



Instituto de Estadística
Ingeniería en Estadística y Ciencia de Datos

Comparación de distintos modelos SARIMA para la serie temporal del precio en dólares por libra del cobre refinado entre 2008-2022

Ricardo Cristian Menares Eyzaguirre
ricardo.menares@postgrado.uv.cl

Valparaíso, Chile
11 de Julio de 2022

Índice general

1. Introducción	2
1.1. Objetivo general	3
1.2. Objetivos específicos	3
1.3. Hipótesis	3
2. Marco teórico	4
2.1. Investigaciones previas	4
2.2. Modelo ARIMA y SARIMA	5
2.3. ACF y PACF	6
2.4. Supuestos del modelo	8
2.5. Estimación de los parámetros de los modelos	10
2.6. Medidas de rendimiento de los modelos	11
3. Metodología	13
3.1. Metodología a usar	13
3.2. Identificación y elección de posibles modelos ARIMA	14
3.3. Ajuste de los modelos ARIMA	20
3.4. Diagnóstico de los modelos ARIMA	24
3.5. Predicción de los modelos ARIMA	29
4. Conclusiones	32
5. Anexo: Códigos en R.	33
6. Referencias	39

Capítulo 1

Introducción

El cobre, elemento metálico de color rojizo, tiene una gran importancia en el desarrollo económico de los países exportadores de esta materia prima, pues es de vital interés para mejorar las condiciones de vida de las personas.

Específicamente en Chile, uno de los países más productores y exportadores del cobre, alcanzando el año 2020 la primera posición de producción de cobre refinado con 2,4 toneladas de este material según la organización US Geological Survey(USGS), el impacto del precio del cobre en relación con las finanzas del país es bastante alto, incluso según el servicio nacional de geología y minería(SERNAGEOMIN) a través del anuario de la minería en Chile 2020, la producción de cobre representa un 11,2 % del PIB de Chile, lo cual es una cifra significativa y por ende es notorio que es completamente importante este metal.

La gran magnitud de inversiones realizadas y la cantidad creciente de proyectos que giran en torno a esta actividad en Chile exigen la continua realización de estudios que permitan a las empresas mejorar permanentemente sus sistemas productivos. Es así, que es indispensable indagar acerca de la evolución a través del tiempo del precio del cobre refinado en dólares por libra, surgiendo la interrogante sobre si es posible obtener uno o varios modelos adecuados de series de tiempo que permita describir, explicar, predecir y controlar los datos.

Debido a lo anterior, la presente investigación tiene como objetivo elegir el o los mejores modelos SARIMA(p,d,q)x(P,D,Q)[s] que se ajusten a la serie de tiempo del precio en dólares por libra del cobre refinado entre los años 2008 y 2022.

1.1. Objetivo general

Elegir y comparar los mejores modelos SARIMA(p,d,q)x(P,D,Q)[s], ajustados para la serie de tiempo del precio en dólares por libra del cobre refinado entre los años 2008 y 2022.

1.2. Objetivos específicos

1. Proponer distintos modelos de series de tiempo SARIMA(p,d,q)x(P,D,Q)[s].
2. Ajustar los modelos de series de tiempo propuestos mediante el método de máxima verosimilitud.
3. Elegir el o los modelos con mejor ajuste a la serie mediante distintos tests de hipótesis y métricas de desempeño tradicionales de series de tiempo.

1.3. Hipótesis

Es posible elegir el o los mejores modelos ARIMA(p,d,q) ajustados para la serie de tiempo del precio en dólares por libra del cobre refinado entre los años 2008 y 2022.

Capítulo 2

Marco teórico

2.1. Investigaciones previas

En cuanto al estudio del ajuste de modelos estocásticos para la serie del cobre refinado se ha encontrado distintas investigaciones, tales como la de Engel & Valdés (2001) que comparan la capacidad predictiva de mediano plazo de una variada gama de modelos de series cronologicas para el precio del cobre, tales como modelos MA(q), AR(p) y ARIMA(p,d,q). Para ello, usaron una base de datos desde el año 1908 a 1999 y usaron distintas métricas de desempeño, tales como el error cuadrático medio y la raíz del error cuadrático medio, lo cual, esta última, se ha decidido incluir en este trabajo. Como conclusiones, se dice que el modelo AR(1) ha sido el que mejor ajuste ha tenido dentro de todos los modelos propuestos, lo cual es llamativo.

