

Modelado ARIMA para el PIB de Colombia (1960 - 2023)

Juan Andrés Torres Contreras

2025-03-17

```
library(FinTS)
```

```
## Loading required package: zoo
```

```
##
```

```
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      as.Date, as.Date.numeric
```

```
library(forecast)
```

```
## Registered S3 method overwritten by 'quantmod':
```

```
##      method      from
```

```
##      as.zoo.data.frame zoo
```

```
##
```

```
## Attaching package: 'forecast'
```

```
## The following object is masked from 'package:FinTS':
```

```
##
```

```
##      Acf
```

```
library(TSA)
```

```
## Registered S3 methods overwritten by 'TSA':
```

```
##      method      from
```

```
##      fitted.Arima forecast
```

```
##      plot.Arima   forecast
```

```
##
```

```
## Attaching package: 'TSA'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      acf, arima
```

```
## The following object is masked from 'package:utils':
##
##      tar

library(tseries)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##      filter, lag

## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union

library(lmtest)
library(MASS)

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##      select
```

Contexto inicial

Este análisis se basa en la serie temporal del Producto Interno Bruto (PIB) de Colombia, medida en dólares estadounidenses a precios constantes de 2015. Los datos fueron extraídos del Banco Mundial y abarcan el período de 1960 a 2023.

Fuente de los datos

Los datos provienen de las cuentas nacionales del Banco Mundial y de los archivos de datos de Cuentas Nacionales de la OCDE.

Unidades de medida

El PIB está expresado en dólares estadounidenses constantes de 2015, lo que permite eliminar los efectos de la inflación y facilita comparaciones a lo largo del tiempo. La metodología utilizada convierte las cifras de PIB a partir de monedas nacionales utilizando los tipos de cambio oficiales de 2015. En algunos casos, se aplica un factor de conversión alternativo si el tipo de cambio oficial no refleja las transacciones reales de divisas.

Definición del indicador

Según la metadata, el PIB a precios de comprador se define como la suma del valor agregado bruto de todos los productores residentes en la economía, más los impuestos sobre los productos y menos los subsidios no incluidos en el valor de los productos. No se hacen deducciones por depreciación de activos ni por agotamiento de recursos naturales.

Serie sin Transformar

```
library(readr)

##
## Attaching package: 'readr'

## The following object is masked from 'package:TSA':
##
## spec

df <- read_csv("API_NY.GDP.MKTP.KD_DS2_es_csv_v2_8100.csv",
               skip = 4)

## New names:
## * '' -> '...69'

## Rows: 266 Columns: 69
## -- Column specification -----
## Delimiter: ","
## chr (4): Country Name, Country Code, Indicator Name, Indicator Code
## dbl (64): 1960, 1961, 1962, 1963, 1964, 1965, 1966, 1967, 1968, 1969, 1970, ...
## lgl (1): ...69
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.

print(df)

## # A tibble: 266 x 69
##   'Country Name'      'Country Code' 'Indicator Name' 'Indicator Code' '1960'
##   <chr>             <chr>         <chr>           <chr>           <dbl>
## 1 Aruba             ABW           PIB (US$ a prec~ NY.GDP.MKTP.KD    NA
## 2 <NA>              AFE           PIB (US$ a prec~ NY.GDP.MKTP.KD  1.52e11
## 3 Afganistán       AFG           PIB (US$ a prec~ NY.GDP.MKTP.KD    NA
## 4 <NA>              AFW           PIB (US$ a prec~ NY.GDP.MKTP.KD  1.08e11
## 5 Angola           AGO           PIB (US$ a prec~ NY.GDP.MKTP.KD    NA
## 6 Albania           ALB           PIB (US$ a prec~ NY.GDP.MKTP.KD    NA
## 7 Andorra           AND           PIB (US$ a prec~ NY.GDP.MKTP.KD    NA
## 8 El mundo árabe    ARB           PIB (US$ a prec~ NY.GDP.MKTP.KD    NA
## 9 Emiratos Árabes Un~ ARE           PIB (US$ a prec~ NY.GDP.MKTP.KD    NA
## 10 Argentina        ARG           PIB (US$ a prec~ NY.GDP.MKTP.KD  1.51e11
```

```
## # i 256 more rows
## # i 64 more variables: '1961' <dbl>, '1962' <dbl>, '1963' <dbl>, '1964' <dbl>,
## #   '1965' <dbl>, '1966' <dbl>, '1967' <dbl>, '1968' <dbl>, '1969' <dbl>,
## #   '1970' <dbl>, '1971' <dbl>, '1972' <dbl>, '1973' <dbl>, '1974' <dbl>,
## #   '1975' <dbl>, '1976' <dbl>, '1977' <dbl>, '1978' <dbl>, '1979' <dbl>,
## #   '1980' <dbl>, '1981' <dbl>, '1982' <dbl>, '1983' <dbl>, '1984' <dbl>,
## #   '1985' <dbl>, '1986' <dbl>, '1987' <dbl>, '1988' <dbl>, '1989' <dbl>, ...
```

```
# Filtramos solo los datos de Colombia
```

```
df_colombia <- df %>% filter(`Country Code` == "COL")
print(df_colombia)
```

```
## # A tibble: 1 x 69
##   'Country Name' 'Country Code' 'Indicator Name'      'Indicator Code' '1960'
##   <chr>          <chr>          <chr>          <chr>          <dbl>
## 1 Colombia      COL          PIB (US$ a precios con~ NY.GDP.MKTP.KD  3.06e10
## # i 64 more variables: '1961' <dbl>, '1962' <dbl>, '1963' <dbl>, '1964' <dbl>,
## #   '1965' <dbl>, '1966' <dbl>, '1967' <dbl>, '1968' <dbl>, '1969' <dbl>,
## #   '1970' <dbl>, '1971' <dbl>, '1972' <dbl>, '1973' <dbl>, '1974' <dbl>,
## #   '1975' <dbl>, '1976' <dbl>, '1977' <dbl>, '1978' <dbl>, '1979' <dbl>,
## #   '1980' <dbl>, '1981' <dbl>, '1982' <dbl>, '1983' <dbl>, '1984' <dbl>,
## #   '1985' <dbl>, '1986' <dbl>, '1987' <dbl>, '1988' <dbl>, '1989' <dbl>,
## #   '1990' <dbl>, '1991' <dbl>, '1992' <dbl>, '1993' <dbl>, '1994' <dbl>, ...
```

```
# Seleccionamos solo los datos de la serie
```

```
df_colombia <- df_colombia %>% dplyr::select(-`Country Name`, -`Country Code`, -`Indicator Name`, -`Ind
```

```
# Asegurar que df_colombia contiene solo los valores de los años
```

```
valores_pib <- as.numeric(unlist(df_colombia))
```

```
# Crear la serie temporal desde 1960 hasta 2003 con frecuencia anual
```

```
ts_colombia <- ts(valores_pib, start = 1960, end = 2023, frequency = 1)
```

```
# Verificar que la serie temporal se creó correctamente
```

```
print(ts_colombia)
```

```
## Time Series:
```

```
## Start = 1960
```

