Modelo ARIMA aplicado al PIB de México

José Julián Mateo López

2022-9-22

Introducción

Los modelos econométricos son una de las principales herramientas de análisis económico cuantitativo para los economistas. Una de las principales funciones de los modelos econométricos es poder predecir el comportamiento futuro de múltiples fenómenos económicos. Uno de los modelos clásicos para analizar las series temporales económicas es la aplicación del modelo autorregresivo integrado de medias móviles (ARIMA).

Obtención de la información

```
# Activamos la libreria readxl para que el lenguaje nos permita importar nuestra base de datos
library(readxl)
# El objeto que guarda nuestra base de datos se denomina "df", que recoge nuestro archivo en formato ex
df <- read_excel("Modelo ARIMA PIB.xlsx")
#atamos los datos
attach(df)
#Le indicamos a R que nos muestre las primeras observaciones de la serie
head(df)
```

```
## # A tibble: 6 x 1
##          PIB
##        <dbl>
## 1 12725022.
## 2 12994569.
## 3 13008792.
## 4 13003302.
## 5 12796119.
## 6 12967530.
```

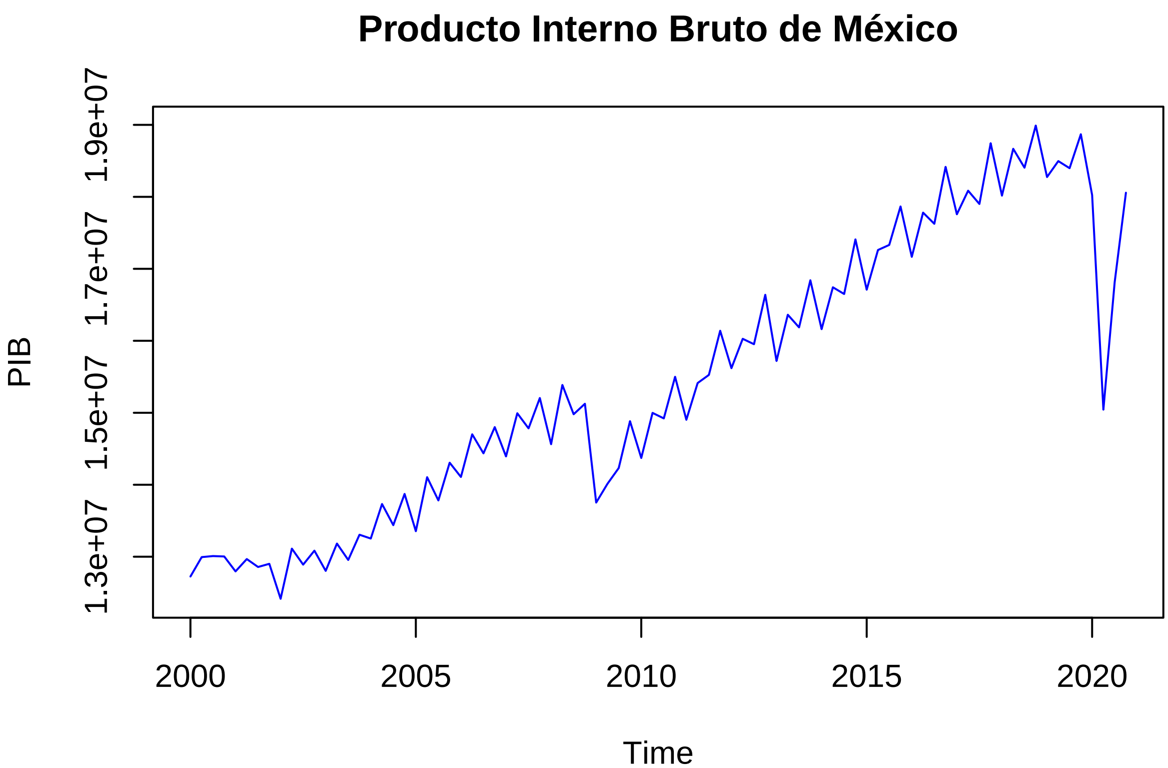
Análisis preeliminar de la serie

Análisis gráfico de la serie PIB de México

Una de las primeras condiciones para poder realizar un modelo de este tipo, es que la serie a analizar debe ser estacionaria, es decir, debe tener media y varianza constante. O sea, debe ser un proceso estocástico puramente aleatorio de ruido blanco. Por lo general, el análisis de una serie empieza por la visualización del comportamiento estadístico de la serie a través del tiempo, que para nuestro ejercicio, empieza a partir del primer trimestre del 2000 y termina hasta el ultimo trimestre del 2020.

```r
# Para poder gráficar la serie, previamente necesitamos darle un formato de series de tiempo, de forma
#que el lenguaje me permita trabajar mejor con los procesos de modelamiento estadístico.
#El nuevo objeto de denominará "PIB_ts" que guarda nuestra base de datos
#en un formato de time series. Tal como se muestra en la siguiente linea de código.

PIB_ts <- ts(PIB, start = 2000, frequency = 4)
plot(PIB_ts, main = "Producto Interno Bruto de México",
     col = "blue", ylab = "PIB")
```



