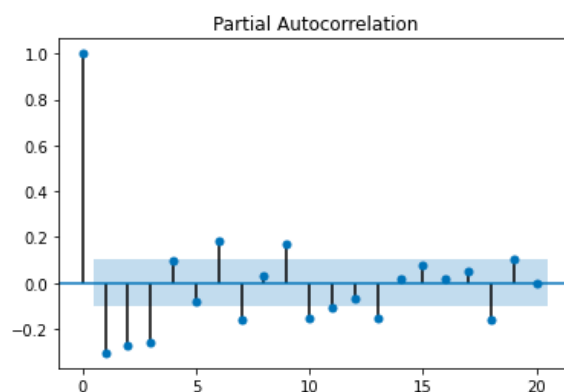
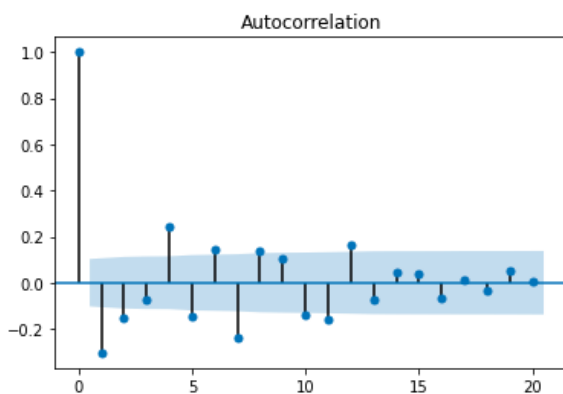
Se procede a realizar la primera diferencia de la serie en un intento de obtener una serie estacionaria. Al diferenciar la serie se pierde un dato (una observación), la serie queda transformada en un proceso que se comporta de la siguiente manera:



**Figura 5:** Serie de tiempo de la primer diferencia.

Es posible observar que, en esta nueva serie, el comportamiento se mantiene en torno a la media. A simple vista, la serie se comporta de forma estacionaria en torno a su media.

Se analizarán los gráficos de la FAC y FACP de la serie diferenciada y también se realizará el test de Dickey Fuller para corroborar que la serie tenga un comportamiento estacionario.



**Figuras 6 y 7:** Funciones de autocorrelación y autocorrelación parcial respectivamente de la serie diferenciada

De los gráficos presentados anteriormente se observa que la función de autocorrelación decrece de manera sinusoidal muy rápidamente lo cual indica que el proceso es de memoria corta, es decir, estacionario.

Al analizar la presencia de raíces unitarias mediante el test de Dickey Fuller para la serie diferenciada, se plantea nuevamente como hipótesis nula que la serie es integrada de orden 1 (no estacionaria) y se obtienen los siguientes resultados:

```
Valor del test: -6.1715175970750895
p-value: 6.787241622863207e-08
El p-value (6.787241622863207e-08) es menor que el alpha (0.05).
Rechazo H0.
(-6.1715175970750895, 6.787241622863207e-08)
```

Debido a que el p-valor es muy pequeño, y menor a alpha, se rechaza la hipótesis nula. Entonces, dada la evidencia empírica, se puede afirmar que la primera diferencia de la serie no posee raíces unitarias, lo que es equivalente a decir que es estacionaria.

## Modelos ARIMA

Se estimaron distintos modelos ARIMA partiendo de la serie original y con el objetivo de decidir cuál de ellos es el que se ajusta mejor a los datos se optó por analizar los criterios de información: Akaike, Schwarz y Hannan-Quinn.

Los datos obtenidos fueron los siguientes:

	ARIMA(1,2,1)	ARIMA(0,1,0)	ARIMA(1,1,1)	ARIMA(2,1,1)	ARIMA(3,1,2)	ARIMA(4,1,3)
<b>AIC</b>	2465,181	2501,430	2397,821	2395,788	2378,980	2355,600
<b>BIC</b>	2480,759	2509,225	2413,410	2415,274	2406,260	2390,675
<b>HQIC</b>	2471,373	2504,528	2404,017	2403,533	2389,823	2369,541

**Tabla 2:** Modelos ARIMA estimados

Los modelos ARIMA con  $d > 1$  tienen mayores coeficientes de AIC, BIC y HQ, cualquiera sea la parte autorregresiva y de medias móviles.