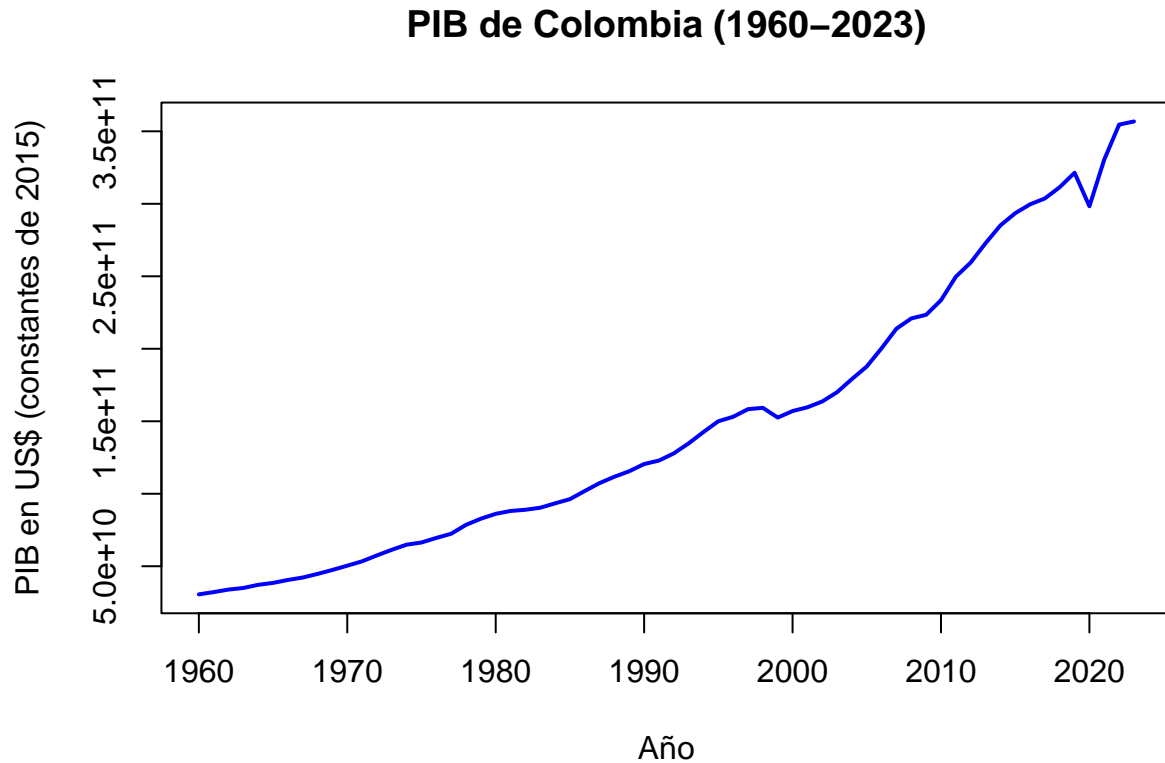
```
## End = 2023
```

```
## Frequency = 1
```

```
## [1] 30607351921 32165026473 33905762200 35020145448 37179817269
## [6] 38518294925 40535396289 42211479865 44715936774 47444734556
## [11] 50390007346 53393618596 57488453985 61353376859 64878592920
## [16] 66385888257 69524857841 72416052846 78549620713 82774898276
## [21] 86158201123 88119913722 88955661301 90355784303 93383320875
## [26] 96284839824 101892542204 107363083659 111726505380 115540955654
## [31] 120488417446 122900122762 127871346008 134757742185 142593449287
## [36] 150011784498 153095808839 158347444691 159249683235 152554802278
## [41] 157016818929 159651401477 163649041384 170061255793 179130660093
## [46] 187780451739 200393418124 213896316785 220919487241 223437193185
## [51] 233479933032 249701866578 259471791121 272793056064 285066097497
## [56] 293492370193 299618678572 303691577642 311479214512 321405606656
## [61] 298309675728 330530695025 354622693595 356787344439
```

```
# Graficamos la serie para visualizarla
```

```
plot(ts_colombia, main = "PIB de Colombia (1960-2023)", ylab = "PIB en US$ (constantes de 2015)", xlab = "Año")
```



La serie muestra una clara tendencia creciente en el PIB de Colombia desde 1960 hasta 2023. A simple vista, se puede observar un crecimiento acelerado a partir de los años 2000, con ciertas desaceleraciones y pequeñas caídas en algunos periodos.

Tendencias y anomalías

- Desde 1960 hasta finales de los 90, el PIB crece de forma sostenida con ligeras variaciones.
- A partir del 2000, el crecimiento se vuelve más pronunciado.
- Se observa una caída notable alrededor de 2020, probablemente asociada al impacto económico de la pandemia de COVID-19, seguida de una rápida recuperación en los años posteriores.

