

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA

MAT267

SERIES DE TIEMPO

PROYECTO FINAL

Estudiante:

Clemente Ferrer

Profesor:

Ronny Vallejos

2022-1

Índice


INTRODUCCIÓN	3
CONTEXTO Y OBJETIVOS	3
ANÁLISIS DESCRIPTIVO	3
MODELO INGENUO	4
MODELO ARIMA	5
ANÁLISIS ESPECTRAL DE LA SERIE	7
EXTRA: ARIMA ESTACIONAL	8
DIAGNÓSTICO DEL MODELO	8
PREDICCIÓN	9
DEPENDENCIA DE FACTORES EXTERNOS Y CORRELACIÓN	10
CONCLUSIÓN	11
ANEXO	12

INTRODUCCIÓN


El oro es un metal precioso y es completamente diferente de otros activos y metales. En particular, es muy sensible a los cambios de precios.

Por lo general, el rendimiento del oro está casi en línea con la tasa de inflación. Es por esto, que las personas que invierten en él tienen principalmente dos objetivos, proteger su patrimonio contra la inflación a lo largo del tiempo y mezclar su canasta de inversiones generando diversificación del riesgo y reducción parcial en volatilidad de la cartera.

CONTEXTO Y OBJETIVOS

A lo largo de este proyecto nos enfocaremos en predecir el precio de este activo a través de diferentes modelos vistos durante el curso *Series de Tiempo I* usando el software . Para ello, recurriremos a una base de datos de aproximadamente cuarenta años con mediciones diarias en un calendario anual de trading.

Como objetivos secundarios se planea realizar una correcta validación cruzada y posterior chequeo de los supuestos, como la normalidad de los residuos a través del test y gráficos asociados.

Los códigos serán subidos al siguiente repositorio público para cualquier revisión  FIS267 y serán adjuntados al anexo que se encontrará al final de este documento.

ANÁLISIS DESCRIPTIVO

Al visualizar el set de datos `gold_price_data.csv` vislumbramos que existen ciertos outliers de datos trimestrales desde el año 1970 a 1978, iniciando de forma diaria desde el año 1979 hasta marzo del 2020.

Esto fue resuelto eliminando los datos iniciales al tener una nula representatividad para los efectos de la predicción.

Además, el segundo percance derivó de la frecuencia con la cual se tomarían los datos. Dado que al tratarse de una base de datos de un activo financiero, no se encuentra información de los fines de semana y festivos al estar el mercado cerrado. Esto fue solucionado agregando de forma manual los días restantes y rellenando con el valor que cerró el día anterior. Por ejemplo: Si el viernes cerró con un precio de 1212,3 [USD], los días sábado y domingo fueron rellenados con este mismo valor.

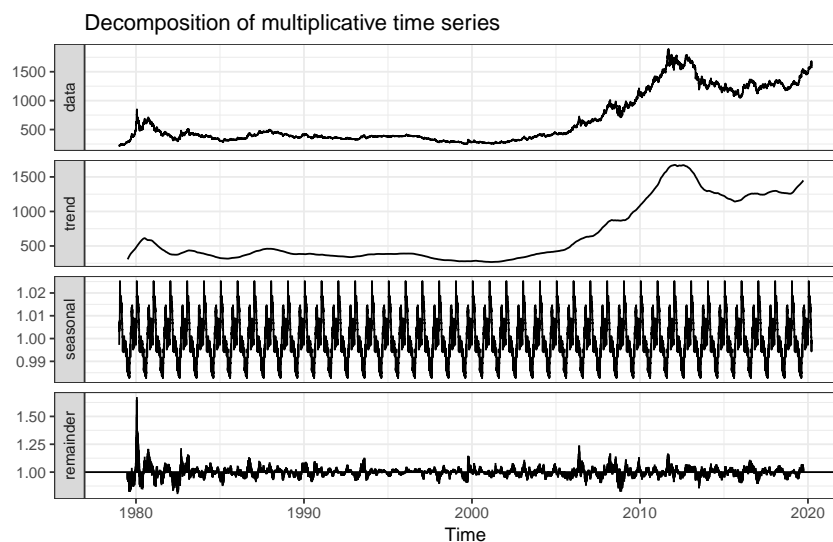
De esta manera, el gráfico de la serie de tiempo habiendo pasado el proceso de limpieza de los datos se presenta en la siguiente figura



Posterior a esto, se realizó una partición de los datos en dos submuestras: el conjunto de entrenamiento consta con el 95 % de los datos, mientras que el de testeo con el 5 % restante. A continuación los modelos fueron entrenados con el primero de ellos y la predicción entregada contrastada con el conjunto de prueba.