Como se puede apreciar en la gráfica anterior, el PIB de México presenta una tendencia estocástica o determinista. Cuando se tiene experiencia trabajando con series de tiempo, rápidamente a partir de la visualización gráfica se puede inferir el comportamiento de esta. El análisis gráfico representa el primer paso en el análisis exploratorio de los datos. En este sentido, el comportamiento estadístico del PIB de México no contiene una media o varianza constante, por el contrario, presenta una amplia volatilidad durante el tiempo, principalmente en los años 2008-2009 y en el año 2020, que se puede apreciar una significativa caída del crecimiento económico en esos años, por supuesto, esto se le puede atribuir a las fuertes crisis económicas de esos años. Ahora dirigiéndonos a pruebas estadísticas mas formales, podemos aplicar la prueba de Dickey Fuller Aumentada para cerciorarnos de la estacionariedad de la serie. Es cierto que en la literatura econométrica existe una amplia gama de pruebas que se podrian aplicar para determinar la existencia de una raíz unitaria, sin embargo, en este trabajo nos enfocaremos principalmente en el test ADF, que por cuestiones practicas, nos resulta fácil llevarlo a cabo en R.

```r
# Primero activamos la librería URCA
library(urca)
# Una vez activada, llevamos a cabo la aplicación del test.
summary(ur.df(PIB_ts))
```

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3465021  -186154    44932   294956  1989726
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## z.lag.1     0.004898   0.004143   1.182    0.241
## z.diff.lag -0.468061   0.101820  -4.597 1.58e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 581500 on 80 degrees of freedom
## Multiple R-squared:  0.2133, Adjusted R-squared:  0.1936
## F-statistic: 10.85 on 2 and 80 DF,  p-value: 6.792e-05
##
##
## Value of test-statistic is: 1.1823
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau1 -2.6 -1.95 -1.61
```

Para llevar a cabo el contraste de este test, formulamos la siguiente hipótesis nula:

Ho: La serie presenta al menos una raíz unitaria.

Ha: La serie no presenta una raíz unitaria.

La regla de decisión es la siguiente:

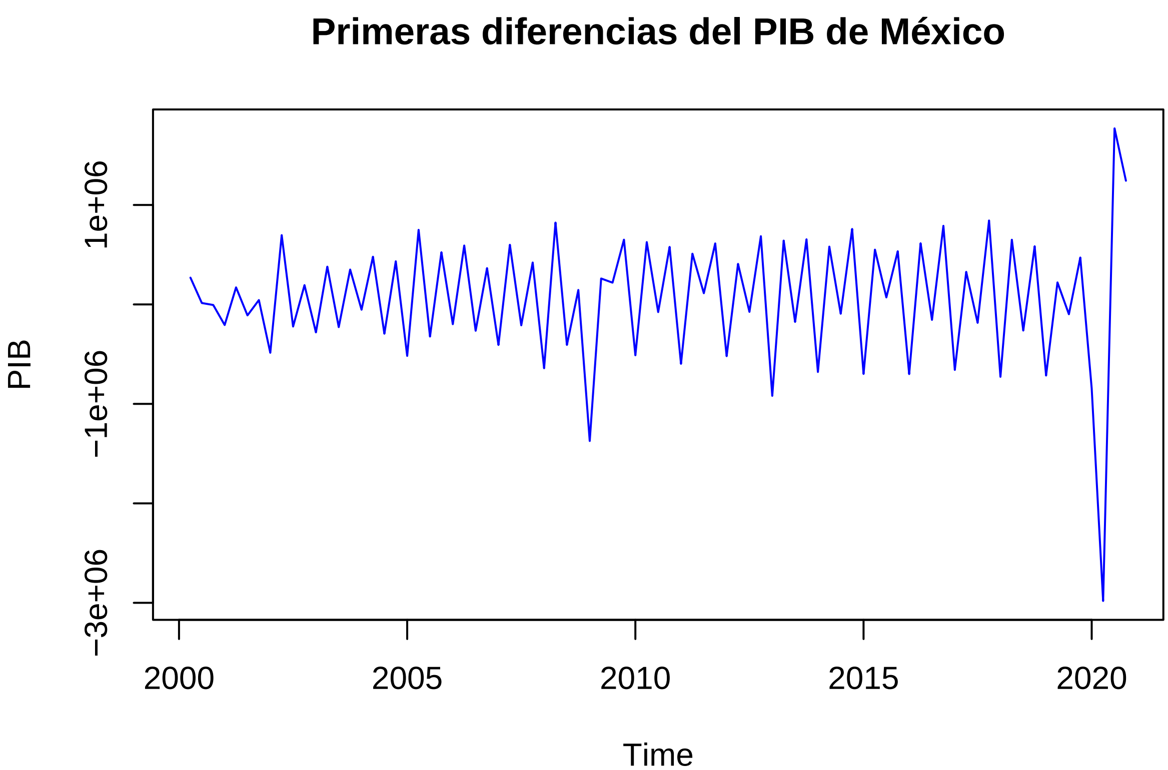
si el estadístico calculado es mayor a los valores críticos de tau, se rechaza la hipótesis nula. Como se puede apreciar, de acuerdo a los resultados, el estadístico ADF recogió un valor de 1.18, por consiguiente, no se rechaza la hipótesis nula, y se concluye que la serie presenta al menos una raíz unitaria. Es decir, es una serie de caminata aleatoria, que de ser el caso contrario, sería una serie puramente aleatoria de ruido blanco.

Por consiguiente, al verificar que la serie PIB de México es no estacionaria, es necesario aplicar las técnicas necesarias que permitan transformar la serie no estacionaria a una serie estacionaria. Para ello, procedemos a diferenciar la serie. Primero, llevaremos a cabo las primeras diferencias para verificar si la seria ya cumple con los supuestos de ruido blanco. Es importante mencionar que si en sus primeras diferencias la serie continua siendo no estacionaria, nos veremos en la necesidad de aplicar las segundas diferencias, hasta poder convertir la serie a estacionaria, puesto que, es una de las principales condiciones para trabajar con un modelo autorregresivo integrado de medias móviles.

Primeras diferencias del PIB de México

Para obtener las primeras diferencias de la serie hacemos uso de la función `diff()`, que para guardar nuestro resultado creamos la variables **d1PIB**, que me recoge las primeras diferencias de la serie.