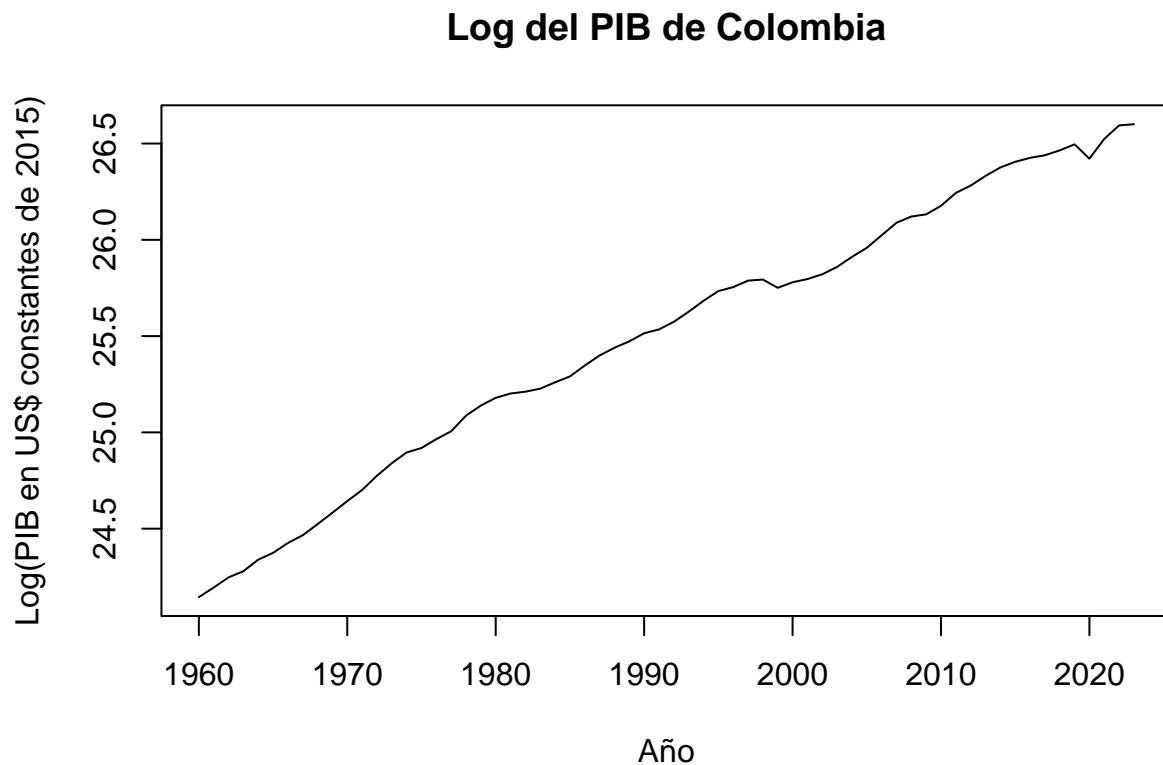
Estacionalidad y estacionariedad

- No se observa estacionalidad evidente, lo que es esperable en una serie de PIB anual.
- La serie no es estacionaria, ya que presenta una tendencia clara de crecimiento en el tiempo. Para aplicar modelos ARMA o ARIMA, será necesario transformar la serie para hacerla estacionaria.

Transformación logarítmica y diferenciación de la serie

```
# Aplicamos logaritmo natural a la serie
log_ts_colombia <- log(ts_colombia)

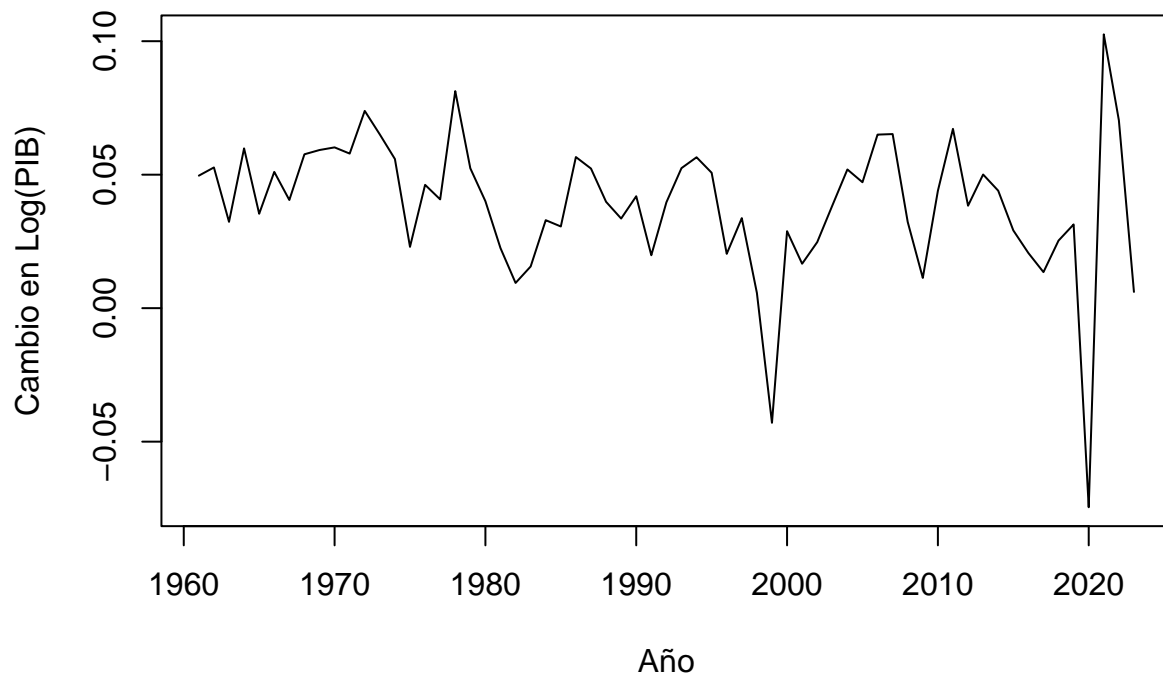
# Graficamos la serie transformada
plot(log_ts_colombia, main = "Log del PIB de Colombia",
      ylab = "Log(PIB en US$ constantes de 2015)", xlab = "Año")
```



```
# Diferenciamos la serie logarítmica para hacerla estacionaria
diff_log_ts_colombia <- diff(log_ts_colombia)

# Graficamos la serie diferenciada
plot(diff_log_ts_colombia, main = "Diferencia del Log(PIB) de Colombia",
      ylab = "Cambio en Log(PIB)", xlab = "Año")
```