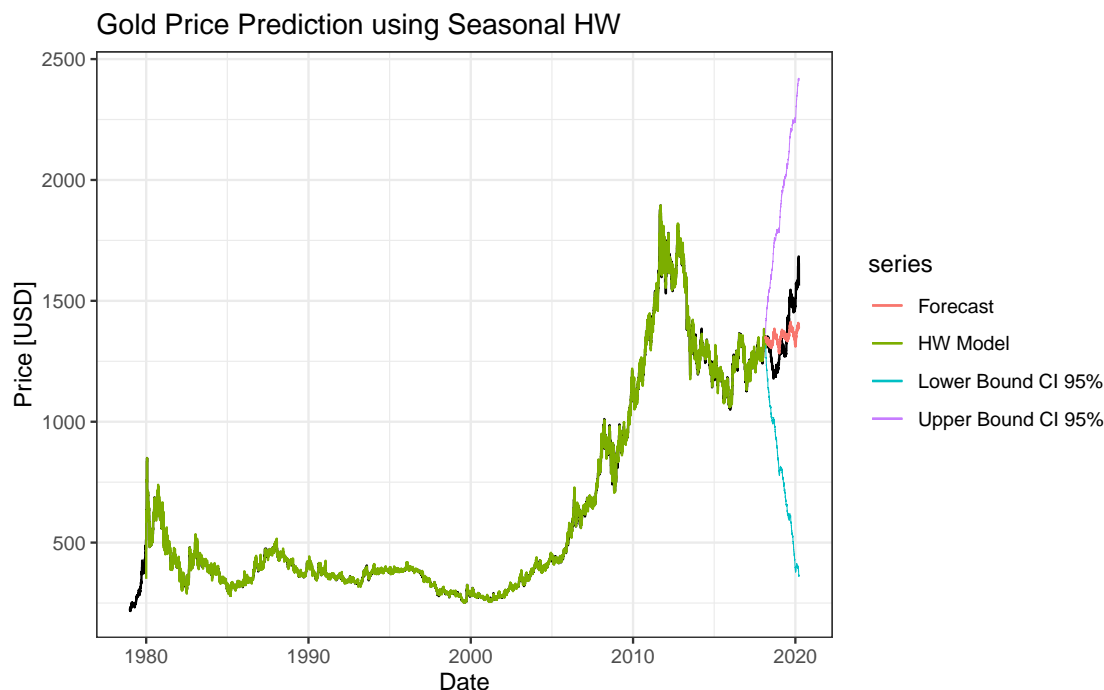
MODELO INGENUO

En primer lugar, vemos una clara tendencia en el gráfico, junto a una marcada componente estacional. Por ello, realizaremos una descomposición en sus partes tendenciales, estacionales y residuales usando un modelo multiplicativo por el crecimiento que presenta:



Esto ratifica nuestra observación preliminar. Por lo tanto, ajustaremos un modelo Holt-Winters con componente estacional multiplicativa y realizaremos la predicción con unas

bandas del 95 % de confianza.



donde su coeficiente R^2 y error cuadrático medio, RMSE vienen dados por

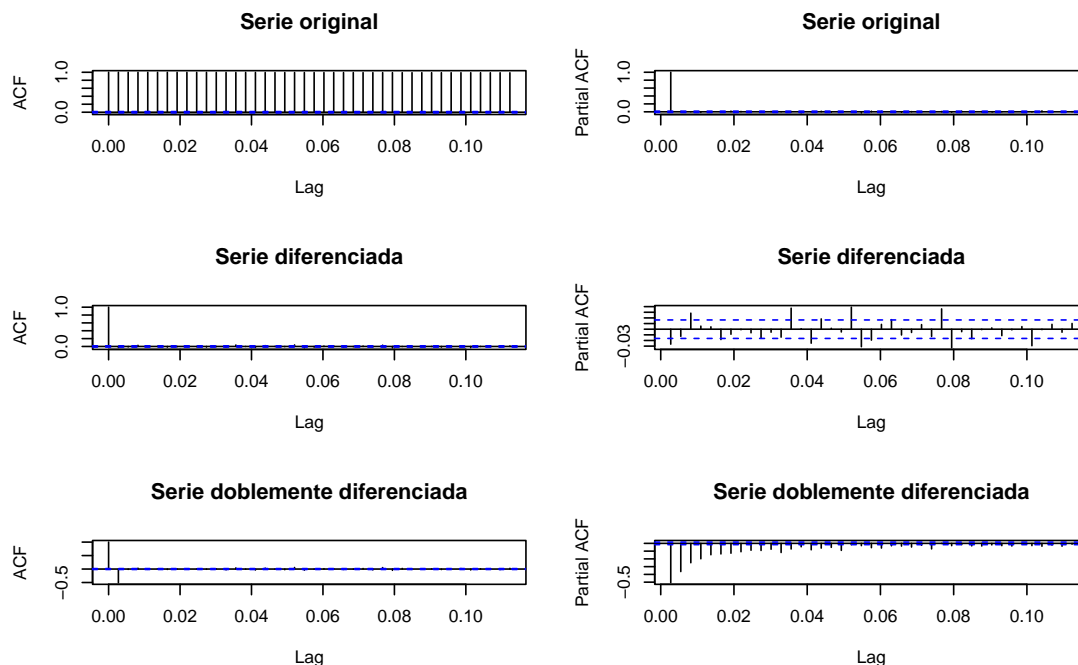
$$\text{RMSE HW-Seasonal} = 110.9195, \quad R^2 = 0.1877348$$