Se observa que de los modelos ARIMA generados, aquellos con  $d=1$  son los que tienen menor AIC, BIC y HQ, lo cual nos permite concluir que el proceso es integrado de orden 1, es decir, posee una única raíz unitaria.

Según los tres criterios y en base a las estimaciones realizadas, el modelo más adecuado es el ARIMA (4,1,3).

Además, se testeó la significatividad individual de los coeficientes de cada uno de los modelos estimados.

En el caso del modelo que se eligió como más adecuado, es decir el ARIMA (4,1,3) los resultados del test de significatividad individual de los coeficientes arrojaron que de los 7 coeficientes cuatro eran significativos, mientras que los tres restantes no.

Específicamente, tanto el coeficiente término independiente  $\phi_0$ , como el coeficiente  $\phi_3$  y el coeficiente  $\phi_4$  resultan no ser significativos, por lo cual podrían omitirse del modelo.

Al analizar el resto de los modelos estimados, se observa que el modelo ARIMA(3,1,2) si bien tiene AIC, BIC y HQ un poco más alto, posee la totalidad de coeficientes significativos, es decir que todos son importantes a la hora de predecir el modelo.

Se podría entonces analizar más a fondo ésta situación para, teniendo en cuenta los resultados que arrojó el test de significatividad individual, determinar el modelo más adecuado.

## Análisis de los residuos

De la serie diferenciada:

A través del test de Ljung Box, se rechaza la Hipótesis Nula por lo que los residuos están correlacionados.

```
Valor del test: 13.33333333333332
p-value: 0.0002607296328553169
El p-value (0.0002607296328553169) es menor que el alpha (0.05).
Rechazo H0.
```

A través del test de Jarque Bera, también se rechaza la Hipótesis nula, por lo que los errores no se distribuyen como una distribución normal.

```
Valor del test: 7359.259989600171
p-value: 0.0
El p-value (0.0) es menor que el alpha (0.05).
Rechazo H0.
(7359.259989600171, 0.0)
```

## Predicción

A través de los modelos estimados, y dado el pasado de una serie de tiempo, se intentará predecir los siguientes valores que tomará la serie. Por lo tanto se tendrá un predictor y un término de error. A medida que se amplía el rango del horizonte de predicción los errores irán incrementándose y la predicción será cada vez menos certera. Esto se demostrará con la realización de intervalos de confianza.



Con la serie observada  $Y_{t-n}, Y_{t-n+1}, \dots, Y_t$  se intentará predecir  $Y_{t+l}$ , que representa el parámetro a estimar. A  $l$  se lo denomina horizonte de predicción y el estimador del predictor se detona de la siguiente forma:

$$\hat{Y}_{t+l/t} \equiv f(t, l)(Y_{t-n}, Y_{t-n+1}, \dots, Y_t)$$

El predictor es aquel que minimiza la varianza del error de predicción.

El valor desconocido que se intenta predecir se puede descomponer con un predictor sumado a un término de error. Se puede observar en la siguiente fórmula:

$$Y_{t+l} = \hat{Y}_{t+l/t} + e_{l/t}$$

Los resultados obtenidos para el valor de predicción con un horizonte de predicción de 10 periodos del ARIMA(4,1,3) son los siguientes:

```
El valor de la serie en t + 1 es 47.2047
El valor de la serie en t + 2 es 62.6832
El valor de la serie en t + 3 es 72.8735
El valor de la serie en t + 4 es 48.0026
El valor de la serie en t + 5 es 34.3688
El valor de la serie en t + 6 es 44.136
El valor de la serie en t + 7 es 40.777
El valor de la serie en t + 8 es 39.6847
El valor de la serie en t + 9 es 42.6628
El valor de la serie en t + 10 es 41
```