```
# creamos un nuevo objeto para almacenar los resultados
d1PIB <- diff(PIB_ts)
# Ahora, gráficamos las primeras diferencias del PIB
plot(d1PIB, main = "Primeras diferencias del PIB de México",
     ylab = "PIB", col = "Blue")
```



Al parecer la serie ya oscila alrededor de una media. Como se puede apreciar, si se compara con el gráfico anterior de la serie original, la serie transformada tiene una media alrededor del 0 y varianza constante, obviamente, tiene errores, principalmente en el 2020 que se observa esa volatilidad en el segundo trimestre, pero dejando de lado eso, la serie parece ser de ruido blanco, falta determinar si no esta serialmente correlacionada. Pero, al menos por el momento, los resultados parecen ya relativamente satisfactorios, por lo que, procedemos a aplicar la prueba ADF para corroborar la estacionariedad de la serie.

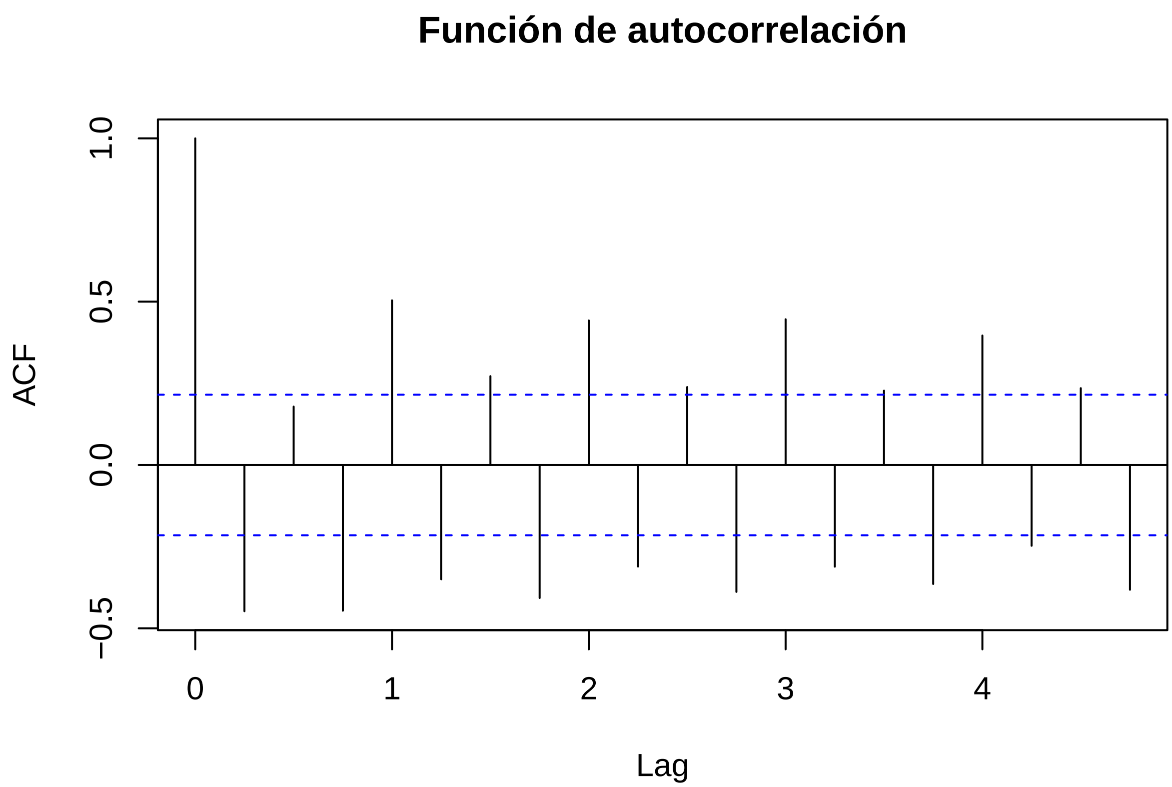
```
summary(ur.df(d1PIB))
```

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression none
```