Notamos que la predicción es aceptable para ser un modelo ingenuo y logra captar la tendencia de la serie.

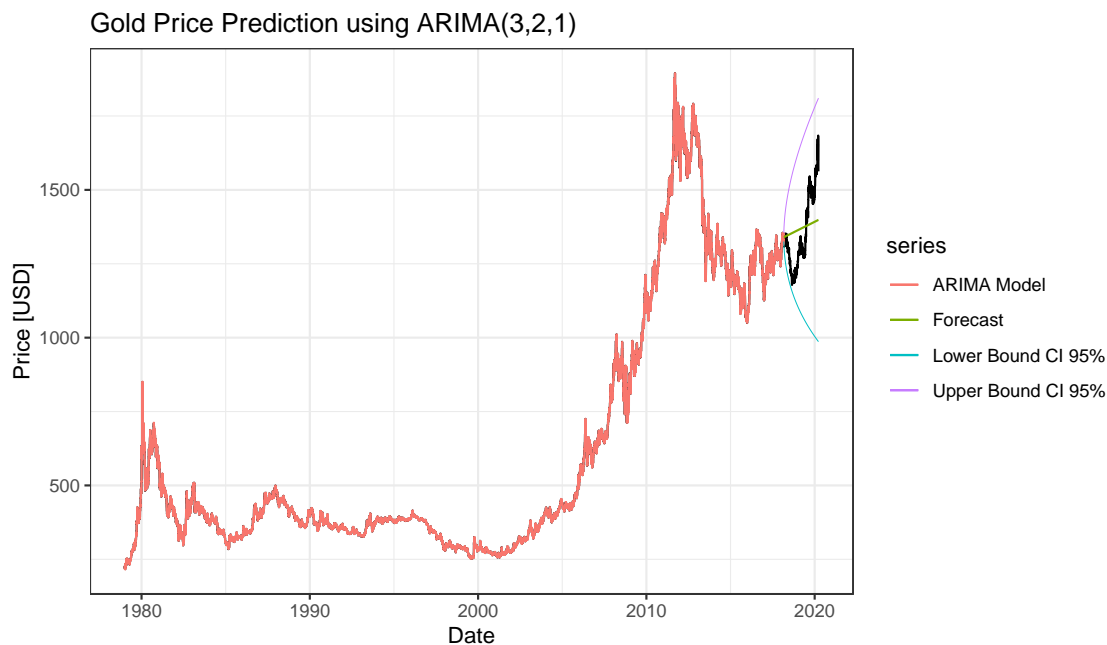
MODELO ARIMA

Ahora bien, nos adentraremos en la mitad del curso y veremos la factibilidad de ajustar un modelo ARIMA. Para ello, obtenemos la función de autocorrelación y autocorrelación parcial para determinar los órdenes de p , q y d . En particular, diferenciaremos la serie hasta tres veces para identificar lo anterior.

Cabe destacar que no se verifica para un modelo de media móvil o uno autorregresivo, pues el ARIMA los contiene a ambos.



En base a la figura, pensamos en ajustar un $ARIMA(3,2,1)$, pues se presentan lags más claros que en la primera y segunda diferenciación.