Por otro lado, también se destaca la investigación de Contreras & Idrovo (2021) donde realizan un estudio de funciones de impulso-respuesta de un modelo vectorial autorregresivo para captar la dinámica entre el precio del cobre y distintos permisos de construcción de viviendas, además de la autocorrelación cruzada entre estas series. Esta investigación ha sido de gran relevancia, pues ha ayudado a poder entender de mejor manera la importancia del precio del cobre en distintas materias, en este caso los permisos de vivienda.

2.2. Modelo ARIMA y SARIMA

En esta investigación, se trabajará con 2 modelos estocásticos tradicionales de series de tiempo, como lo son el proceso autorregresivo integrado de media móvil y el autorregresivo integrado de media móvil estacional, ARIMA y SARIMA por sus siglas en inglés, respectivamente.

El proceso estocástico ARIMA(p,d,q), sistematizado por Box y Jenkins(1970), con p el orden de la parte autorregresiva estacionaria, d el número de raíces unitarias (orden de integración del proceso) y q el orden de la parte de media móvil, definido como:

$$(1 - \alpha_1 B - \dots - \alpha_p B^p)(1 - B)^d X_t = c + (1 - \theta_1 B - \theta_q B^q) Z_t \quad (2.1)$$

En donde Z_t sigue un proceso de ruido blanco(μ, σ^2) y $Cov(Z_t, Z_{t+k}) = 0$, además, α_i y θ_j , $i=1, \dots, p$; $j=1, \dots, q$ son parámetros, c constante y B un operador de retardo tal que $BX_t = X_{t-1}$.

El modelo ARIMA presentado puede escribirse de manera más simple con:

$$\phi_p(B) \nabla^d X_t = c + \theta_q(B) Z_t \quad (2.2)$$

Donde:

$$\begin{aligned} \phi_p(B) &= (1 - \alpha_1 B - \dots - \alpha_p B^p), \\ \nabla^d &= (1 - B)^d, \\ \theta_q(B) &= (1 - \theta_1 B - \theta_q B^q) \end{aligned}$$

Por otro lado, según Peña(2005, pág. 200) un modelo SARIMA multiplicativo viene dado por la forma:

$$\Phi_P(B^s) \phi_p(B) \nabla_s^D \nabla^d X_t = \Theta_Q(B^s) \theta_q(B) Z_t \quad (2.3)$$

Con

$$\begin{aligned} \Phi_P(B^s) &= (1 - \Phi_1 B^s - \dots - \Phi_P B^{sP}), \\ \Delta_s^D &= (1 - B^s)^D, \end{aligned}$$

$$\Theta_Q(B^s) = (1 - \Theta_1 B^s - \dots - \Theta_P B^{sP})$$

Así, un modelo SARIMA multiplicativo se escribirá como $\text{ARIMA}(p,d,q) \times (P,D,Q)[s]$, en donde p,d,q serán los operadores autorregresivos, integrados y media móvil regulares y P,D,Q los mismos operadores pero "estacionales". Además, s denotará el periodo estacional en cuestión.

2.3. ACF y PACF

En consiguiente, se considerarán 2 funciones de autocorrelación, la función autocorrelación simple(ACF) y la función de autocorrelación parcial(PACF), por sus siglas en inglés.

En cuanto a la ACF, va a estar dado por:

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)} \quad (2.4)$$

Donde $\gamma(\tau)$ es la autocovarianza de orden τ .

Sin embargo, lo anterior es descrito para una representación teórica. En el caso de este trabajo, al trabajar con una muestra, se deberá ocupar la ACF muestral. Esta ACF muestral será de gran ayuda para el trabajo, pues con ella es posible determinar si los residuos de los modelos propuestos son no correlacionados, además de poder determinar, heurísticamente, el orden de los coeficientes $\text{MA}(q)$ del modelo, según la cantidad de rezagos significativos que existan, también sirve para determinar el orden estacional $\text{MA}(Q)$, viendo los rezagos significativos que aparezcan con un comportamiento repetitivo de orden s .