Diferencia del Log(PIB) de Colombia



```
# Verificamos estacionariedad con prueba de Dickey-Fuller  
adf.test(diff_log_ts_colombia)
```

```
##  
## Augmented Dickey-Fuller Test  
##  
## data: diff_log_ts_colombia  
## Dickey-Fuller = -3.8133, Lag order = 3, p-value = 0.02368  
## alternative hypothesis: stationary
```

Para estabilizar la varianza y facilitar el modelado, se aplicó el logaritmo natural a la serie de tiempo del PIB. Posteriormente, se realizó una diferenciación de primer orden para eliminar la tendencia y hacer la serie estacionaria.

Prueba de Dickey-Fuller Aumentada (ADF)

La prueba ADF se utilizó para evaluar la estacionariedad de la serie diferenciada. Esta prueba contrasta la hipótesis nula de que la serie tiene una raíz unitaria (es decir, no es estacionaria) contra la alternativa de que es estacionaria.

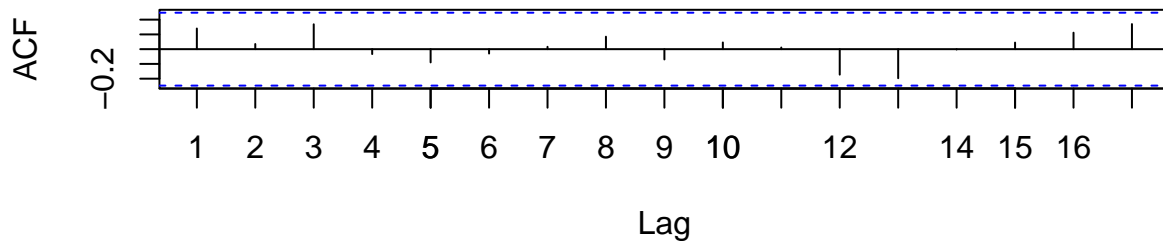
Resultado de la prueba: Dado que el **p-valor es 0.02368**, menor al umbral típico de **0.05**, se rechaza la hipótesis nula, indicando que la serie diferenciada en logaritmos es estacionaria.

```
# Graficamos ACF y PACF de la serie diferenciada en logaritmos

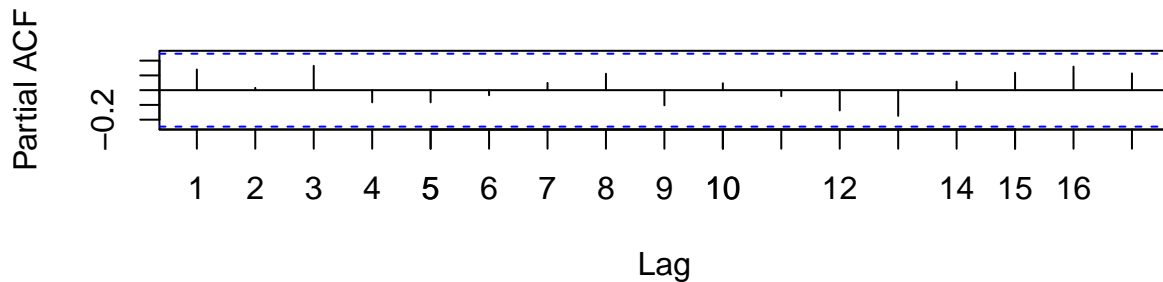
par(mfrow = c(2,1)) # Para mostrar ambas gráficas juntas

acf(diff_log_ts_colombia, main = "ACF del PIB Diferenciado en Log")
axis(1, at = seq(0, 30, by = 1), labels = seq(0, 30, by = 1))
pacf(diff_log_ts_colombia, main = "PACF del PIB Diferenciado en Log")
axis(1, at = seq(0, 30, by = 1), labels = seq(0, 30, by = 1))
```

ACF del PIB Diferenciado en Log



PACF del PIB Diferenciado en Log



De acuerdo a las gráficas ACF y PACF, no existe una autocorrelación significativa bajo ningún rezago. Esto puede ser debido a la influencia de los valores atípicos de la serie en la pandemia.