$$ARIMA(3,2,1) = 110.8548, \quad R^2 = 0.1886824$$

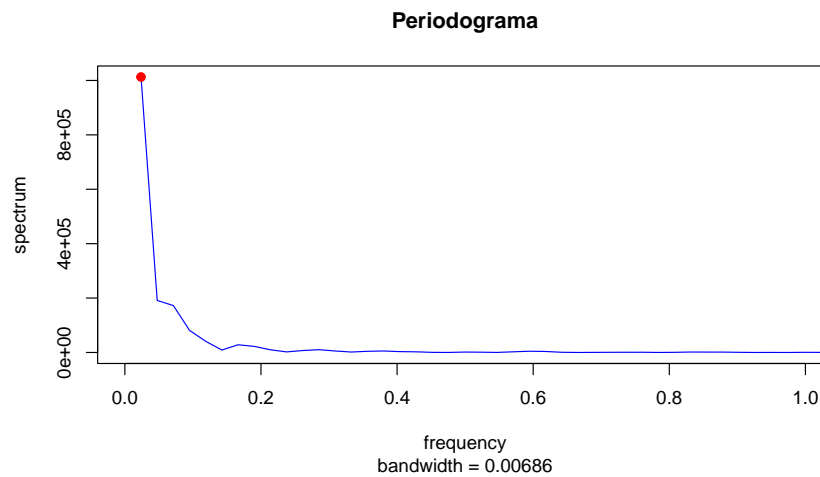
Obtenemos nuevamente una predicción que logra captar la tendencia correcta, mas no los patrones estacionales. Además, los coeficientes RMSE y R^2 son similares al modelo anterior.

ANÁLISIS ESPECTRAL DE LA SERIE

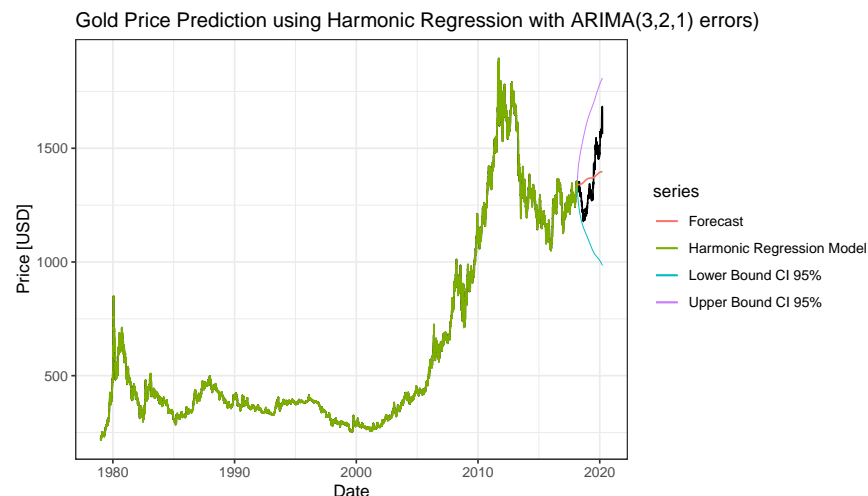
En tercer lugar y en vista de los resultado del ARIMA, pensamos en cómo solucionar el tema de la parte estacionaria la cual no es captada. Inmediatamente, surge la alternativa de agregar regresores externos al modelo como suma de senos y cosenos, es decir, realizar un modelo armónico de la forma

$$y_t = \beta_0 + \sum_{k=1}^K [\alpha_k s_k(t) + \gamma_k c_k(t)] + \varepsilon_t,$$

donde $s_i(t) = \sin(2\pi\omega_i t)$, $\gamma_i(t) = \cos(2\pi\omega_i t)$ y ε_t es modelado ahora como el ARIMA(3,2,1) del caso anterior. Observemos que debemos estimar el número del armónico K . Para ello recurrimos al análisis espectral



Así, obtenemos tan solo una frecuencia significativa $\omega = 0,02376302$. Con esto en mente, procedemos a construir el modelo y realizar su posterior testeo

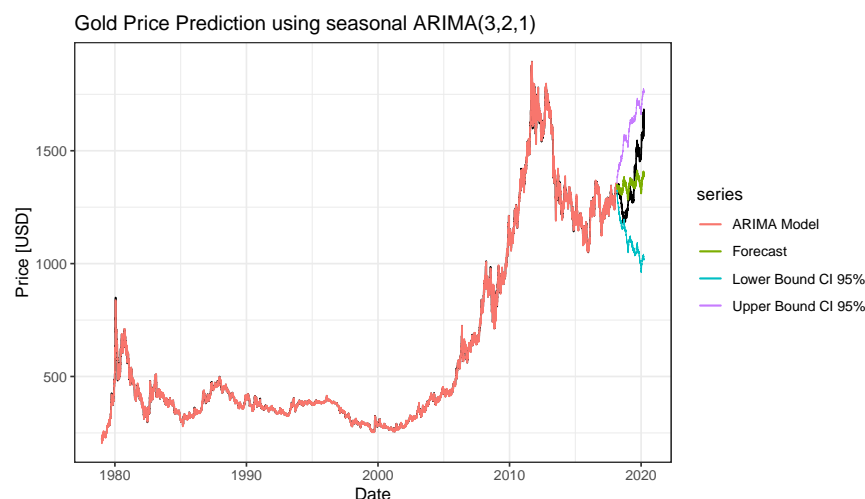


$$\text{Regresión armónica} = 110.0005, \quad R^2 = 0.2011387$$

Del gráfico, se desprende una leve mejora en la parte estacional.