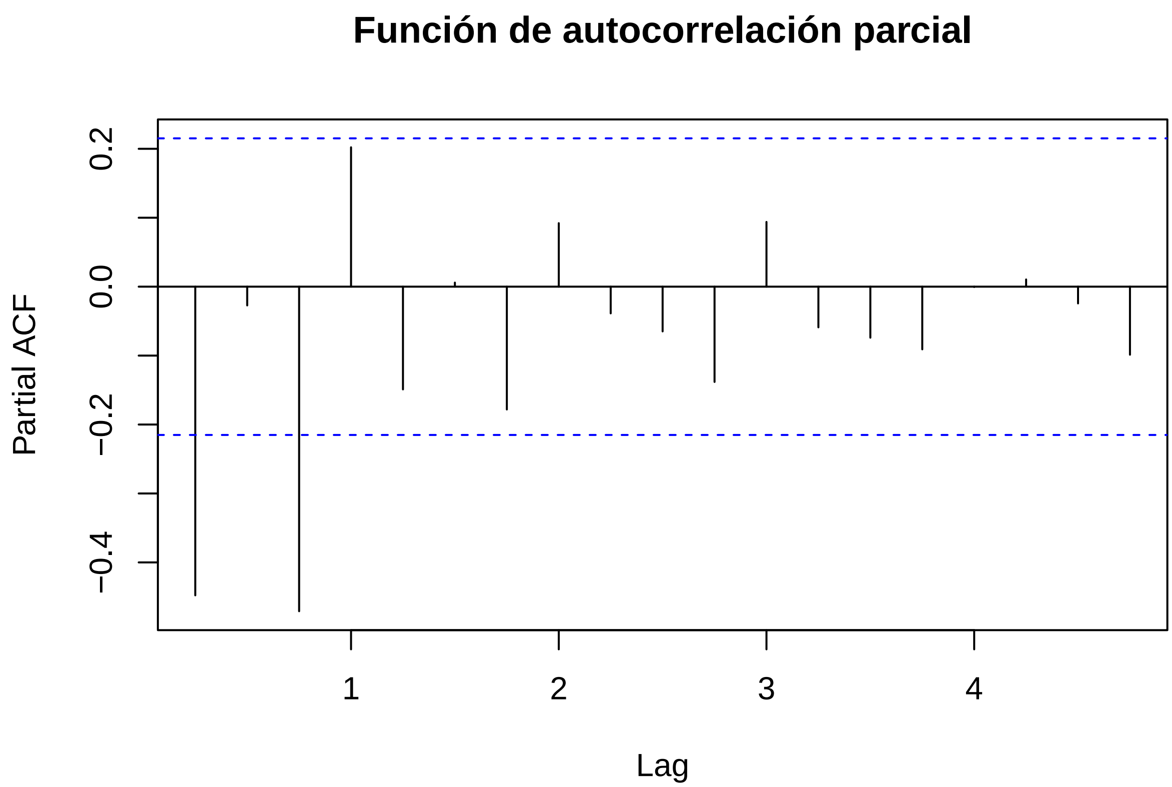
```
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3365766   -89038    95959   364422  1890584
##
## Coefficients:
##            Estimate Std. Error t value Pr(>|t|)
## z.lag.1    -1.57902    0.22224  -7.105  4.7e-10 ***
## z.diff.lag  0.07954    0.12926   0.615     0.54
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 588700 on 79 degrees of freedom
## Multiple R-squared:  0.7186, Adjusted R-squared:  0.7115
## F-statistic: 100.9 on 2 and 79 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: -7.1051
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau1 -2.6 -1.95 -1.61
```

El estadístico ADF recoge un valor de -7.10, esto quiere decir que, manteniendo la misma hipótesis nula anterior, se rechaza la hipótesis nula, que por ende, se acepta la hipótesis nula y se concluye que las primeras diferencias de la serie son estacionarias. Es decir, es un proceso estocástico puramente aleatorio. Una vez que se ha logrado esta importante característica, estamos en condiciones de llevar a cabo la función de autocorrelación y la función de autocorrelación parcial.

```
# Para obtener la función de autocorrelación usamos la función acf()
# tal como se muestra en la siguiente linea de código y le insertamos en # el parámetro la serie difere
acf(d1PIB, main = "Función de autocorrelación")
```