Para mitigar el impacto de valores atípicos en la identificación de p y q , aplicaremos un preprocesamiento temporal a la serie diferenciada del logaritmo del PIB de Colombia, utilizando `tsoutliers`, que permitirá identificar posibles observaciones extremas que podrían distorsionar el análisis. Luego, se aplicará `tsclean`, para corregir la serie, reemplazando estos valores con estimaciones interpoladas y suavizando su comportamiento sin alterar la estructura subyacente.

```
outliers <- tsoutliers(diff_log_ts_colombia)
ts_colombia_clean <- tsclean(diff_log_ts_colombia)

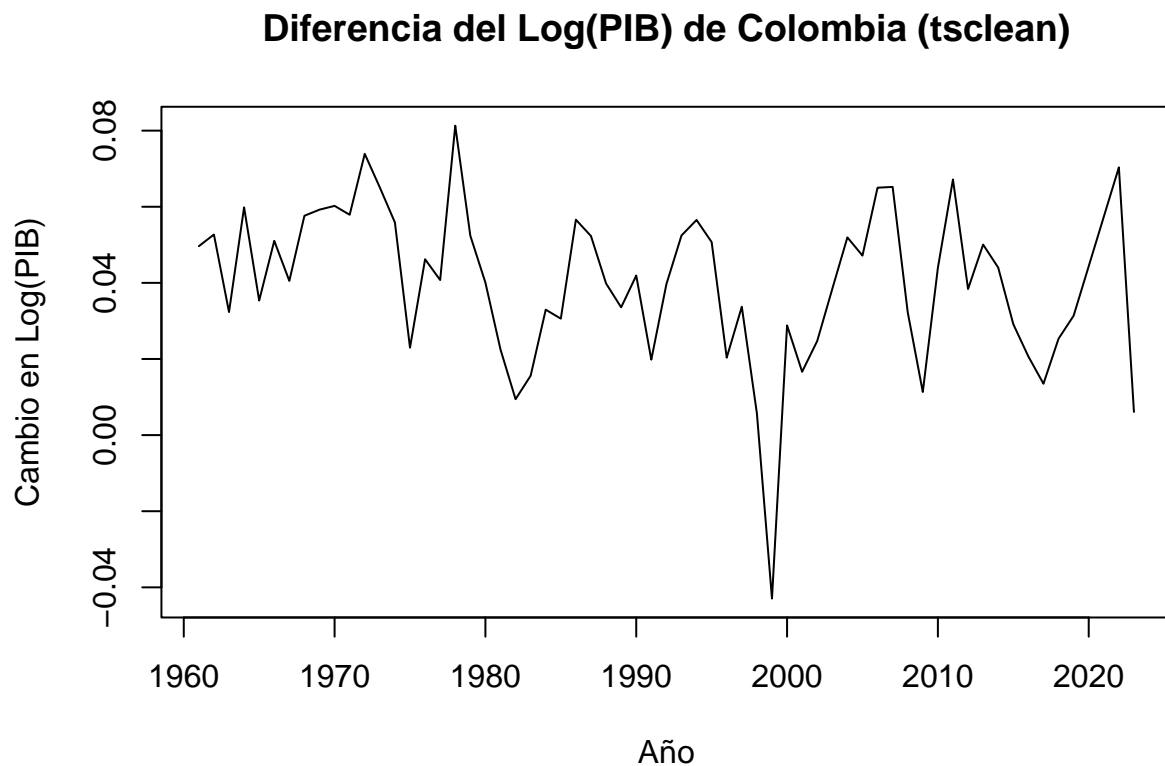
print(outliers)
```

```
## $index
## [1] 60 61
##
```



```
## $replacements
## [1] 0.04436581 0.05736034
```

```
# Graficamos la serie diferenciada
plot(ts_colombia_clean, main = "Diferencia del Log(PIB) de Colombia (tsclean)",
     ylab = "Cambio en Log(PIB)", xlab = "Año")
```



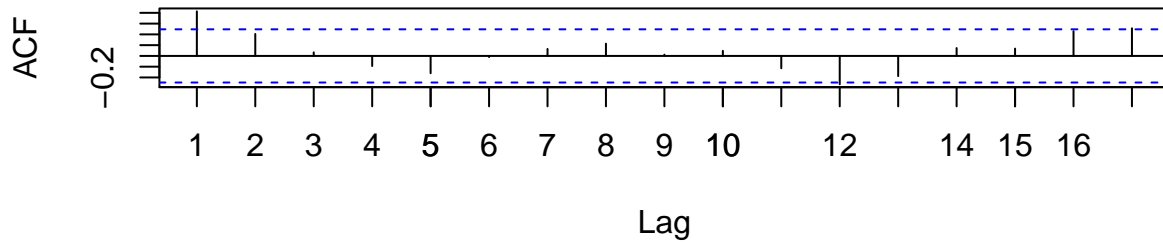
Se observa que los datos que cambian son solamente los de 2020 y 2021, por lo que la corrección de los datos atípicos generados en la pandemia es exitosa.

```
# Graficamos ACF y PACF de la serie diferenciada en logaritmos

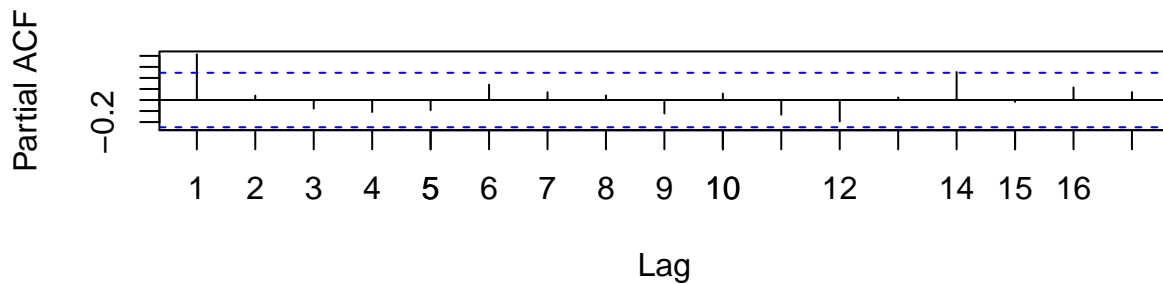
par(mfrow = c(2,1)) # Para mostrar ambas gráficas juntas

acf(ts_colombia_clean, main = "ACF del PIB Diferenciado en Log (tsclean)")
axis(1, at = seq(0, 30, by = 1), labels = seq(0, 30, by = 1))
pacf(ts_colombia_clean, main = "PACF del PIB Diferenciado en Log (tsclean)")
axis(1, at = seq(0, 30, by = 1), labels = seq(0, 30, by = 1))
```

ACF del PIB Diferenciado en Log (tsclean)



PACF del PIB Diferenciado en Log (tsclean)



Tras aplicar la corrección, se identifica un modelo ARMA(1,1) como la mejor representación de la estructura temporal de la serie.