EXTRA: ARIMA ESTACIONAL

Habiendo solucionado el tema de la estacionariedad del ARIMA usando un modelo armónico, surge la inquietud de poder arreglar lo anterior sin la necesidad de recurrir a regresores externos como funciones trigonométricas. Es por ello, que ajustaremos directamente su estacionariedad a través de una descomposición STL con función principal ARIMA(3, 2, 1). Esto es, una descomposición de la tendencia y la estacionariedad usando Loess, el cual es un método que permite estimar tendencia no lineal.



$$\text{seasonal ARIMA}(3,2,1) = 110.8653, \quad R^2 = 0.1885286$$

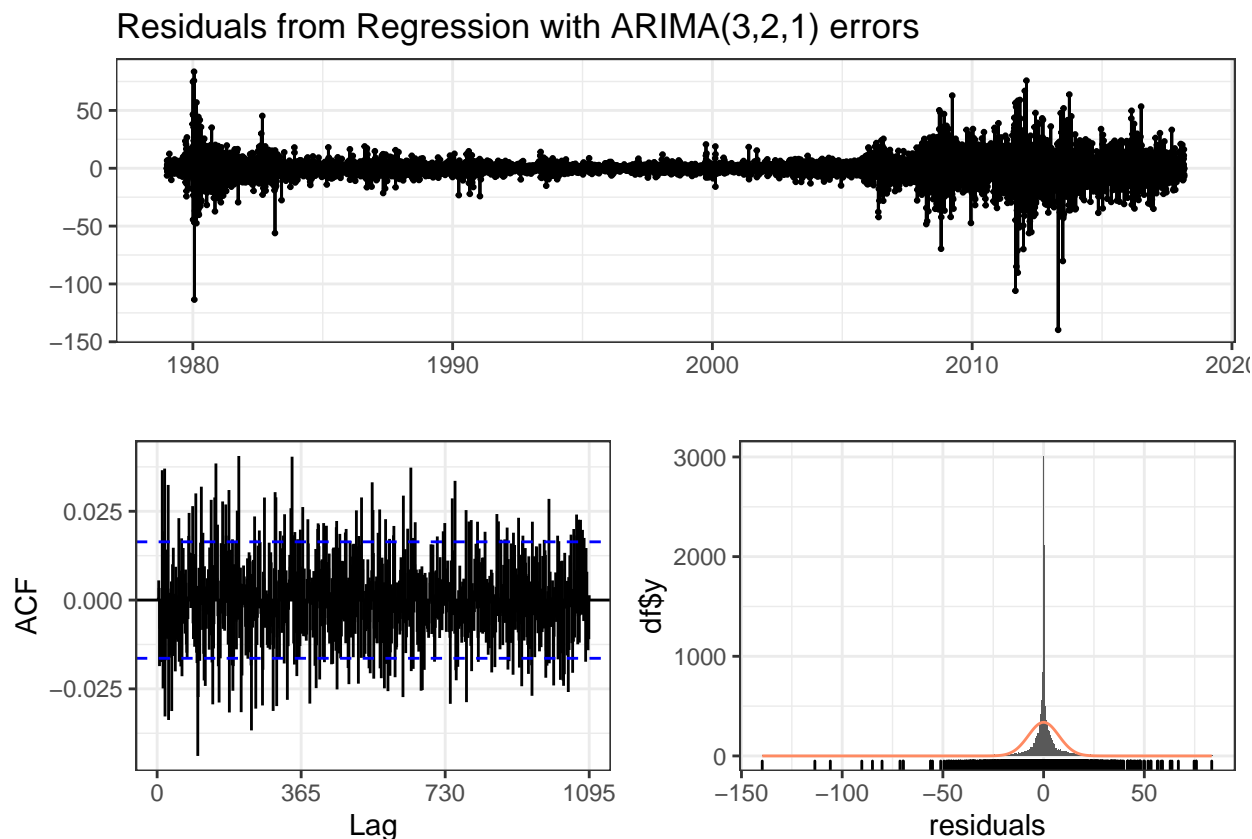
DIAGNÓSTICO DEL MODELO

La información de los modelos anteriores se resume en la siguiente tabla

	ARIMA(3,2,1)	Armónico	Holt-Winters	ARIMA(3,2,1) Seasonal
RMSE	110.8548	110.0005	110.9195	110.8653
R^2	0.1886824	0.2011387	0.1877348	0.1885286

En función de los resultados, la mayoría de los modelos presenta valores similares de RMSE y R^2 . Sin embargo, el que exhibe mejores resultados es el modelo armónico. Procedemos a realizar el diagnóstico del mismo y chequear la normalidad de los residuos.