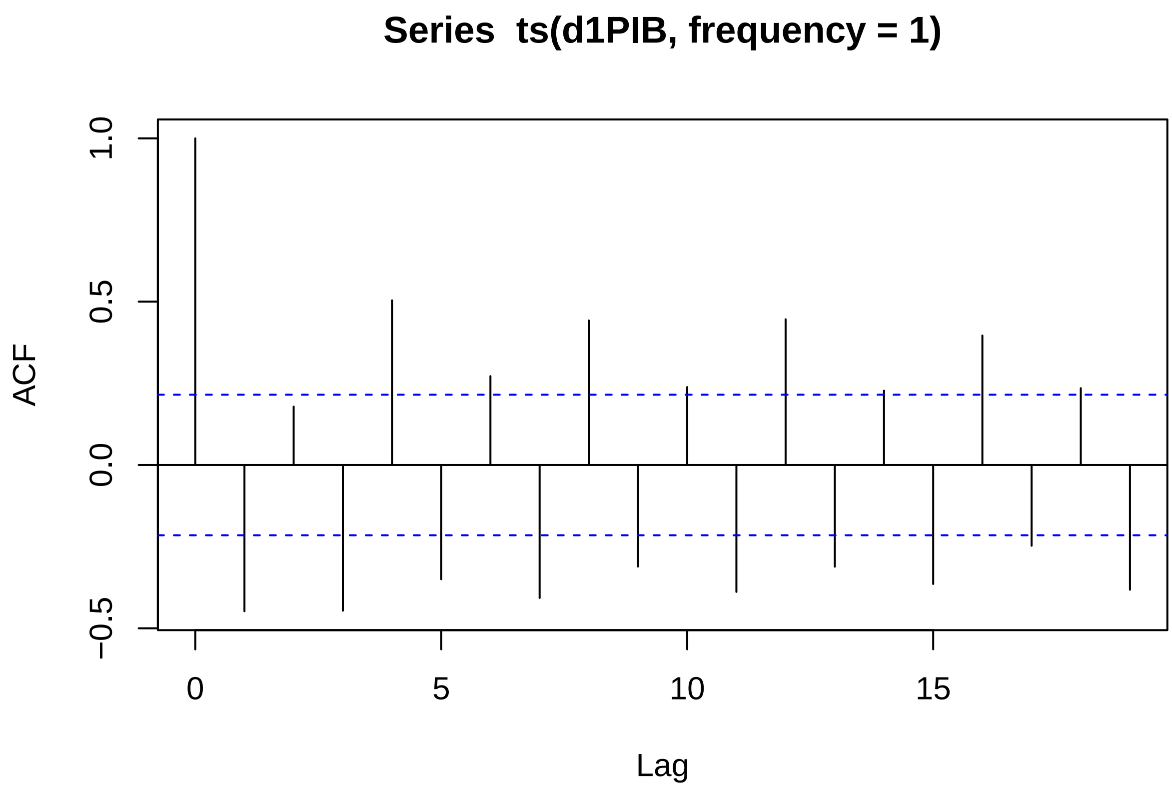

```
pacf(d1PIB, main = "Función de autocorrelación parcial")
```