## Intervalos de confianza

Para calcular los intervalos de confianza en los que se encontraría el valor estimado, con cierto nivel de confianza se utiliza la siguiente fórmula:

$$P\left(\hat{Y}_{t+l/t} - Z_{1-\frac{\alpha}{2}}\sigma[e(l, t)] \leq Y_{t+l} \leq \hat{Y}_{t+l/t} + Z_{1-\frac{\alpha}{2}}\sigma[e(l, t)]\right) = 1 - \alpha$$

En el caso que se está analizando, se realizarán los intervalos con un horizonte de predicción  $l=10$  y con distintos niveles de significatividad: 95%, 97,5% y 99%. A continuación se muestra los intervalos de confianza con un 95%. Podría observarse el resto en el archivo de Google Colab adjuntado en el Anexo.

```
El Intervalo de Confianza de la serie en t + 1, con un alpha de 5%, es: [35.48366247 58.9257 ]
El Intervalo de Confianza de la serie en t + 2, con un alpha de 5%, es: [49.69032371 75.67598785]
El Intervalo de Confianza de la serie en t + 3, con un alpha de 5%, es: [59.71903436 86.02793528]
El Intervalo de Confianza de la serie en t + 4, con un alpha de 5%, es: [34.76562301 61.23957941]
El Intervalo de Confianza de la serie en t + 5, con un alpha de 5%, es: [20.21918718 48.51840432]
El Intervalo de Confianza de la serie en t + 6, con un alpha de 5%, es: [29.97564004 58.29644077]
El Intervalo de Confianza de la serie en t + 7, con un alpha de 5%, es: [26.16530373 55.38870462]
El Intervalo de Confianza de la serie en t + 8, con un alpha de 5%, es: [24.84148138 54.52787989]
El Intervalo de Confianza de la serie en t + 9, con un alpha de 5%, es: [27.67046164 57.65508098]
El Intervalo de Confianza de la serie en t + 10, con un alpha de 5%, es: [25.24962167 55.86705581]
```

Dado los resultados es importante mencionar algunas consideraciones: los intervalos de predicción de los modelos ARIMA se basan en suposiciones acerca de que los residuos no están correlacionados y distribuidos normalmente, situación que no ocurre en la serie de tiempo que se utilizó para predecir, por lo cual, los intervalos de predicción podrían ser incorrectos. Además, como  $d=1$ , los intervalos de predicción seguirán creciendo en un futuro.

## Conclusiones

En conclusión, es posible enunciar que la mayoría de las series de tiempo del mundo real están atravesadas por tendencias o ciclos los cuales es importante desglosar y analizar. De aquí surge el estudio de las series de tiempo, para intentar obtener algún comportamiento de un fenómeno determinado y de esta manera poder predecir. Los modelos ARIMA( $p,d,q$ ) sirven para intentar analizar y predecir el comportamiento de un fenómeno a través del tiempo.

En el presente trabajo, se auto generó una semilla y se simuló una serie de tiempo que fue descompuesta y analizada mediante las herramientas adquiridas en el curso de estadística actuarial. Es muy útil para el aprendizaje poder aplicar la teoría y tener luego la posibilidad de visualizar de manera práctica cómo de una serie original de datos no estacionarios, haciendo ciertas transformaciones, es posible transformar la serie estacionaria. Y también es muy interesante poder contrastar los resultados de pruebas estadísticas y otro tipo de tests en series estacionarias y series no estacionarias. Además, la herramienta Python permite realizar todo el trabajo en términos muy dinámicos. Otorga la posibilidad de realizar análisis que quizás serían imposibles de hacer manualmente, o requerirían de muchas horas de trabajo.

De esta forma, se concluye con el análisis completo de la serie de tiempo, siguiendo la metodología de Box Jenkins: identificación del modelo, a través de la estimación de los distintos ARIMA; estimación de los parámetros y el análisis de su significatividad; diagnóstico comparativo y el pronóstico o predicción de la serie.