Dado que tenemos muchos datos (más de 5000), el test de Shapiro-Wilk sucumbe en este escenario, por lo que usaremos el test de normalidad de Anderson-Darling.



Ljung-Box test data: Residuals from ARIMA(3,2,1) $Q^* = 1813.2$, $df = 726$,
 $p\text{-value} < 2.2e-16$ Model $df: 4$. Total lags used: 730

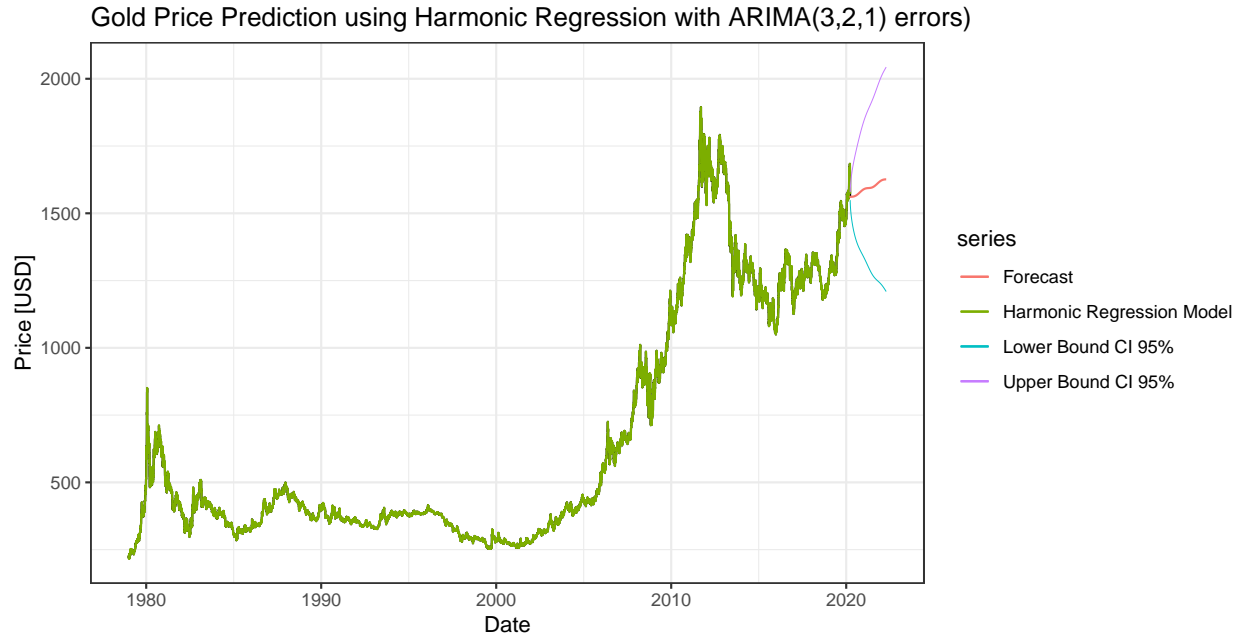
Anderson-Darling normality test data: $\text{fit}\$residuals$ $A = 1353.7$, $p\text{-value} < 2.2e-160$

Claramente los residuos no poseen normalidad, lo cual se confirma al ver el $p\text{-value}$ del test de Anderson-Darling. Más aún, el test Ljung Box ratifica la no independencia de los mismos.

Observación: Se realizó el chequeo de los supuestos para el resto de los modelos, pese a tener un desempeño menor. En todos los casos se obtuvo la no normalidad de los residuos.

PREDICCIÓN

Con todo el análisis realizado, predecimos los valores de la serie para dos años al futuro con el modelo armónico. En la figura, podemos notar que el precio del oro seguirá una tendencia alcista prontamente a realizar una lateralización.