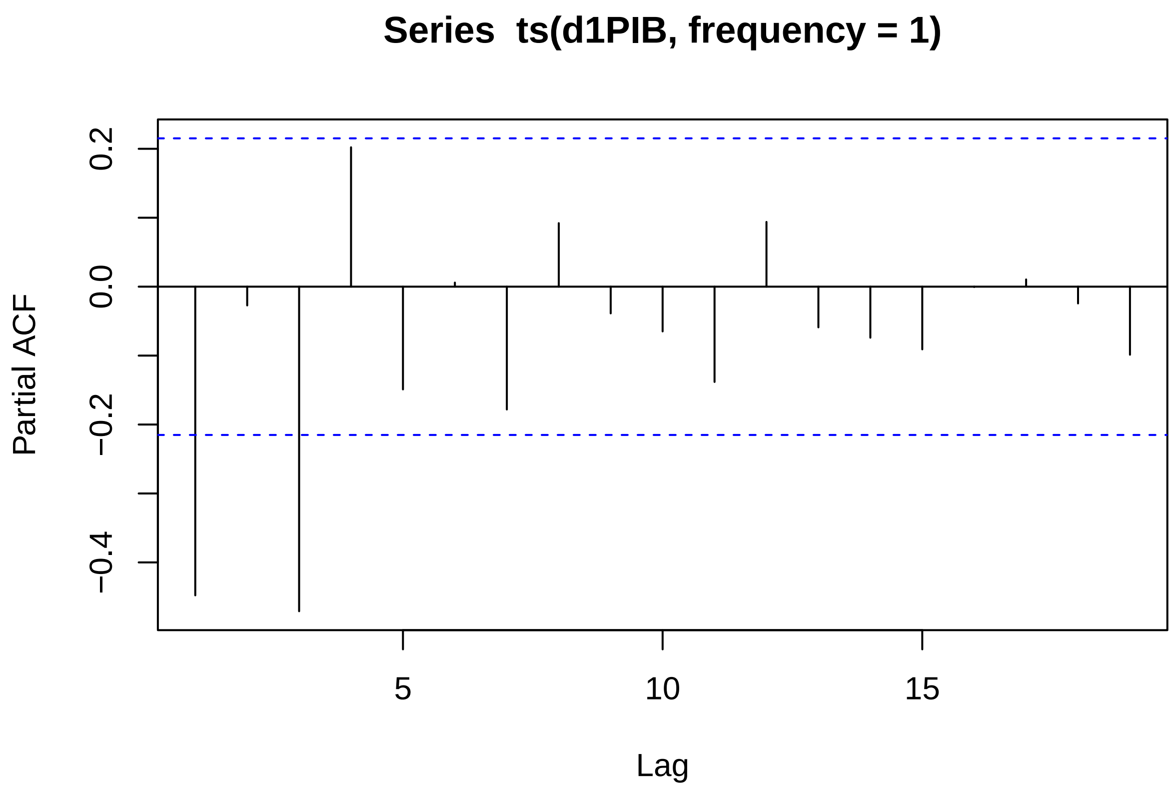
Estas dos funciones nos sirven para saber cuantas medias móviles y cuantos autorregresivos vamos a utilizar en nuestro modelo ARIMA. La función de autocorrelación nos indica el numero de medias móviles que necesita el modelo. Por el contrario, la función de autocorrelación parcial nos indica el numero de autorregresivos que se necesitan para estimar el modelo.

Sin embargo, es necesario previo a la estimación ajustar las funciones de autocorrelación de forma que el numero de rezagos coincida con las frecuencias. Para ello, en R ejecutamos nuevamente las funciones pero agregamos el parámetro de frequency = 1. Por defecto el lenguaje te arroja 20 rezagos.

```
acf(ts(d1PIB, frequency = 1))
```




```
pacf(ts(d1PIB, frequency = 1))
```



De acuerdo a la función de autocorrelación el numero de medias móviles debería ser un aproximado a 16, por su parte, la función de autocorrelación parcial nos indica un posible de 2 términos autorregresivos. Por consiguiente, llevaremos a cabo la primer estimación con dos términos autorregresivos, una diferencia y 16 medias móviles, que al menos por el momento, parece los mas adecuado al modelo.

Estimación del modelo ARIMA

Para poder estimar el modelo ARIMA utilizamos la función `arima()` de la paquetería stats. Guardamos el resultado en el objeto llamado "modelo1", tal como se muestra en el comando. Es importante mencionar que uno de los parámetros de las función es la introducción de la serie a explicar, sin embargo, aquí introducimos la serie original, no la serie diferenciada. El modelo ARIMA se basa en series no estacionarias, por lo que, es necesario introducir la serie original (no estacionaria) para poder estimar el modelo. En el parámetro order =, introducimos el numero de autorregresivos, diferencias y medias móviles.