Modelo ARIMA(1,1,1)

```
modelo_arma <- Arima(log_ts_colombia, order = c(1,1,1))
summary(modelo_arma)
```

```
## Series: log_ts_colombia
## ARIMA(1,1,1)
##
## Coefficients:
##      ar1      ma1
##    0.9988 -0.9233
## s.e.  0.0026  0.0620
##
## sigma^2 = 0.0007317:  log likelihood = 138.07
## AIC=-270.14  AICc=-269.73  BIC=-263.71
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE
## Training set -0.001078681 0.02640878 0.01852064 -0.003846674 0.07223057
##              MASE      ACF1
## Training set 0.4336029 0.1021898
```

Para evaluar si este es el modelo más adecuado, compararemos sus criterios de información con los obtenidos mediante la función `auto.arima()`. Esto nos permitirá determinar si el modelo seleccionado manualmente ofrece un mejor ajuste en términos de AIC y BIC.

```
# Modelo óptimo según auto.arima()
modelo_auto <- auto.arima(log_ts_colombia)

summary(modelo_auto)

## Series: log_ts_colombia
## ARIMA(0,1,0) with drift
##
## Coefficients:
##      drift
##      0.0390
## s.e.  0.0033
##
## sigma^2 = 0.000703:  log likelihood = 140.23
## AIC=-276.46   AICc=-276.26   BIC=-272.17
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set 0.0003766486 0.02609667 0.01841923 0.00224527 0.07189262 0.4312286
##              ACF1
## Training set 0.1437973
```

El modelo seleccionado por `auto.arima()` fue un ARIMA(0,1,0) con drift, lo que indica que la mejor opción sin considerar regresores externos es un modelo de paseo aleatorio con tendencia. Esto se ve reflejado en los criterios de información, donde obtuvo un AIC de -276.46, mejor que el del modelo ARMA(1,1) considerado anteriormente.

Dado que este modelo no incluye términos AR ni MA, exploraremos la posibilidad de mejorar la especificación incorporando una variable externa que capture el impacto de la pandemia en la serie. Para ello, construiremos un modelo ARIMA(1,1,1) con un regresor indicador para el período posterior a 2020. Esto permitirá evaluar si la inclusión de esta variable mejora el ajuste del modelo y proporciona una mejor interpretación de los efectos de la pandemia en la serie.

Modelo ARIMA(1,1,1) con xreg

```
años <- as.numeric(colnames(df_colombia))

## Warning: NAs introduced by coercion

años <- na.omit(años)
pandemia <- ifelse(años >= 2020, 1, 0)

# Ajustamos modelo ARIMA con regresor externo
modelo_arma_xreg <- Arima(log(ts_colombia), order = c(1,1,1), xreg = pandemia)
summary(modelo_arma_xreg)
```

```
## Series: log(ts_colombia)
## Regression with ARIMA(1,1,1) errors
##
## Coefficients:
##          ar1          ma1          xreg
##          0.9235      -0.2264      -0.1350
## s.e.    0.0649      0.2551      0.0198
##
## sigma^2 = 0.0005181:  log likelihood = 150.3
## AIC=-292.59   AICc=-291.9   BIC=-284.02
##
## Training set error measures:
##              ME          RMSE          MAE          MPE          MAPE          MASE
## Training set 0.004196965 0.02204001 0.01731583 0.01671419 0.06769868 0.4053961
##              ACF1
## Training set 0.008249897
```

El modelo ARIMA(1,1,1) con regresor externo presenta una mejora significativa respecto al ARIMA(0,1,0), dado que su criterio de información AIC (-292.59) es menor que el del modelo sin regresores (-276.46). Esto indica que la inclusión de la variable externa ha permitido capturar mejor la dinámica de la serie, logrando un mejor balance entre ajuste y complejidad del modelo.

En cuanto a los coeficientes, el componente autoregresivo (AR1) es alto (0.9235), lo que sugiere una fuerte dependencia temporal en la serie. El término de media móvil (MA1) es negativo (-0.2264), pero con mayor incertidumbre, lo que indica una corrección más leve de los choques pasados. El regresor externo tiene un coeficiente de -0.1350, lo que sugiere que la variable incluida tiene un impacto significativo en la serie.

Las métricas de error también reflejan una mejora con respecto al modelo anterior. El RMSE (0.0220), el MAE (0.0173) y el MAPE (0.0677) son menores, lo que indica una mejor precisión en las predicciones.

Ecuación del modelo:

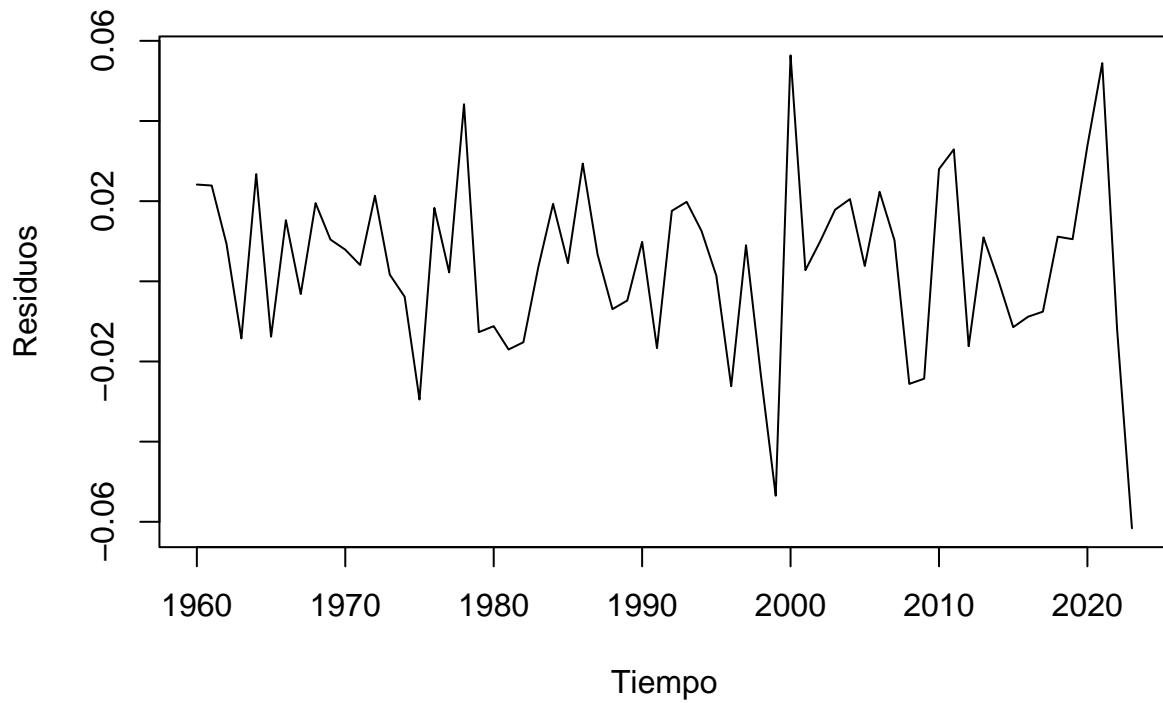
$$\nabla \log(y_t) = \phi_1 \nabla \log(y_{t-1}) + \theta_1 \varepsilon_{t-1} + \beta \cdot \text{pandemia} + \varepsilon_t$$

$$\nabla \log(y_t) = 0.9235 \nabla \log(y_{t-1}) - 0.2264 \varepsilon_{t-1} - 0.1350 \cdot \text{pandemia} + \varepsilon_t$$