DEPENDENCIA DE FACTORES EXTERNOS Y CORRELACIÓN

La plata y el oro son sustitutos perfectos entre sí. Existe evidencia significativa de que estos metales son una inversión atractiva y pueden desempeñar un papel útil en la diversificación del riesgo. Esto sugiere preguntarse si existe alguna relación entre ambos. Para ello recurriremos al coeficiente de Pearson, el cual se define como

$$\rho_{X,Y} = \frac{\mathbb{E}[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

y para datos muestrales como

$$\hat{\rho}_{X,Y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}.$$

Utilizaremos la base de datos proveniente de NASDAQ para obtener mediciones diarias del precio de la plata en el mismo intervalo que el oro y realizaremos la misma limpieza de los datos que antes. Luego, el gráfico obtenido es el siguiente



Comparando con el oro, obtenemos que el coeficiente de Pearson entre ambos es

$$\text{Pearson Coefficient Gold-Silver} = 0.8790895$$

Es decir, se encuentran fuertemente correlacionados. Por lo que podría realizarse un análisis predictivo en esta serie para estimar el comportamiento del precio del oro.

CONCLUSIÓN

A lo largo de este trabajo pudimos aplicar los principales métodos aprendidos durante el curso. Estudiando justificadamente cada uno de ellos en base al gráfico o a funciones relacionadas (ACF y PACF). Dados los resultados, podemos ratificar que la predicción de un activo financiero como el oro es sumamente difícil e inalcanzable con las herramientas enseñadas. Sin embargo, podrían usarse como indicador de tendencia para un trading a largo plazo.

Más aún, como el modelo no pasó el chequeo de los supuestos, la predicción obtenida no es confiable del todo. Ahora bien, para complementar lo anterior, hallamos que este metal comueve de forma positiva frente a la plata, lo que incita a realizar análisis temporales en activos correlacionados y obtener una regresión lineal entre ellos a modo de obtener una predicción.

ANEXO

Librerías

```
library(tidyverse)
library(fpp2)
library(padr)
library(lubridate)
library(Metrics)
library(astsa)
library(MLmetrics)
library(ggplot2)
library(Quandl)
library(SpatialPack)
library(nortest)
theme_set(theme_bw())
```

R2

```
R2 <- function(true, fitted){
  1 - (sum((true - fitted)^2)/sum((true - mean(true))^2))
}
```

Limpieza de los datos

```
gold <- read.csv("C:/Users/ccfer/Downloads/gold_price_data.csv",
  colClasses = c("Date", NA))
gold <- gold[-c(1:37), ]
gold <- mutate(gold, Date = ymd(Date))
gold <- pad(gold, interval = "day")
gold <- na.locf(gold)
```

Convertir a serie de tiempo

```
gold_ts <- ts(data = gold$Value, start = 1979, frequency = 365)

autoplot(gold_ts)+
  geom_line(color = "blue", size = 0.1) +
  labs(x = "Date", y = "Price [USD]",
    title = "Gold Price")
```