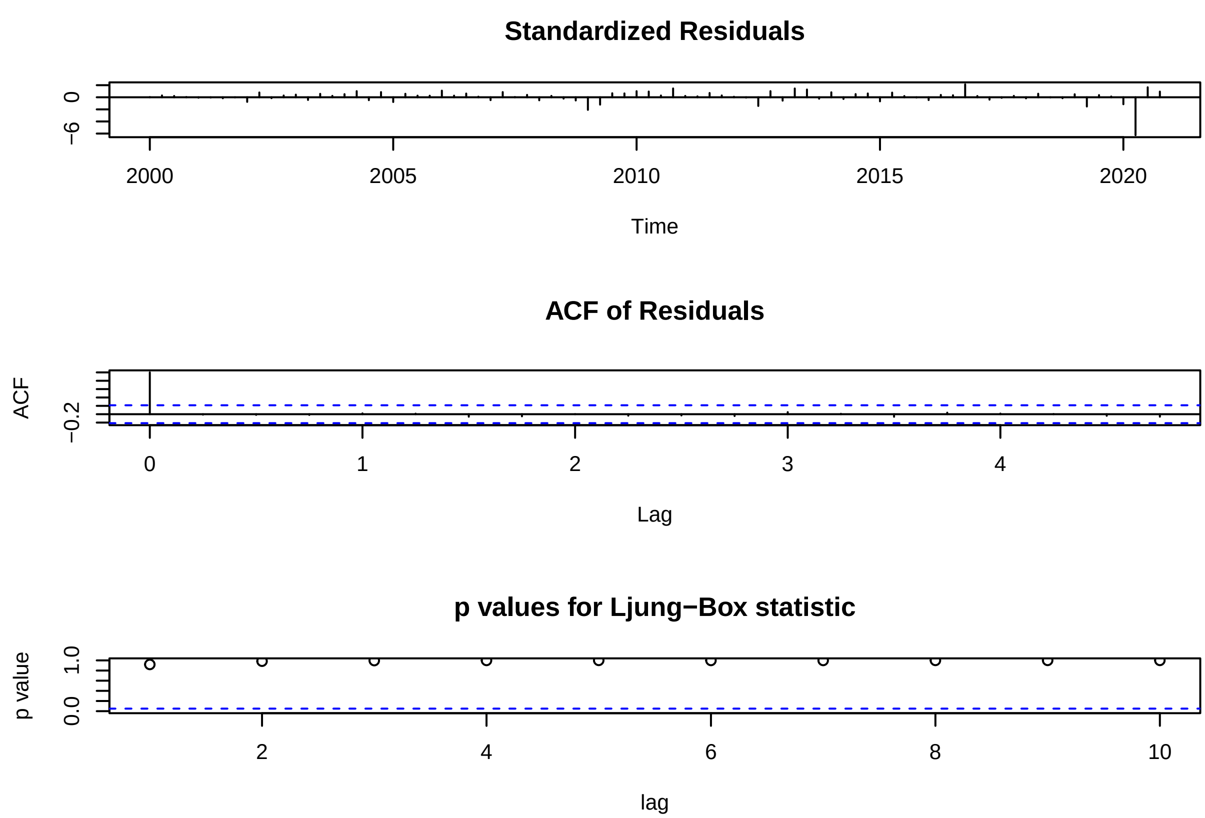
```r
modelo1 <- arima(PIB_ts, order = c(2,1,16))
modelo1
```

```
##
## Call:
## arima(x = PIB_ts, order = c(2, 1, 16))
##
## Coefficients:
##           ar1      ar2     ma1      ma2      ma3     ma4      ma5     ma6
##       -0.4136  -0.2128  0.2047  -0.0689  -0.4454  0.8213  -0.2370  0.2647
## s.e.   0.2128   0.2352  0.2258   0.2141   0.1635  0.1683   0.2264  0.2041
```

```
##          ma7      ma8      ma9     ma10     ma11     ma12     ma13     ma14
##      -0.2762   0.7040  -0.3554   0.2395  -0.3120   0.6483  -0.5877  -0.3478
## s.e.  0.2334   0.1635   0.2132   0.1826   0.2097   0.2084   0.1775   0.1822
##         ma15     ma16
##      -0.2042   0.7458
## s.e.  0.1805   0.2128
##
## sigma^2 estimated as 1.345e+11:  log likelihood = -1194.79,  aic = 2427.59
```

En la parte superior se puede observar los coeficientes estimados, tanto de los autorregresivos como el de los 16 medias móviles junto con sus errores estándar. Ahora, como paso siguiente es evaluar el funcionamiento del modelo, es decir, si el modelo cumple con el supuesto de ser ruido blanco.

```
tsdiag(modelo1)
```



En el gráfico anterior se puede observar la evaluación del modelo, los residuos estandarizados, la función de autocorrelación y las probabilidades del estadístico de Ljung-Box. Los errores estandarizados deben parecerse mucho al ruido blanco, es decir, con media cero, varianza constante y no deben estar autocorrelacionados. No obstante, otra posible alternativa para evaluar el ruido blanco del modelo son las probabilidades asociadas al estadístico Ljung-Box, que como se aprecia en el gráfico están por encima del 5% de la banda de confianza. De acuerdo al análisis gráfico, el modelo parece comportarse como ruido blanco, no obstante, es necesario corroborar esto con la prueba de Box, que en R la podemos aplicar con la función `Box.test()`.

```
# Primero generamos los residuos del modelo
residuals <- modelo1$residuals
Box.test(residuals, type = "Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  residuals
## X-squared = 0.010142, df = 1, p-value = 0.9198
```