Para calcular la ACF a través de una muestra x_1, x_2, \dots, x_n , primero se debe estimar la media mediante:

$$\bar{x} = \frac{\sum_{t=1}^T x_t}{T} \quad (2.5)$$

Donde $T=n$, indicando el último tiempo visto, con el supuesto de que las indexaciones de los tiempos van de 1 hasta n . Además, es posible observar que es un estimador insesgado, pues $E(\bar{x}) = \mu$.

Luego, un estimador para las autocovarianzas de orden k es el siguiente:

$$\gamma(k) = \frac{1}{T} \sum_{t=k+1}^T (x_t - \bar{x})(x_{t-k} - \bar{x}) \quad (2.6)$$

En consiguiente, sabiendo que $\rho(k) = \frac{\gamma(k)}{\gamma(0)}$, entonces la ACF muestral será:

$$\hat{\rho}(k) = \frac{\sum_{t=k+1}^T (x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^T (x_t - \bar{x})^2}, k = 0, |1|, |2|, \dots \quad (2.7)$$

Por otro lado, Las PACF miden la dependencia lineal de una variable después de eliminar el efecto de otra(s) variable(s) que afectan a ambas variables. Por ejemplo, la autocorrelación parcial mide el efecto (dependencia lineal) de y_{t-2} sobre y_t después de eliminar el efecto de y_{t-1} sobre y_t e y_{t-2} . Estas PACF tienen una importancia relevante, pues sirven heurísticamente para determinar el orden del coeficiente de la parte AR(q) del modelo, según la cantidad de rezagos significativos que tenga, también sirve para determinar los coeficientes de la parte estacional de un AR(Q), según la cantidad de rezagos significativos que tengan un periodo s notorio.

Según Box & et al. (2016, p.64). Las autocorrelaciones parciales se pueden describir en términos de p funciones distintas de cero de las autocorrelaciones. Denotando por ϕ_{kj} el j-ésimo coeficiente en una representación autorregresiva de orden k, con ϕ_{kk} el último coeficiente. Entonces se tiene:

$$\rho(j) = \phi_{k1}\rho(j-1) + \phi_{k2}\rho(j-2) + \dots + \phi_{kk}\rho(j-k) \quad (2.8)$$

Por ejemplo, ϕ_{22} dará el valor de la autocorrelación parcial de orden 2. Extendiendo el modelo a k rezagos, la estimación del último ϕ_{kk} dará la estimación de la autocorrelación parcial de orden k.

De lo anterior se desprende que para poder obtener las autocorrelaciones parciales hay que resolver la siguiente ecuación que se le suele llamar recursión de Durbin-Levinson :

$$\begin{bmatrix} \rho_0 & \rho_1 & \dots & \rho_{k-1} \\ \rho_1 & \rho_0 & \dots & \rho_{k-2} \\ \cdot & \cdot & \dots & \cdot \\ \rho_{k-1} & \rho_{k-2} & \dots & \rho_0 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} \quad (2.9)$$

Resolviendo estas ecuaciones para k=1,2,3 se obtiene:

$$\phi_{11} = \rho_1 \quad (2.10)$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} \quad (2.11)$$

$$\frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} \quad (2.12)$$

Luego, para poder calcular las autocorrelaciones muestrales, basta con estimar las autocorrelaciones, mediante la estimación de la ACF vista anteriormente, por tanto queda:

$$\begin{bmatrix} \hat{\rho}_0 & \hat{\rho}_1 & \dots & \hat{\rho}_{k-1} \\ \hat{\rho}_1 & \hat{\rho}_0 & \dots & \hat{\rho}_{k-2} \\ \vdots & \vdots & \dots & \vdots \\ \hat{\rho}_{k-1} & \hat{\rho}_{k-2} & \dots & \hat{\rho}_0 \end{bmatrix} \begin{bmatrix} \hat{\phi}_{k1} \\ \hat{\phi}_{k2} \\ \vdots \\ \hat{\phi}_{kk} \end{bmatrix} = \begin{bmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \vdots \\ \hat{\rho}_k \end{bmatrix} \quad (2.13)$$

2.4. Supuestos del modelo

Se harán supuestos importantes de los residuos del modelo, los cuales vienen dados por:

1. Son no correlacionados para cualquier retardo.
2. Siguen una distribución normal.
3. Media marginal igual a 0.
4. Varianza marginal constante.