Análisis de Residuos

```
residuals_arma <- residuals(modelo_arma_xreg)
plot(residuals_arma, main = "Residuos del Modelo ARMA(2,1)", ylab = "Residuos", xlab = "Tiempo")
```

Residuos del Modelo ARMA(2,1)

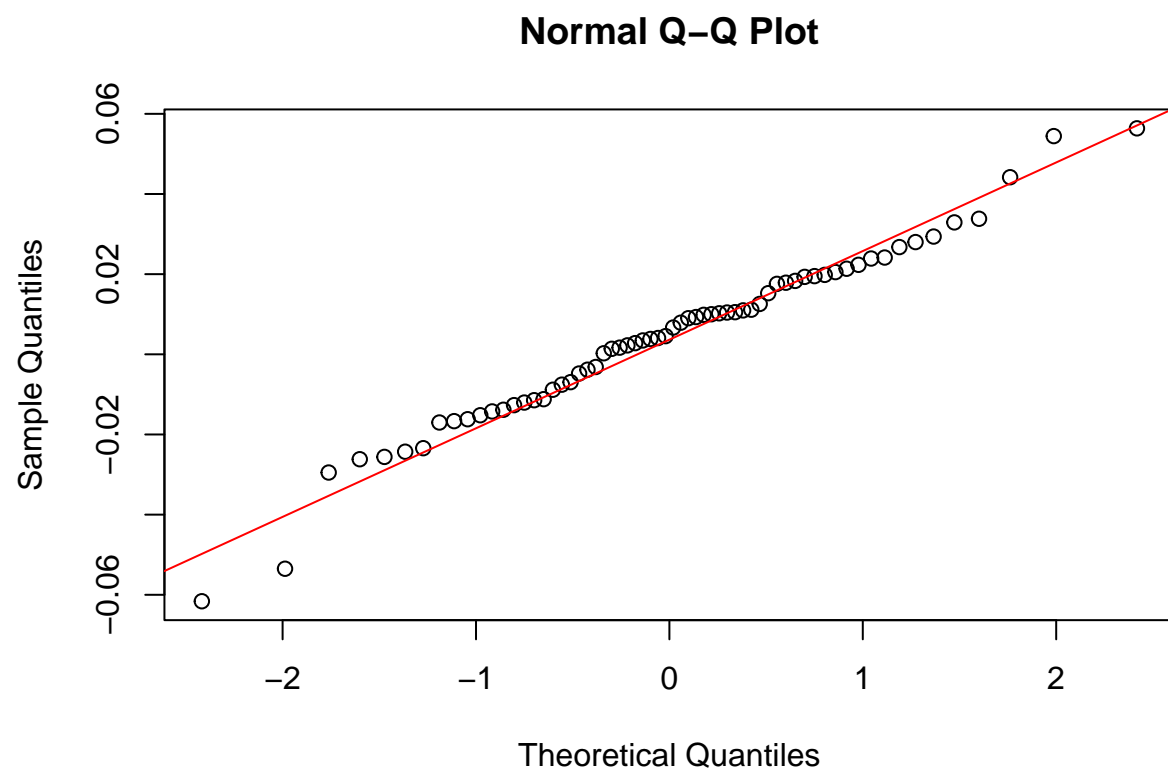


```
jarque.bera.test(residuals_arma)
```

Prueba de Normalidad: Jarque - Bera

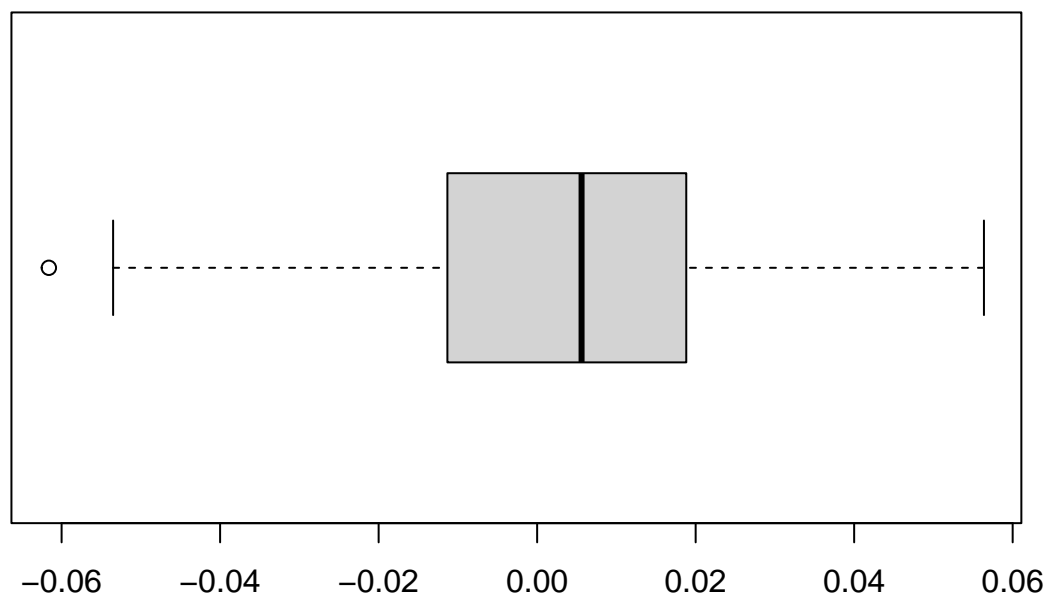
```
##  
## Jarque Bera Test  
##  
## data: residuals_arma  
## X-squared = 3.3279, df = 2, p-value = 0.1894
```

```
qqnorm(residuals_arma); qqline(residuals_arma, col = "red")
```



```
boxplot(residuals_arma, main = "Boxplot de Residuos", horizontal = TRUE)
```