Aun así, quedan temas pendientes por investigar y se podría realizar un análisis mucho más extenso y profundo sobre la serie de datos con temas que nos exceden. Es por eso que se deja abierta la puerta para futuras investigaciones.

## Futuras investigaciones y Limitaciones

Dada la primera diferencia de la serie, el test de DF aseguró que la serie es estacionaria. Aun así, puede haber estacionalidad tanto en la media como en la varianza, y que la serie sea estacionaria en la media, no implica que sea estacionaria en la varianza. Ser estacionario de orden uno, no implica ser estacionario de orden dos.

Al analizar el gráfico de la primera diferencia de la serie, se presenta el problema de heterocedasticidad en la parte final de la serie. Para solucionar este problema y convertir la

serie en homocedástica se podría optar por aplicar una transformación. Hay distintos tipos de transformaciones aplicables a la serie para eliminar la heterocedasticidad y convertir a la serie en estacionaria de segundo orden. A continuación se mencionan algunas de las transformaciones que se podrían realizar.

Transformación de Box-Cox:

$$T^{\lambda}(Y_t) = \begin{cases} \frac{Y_t^{\lambda}-1}{\lambda} & \text{si } 0 < \lambda \leq 1 \\ \ln Y_t & \text{si } \lambda = 0 \\ -Y_t & \text{si } \lambda < 0 \end{cases}$$

Transformación de Tukey:

$$T^{\lambda}(Y_t) = \begin{cases} Y_t & \text{si } \lambda = 1 \\ Y_t^2 & \text{si } \lambda = 1/2 \\ \frac{1}{Y_t} & \text{si } \lambda = -1 \\ \sqrt{Y_t} & \text{si } \lambda = 2 \end{cases}$$

También está la transformación de Wallis (1987), de Anscombe (1948), de Fisher (1925).

Otra cuestión importante que en este trabajo no se analizará es el Test de Zivot & Andrews. En ciertas series, existen cambios estructurales, lo que genera que el test de Dickey Fuller pueda fallar al rechazar la hipótesis nula de raíz unitaria si existe un cambio estructural. Es por este motivo, que Zivot & Andrews elaboraron un test en series donde hay un punto de quiebre determinado de forma exógena. En dicho test, las hipótesis son las siguientes:

H0: posee raíz unitaria con cambios estructurales.

H1: es estacionario con cambios estructurales.

Sería interesante analizar en futuras investigaciones dicho test en profundidad y por su parte, el cambio estructural de la serie.

## Bibliografía

Landro, A. H., & González, M. L. (2009). *Elementos de econometría de los fenómenos dinámicos*. Ediciones Cooperativas.

Canavos, G., Meyer, P., Spiegel, M., & Mendenhall, S. (1988). Probabilidad y estadística. *Licenciatura en Ingeniería en Informática*, 28.

Lincoln L. Chao. (1994). *Estadísticas para las ciencias administrativas*. Tercera edición.

Anderson, Sweeney, Williams. (2008). *Estadística para administración y economía*. Décima edición.

Peña, D. (2005). *Análisis de series temporales*. Alianza.

Uriel, E. (1985). *Análisis de series temporales modelos ARIMA*. Paraninfo.

Zivot, E., & Andrews, D. W. K. (1992). Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of business & economic statistics*, Vol 10, pp. 251-270.

## Anexos:

Se anexa un Google Colab, donde se elaboró el correspondiente lenguaje para el análisis expuesto en el presente trabajo:

[https://colab.research.google.com/drive/1ocZXUKftbRDcSiR6RaklZc5XqfMjt8w8?hl=es#scrollTo=QOTNrcPmz\\_p7](https://colab.research.google.com/drive/1ocZXUKftbRDcSiR6RaklZc5XqfMjt8w8?hl=es#scrollTo=QOTNrcPmz_p7)