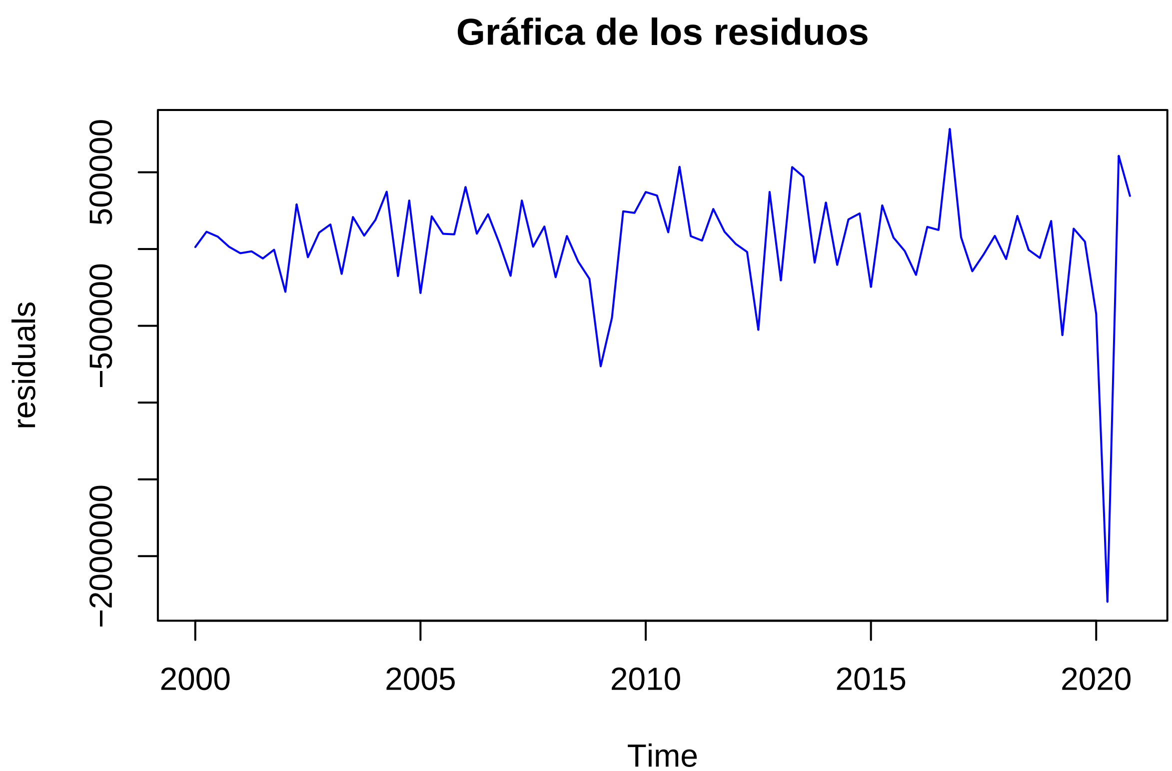
Se plantea la siguiente hipótesis nula:

Ho: Los residuos del modelo son de ruido blanco

Ha: los residuos del modelo no son de ruido blanco

La regla de decisión, como se habrá dado cuenta el lector depende de la probabilidad asociada al estadístico, que recoge un p-value de 0.91. Si la probabilidad asociada es mayor al 5% (0.05), no se rechaza la hipótesis nula y se acepta que los residuos del modelo son ruido blanco. Ahora, basándonos en la prueba estadística de Ljung-Box, y al ver que el p-value es mayor a 0.05, no se rechaza la hipótesis nula y se verifica que el modelo presenta ruido blanco, es decir, cumple con la teoría establecida por el modelo autorregresivo integrado de medias móviles. En el codigo siguiente se puede apreciar el gráfico de los errores del modelo, con media 0, varianza constante y no es serialmente autocorrelacionado.

```
plot(residuals, main = "Gráfica de los residuos", col = "blue")
```



Una vez evaluado el modelo, considerando que la ecuación de predicción aun se puede mejorar, los resultados alcanzados parecen ya relativamente satisfactorios para un ejercicio de este tipo. Por lo que, ya estamos en condiciones de pasar a la etapa de predicción, que es una de las principales funciones de los modelos de series temporales.

Predicción del modelo ARIMA seleccionado

Para llevar a cabo la predicción del modelo utilizaremos la función `forecast()`(predicción), el resultado del pronostico lo guardaremos en un objeto llamado "predicción". Le indicamos a R que nos pronostique 4 valores hacia adelante, es decir, los próximos 4 trimestres correspondientes a los del año 2021.

```
# Como siempre primero activamos la librería correspondiente
library(forecast)
predicción <- forecast(modelo1, h = 4)
predicción
```

```
##         Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 2021 Q1       18242559 17743467 18741651 17479264 19005854
## 2021 Q2       15346620 14713753 15979488 14378733 16314508
## 2021 Q3       17454980 16757372 18152589 16388080 18521881
## 2021 Q4       16611422 15902452 17320393 15527145 17695700
```

Para el primer trimestre de 2021 se proyecta un crecimiento económico de 18,242,559 millones de pesos a precios de 2013, para el segundo trimestre el pronostico recoge un valor de 15,346,620 millones de pesos a precios de 2013, para el tercer trimestre se espera un valor de 17,454,980 millones de pesos a precios de 2013 y para el ultimo trimestre del año se proyecta una cantidad de 16,611,422 millones de pesos a precios de 2013 de acuerdo al modelo estimado. También el programa nos arroja intervalos de confianza de un 80% y 95%, entre los cuales el PIB podría caer.