Cross Validation

```
train <- head(gold_ts, round(length(gold_ts) * 0.95))
h <- length(gold_ts) - length(train)
test <- tail(gold_ts, h)

### Modelo 1 - Holt Winter Multiplicativo ###

gold_decompose <- decompose(gold_ts, type = "multiplicative")
autoplot(gold_decompose)

gold_HWmodel <- HoltWinters(train, seasonal = "multiplicative")

gold_HWforecast <- forecast(train, h, level = 95)

autoplot(gold_ts) +
  autolayer(gold_HWmodel$fitted[,1], lwd = 0.5,
            series = "HW Model", ) +
  autolayer(gold_HWforecast$mean, lwd = 0.5,
            series = "Forecast") +
  autolayer(gold_HWforecast$lower, lwd = 0.1,
            series = "Lower Bound CI 95%") +
  autolayer(gold_HWforecast$upper, lwd = 0.1,
            series = "Upper Bound CI 95%") +
  labs(x = "Date", y = "Price [USD]",
       title = "Gold Price Prediction using Seasonal HW")

rmse(test, gold_HWforecast$mean)
R2(test, gold_HWforecast$mean)

### Modelo 2 - ARIMA ###

par(mfrow = c(3,2))
acf(train, main = "Serie original")
pacf(train, main = "Serie original")
acf(diff(train), main = "Serie diferenciada")
pacf(diff(train), main = "Serie diferenciada")
acf(diff(diff(train)), main = "Serie doblemente diferenciada")
pacf(diff(diff(train)), main = "Serie doblemente diferenciada")

gold_arima <- Arima(train, order = c(3, 2, 1))
gold_ARIMA_NSforecast <- forecast(gold_arima, h, level=95)

autoplot(gold_ts) +
  autolayer(gold_arima$fitted, lwd = 0.5,
            series = "ARIMA Model") +
```

```
autolayer(gold_ARIMA_NSforecast$mean, lwd = 0.5,
          series = "Forecast") +
autolayer(gold_ARIMA_NSforecast$lower, lwd = 0.1,
          series = "Lower Bound CI 95%") +
autolayer(gold_ARIMA_NSforecast$upper, lwd = 0.1,
          series = "Upper Bound CI 95%") +
labs(x = "Date", y = "Price [USD]",
      title = "Gold Price Prediction using ARIMA(3,2,1)")

rmse(test, gold_ARIMA_NSforecast$mean)
R2(test, gold_ARIMA_NSforecast$mean)

### Modelo 3 - SARIMA ###

gold_arima_stl <- stlm(y = train, modelfunction=Arima, order=c(3,2,1))

gold_ARIMAforecast <- forecast(gold_arima_stl, h,level=95)

autoplot(gold_ts) +
  autolayer(gold_arima_stl$fitted, lwd = 0.5,
            series = "ARIMA Model") +
  autolayer(gold_ARIMAforecast$mean, lwd = 0.5,
            series = "Forecast") +
  autolayer(gold_ARIMAforecast$lower, lwd = 0.1,
            series = "Lower Bound CI 95%") +
  autolayer(gold_ARIMAforecast$upper, lwd = 0.1,
            series = "Upper Bound CI 95%") +
  labs(x = "Date", y = "Price [USD]",
        title = "Gold Price Prediction using seasonal ARIMA(3,2,1)")

rmse(test, gold_ARIMAforecast$mean)
R2(test, gold_ARIMAforecast$mean)

### Modelo 4 - ARIMA with Harmonic Regression ###

per <- spec.pgram(gold_ts, taper=0, log = "no", main = "Periodograma", col = "blue",
                  xlim = c(0, 1))

sort(per$spec)
s <- per$spec[per$spec>400000]
f <- per$freq[per$spec>400000]

points(f, s, cex = 1, pch = 19, col ="red")

harmonics <- fourier(train, K = 1)
```

```
fit <- Arima(train, order = c(3, 2, 1), xreg = harmonics)

newharmonics <- fourier(train, K = 1, h)
gold_harmonic <- forecast(fit, xreg = newharmonics, level = 95)

autoplot(gold_ts) +
  autolayer(gold_harmonic$fitted, lwd = 0.5,
    series = "Harmonic Regression Model") +
  autolayer(gold_harmonic$mean, lwd = 0.5,
    series = "Forecast") +
  autolayer(gold_harmonic$lower, lwd = 0.1,
    series = "Lower Bound CI 95%") +
  autolayer(gold_harmonic$upper, lwd = 0.1,
    series = "Upper Bound CI 95%") +
  labs(x = "Date", y = "Price [USD]",
    title = "Gold Price Prediction using Harmonic Regression with
    ARIMA(3,2,1) errors")

rmse(test, gold_harmonic$mean)
R2(test, gold_harmonic$mean)

### Chequear supuestos ###

checkresiduals(fit)
checkresiduals(gold_HWmodel)
checkresiduals(gold_arima_stl)
checkresiduals(gold_arima)

ad.test(fit$residuals)

### Predicción ###

harmonics <- fourier(gold_ts, K = 1)

fit <- Arima(gold_ts, order = c(3, 2, 1), xreg = harmonics)

newharmonics <- fourier(gold_ts, K = 1, h)
gold_harmonic <- forecast(fit, xreg = newharmonics, level = 95)

autoplot(gold_ts) +
  autolayer(gold_harmonic$fitted, lwd = 0.5,
    series = "Harmonic Regression Model") +
  autolayer(gold_harmonic$mean, lwd = 0.5,
    series = "Forecast") +
  autolayer(gold_harmonic$lower, lwd = 0.1,
```

```
      series = "Lower Bound CI 95%") +
autolayer(gold_harmonic$upper, lwd = 0.1,
      series = "Upper Bound CI 95%") +
labs(x = "Date", y = "Price [USD]",
      title = "Gold Price Prediction using Harmonic Regression with
      ARIMA(3,2,1) errors)")

### Silver Comovement ###

silver = Quandl("LBMA/SILVER", api_key="h_TbsUf7pyA9Dj1MpELq")
silver <- silver[nrow(silver):1,]

silver <- silver[-c(1:2791), ]
head(silver)
silver <- silver[-c(10415:11000),]
silver[c(3,4)]<-list(NULL)
silver <- mutate(silver, Date = ymd(Date))
silver <- pad(silver, interval = "day")
silver <- na.locf(silver)
silver <- silver[-c(1:3), ]
silver <- silver[-c(15049:15912),]

x <- silver$USD
y <- gold$Value

print(cor(x, y, method = "pearson"))

silver_ts <- ts(data = x, start = 1979,frequency = 365)
autoplot(silver_ts)+
  geom_line(color = "blue", size = 0.1) +
  labs(x = "Date", y = "Price [USD]",
      title = "Silver Price")
```