Boxplot de Residuos



El p-valor de 0.1894 y el qqplot indican que no hay evidencia suficiente para rechazar la hipótesis nula de normalidad. Esto sugiere que los residuos transformados siguen una distribución normal.

```
ArchTest(residuals_arma, lags = 5)
```

Prueba de Homocedasticidad: ARCH

```
##  
## ARCH LM-test; Null hypothesis: no ARCH effects  
##  
## data: residuals_arma  
## Chi-squared = 5.2315, df = 5, p-value = 0.3883
```

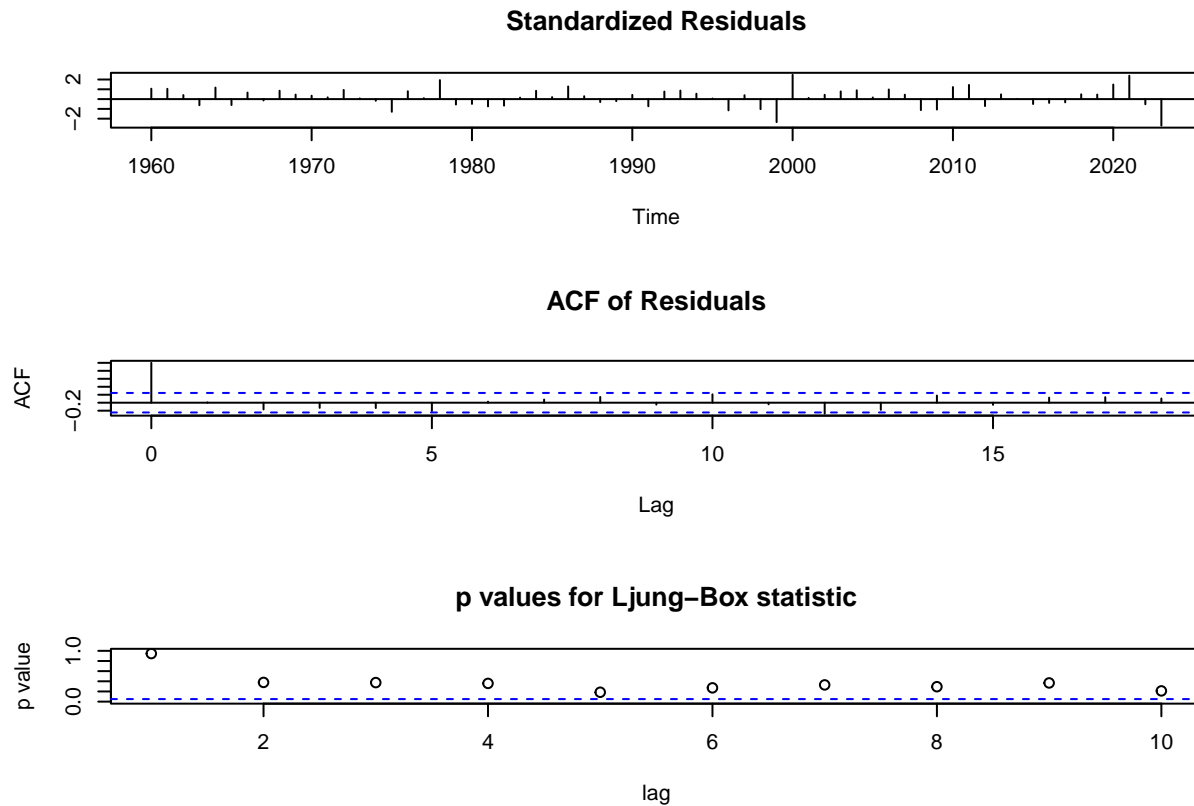
El p-valor de 0.3883 indica que no hay evidencia suficiente para rechazar la hipótesis nula, por lo que no se detectan efectos ARCH en los residuos. Esto sugiere que la varianza es constante en el tiempo.

```
Box.test(residuals_arma, lag = 10, type = "Ljung-Box")
```

Prueba de Correlacion Serial: Ljung-Box

```
##
## Box-Ljung test
##
## data: residuals_arma
## X-squared = 13.234, df = 10, p-value = 0.2109
```

```
tsdiag(modelo_arma_xreg)
```



El p-valor de 0.2109 indica que no hay evidencia suficiente para rechazar la hipótesis nula de ausencia de autocorrelación en los residuos, lo cual se confirma con las gráficas. Esto sugiere que los residuos no presentan una estructura de correlación significativa.

Conclusiones

El modelo final seleccionado para representar la dinámica del PIB de Colombia fue un ARIMA(1,1,1) con un regresor externo, el cual ofreció un mejor ajuste en comparación con el modelo sin regresores. La inclusión de una variable indicadora para la pandemia permitió capturar mejor las fluctuaciones en la serie, reflejándose en una reducción del AIC y en una mejora de las métricas de error.

Los coeficientes estimados muestran que el componente autoregresivo es fuerte, lo que indica una alta persistencia en los valores del PIB, mientras que el término de media móvil tiene un impacto menor en la serie. Además, el coeficiente del regresor externo resultó significativo, confirmando que la pandemia tuvo un efecto relevante en la evolución del PIB.

Las pruebas de diagnóstico confirmaron que los residuos cumplen con los supuestos de normalidad, homocedasticidad y ausencia de autocorrelación, validando la adecuación del modelo. Con este ajuste, se logra

una representación confiable de la evolución del PIB de Colombia en el periodo analizado.