Además, estas propiedades deben verificarse no solo respecto a las distribuciones marginales, sino también a las distribuciones condicionadas a cualquier conjunto de información de valores pasados de la serie.

Para comprobar estos supuestos, se realizan diferentes metodologías propuestas por la literatura. Para empezar, se analiza el gráfico a través del tiempo de los residuos, logrando ver heurísticamente si estos son no correlacionados, tienen media 0 y su varianza es constante.

En consiguiente, como se ha mencionado anteriormente, se analiza la ACF de los residuos, en donde se toma como regla de decisión de que si ningún rezago es significativo, quiere decir que los residuos son no correlacionados.

Para tener un resultado menos heurístico, se considera el test presentado por Ljung & Box (1978), con el nombre Ljung-Box, que es un contraste que tiene como hipótesis nula de que los datos son no correlacionados hasta un retardo h. El estadístico de prueba es:

$$Q(h) = T(T+2) \sum_{j=1}^h \frac{\hat{\rho}_j^2}{T-j} \quad (2.14)$$

Donde T es el tamaño de muestra, $\hat{\rho}_j$ es la autocorrelación simple muestral en el retraso j y h es el número de retrasos de la prueba.

Se rechazará la hipótesis nula cuando la probabilidad:

$$P(\chi^2(h-n)) > Q(h)$$

Sea significativamente pequeña, con n=número de parámetros estimados.

Continuando, para chequear la normalidad de los residuos, se ocupa el gráfico histograma de los residuos. Se analizará la forma que tienen las frecuencias de estos, para así poder realizar un análisis visual del comportamiento de estos resultados. En consiguiente, se procede a realizar un test presentado por Kolgomorov (1933) & Smirnov (1938), llamado test de Kolgomorov-Smirnov, en el caso normal, que tiene como hipótesis nula que la distribución de los datos es normal. Usa la función de distribución empírica dada por:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{[\infty, x]}(X_i) \quad (2.15)$$

Donde $I_{[\infty, x]}$ es la función indicadora, igual a 1 si $X_i \leq x$ y 0 si no se cumple

lo anterior. Luego, el estadístico de prueba para una función de distribución acumulada F(x) va a estar dado por:

$$D_n = \sup_x |F_n(x) - F(x)| \quad (2.16)$$

Para contrastar de que la media de los residuos es 0, se usará una prueba de hipótesis, llamada generalmente prueba t student, presentada por Gosset (1908). Esta prueba tiene como supuesto de que los datos son independientes e

idénticamente distribuidos, además de seguir una distribución normal. Entonces, si se busca contrastar que la media de los residuos es 0, se utiliza la prueba t student con hipótesis nula de que la media de los residuos es 0. Entonces, se tiene el siguiente estadístico de prueba:

$$T = \frac{\bar{Z}\sqrt{T}}{\hat{\sigma}} \quad (2.17)$$

Que bajo H_0 se distribuye $N(0,1)$. T en este caso, es el número de datos,

$\bar{Z} = \frac{\sum_{t=1}^T Z_t}{T}$ y $\hat{\sigma}^2 = \frac{\sum_{t=1}^T (Z_t - \bar{Z})^2}{T-1}$. Entonces, si T es significativamente mayor a un cuantil específico de la distribución $N(0,1)$, se rechazará la hipótesis nula.

Finalizando, se ocupa el test de Dickey-Fuller aumentado, ADF por sus siglas en inglés, para poder identificar si una serie es estacionaria o no. Su cálculo viene dado por:

$$\Delta X_t = \alpha + \beta t + \gamma x_{t-1} + \delta_{t-1} + \dots + \delta_{p-1} \Delta X_{t-p+1} + \epsilon_t \quad (2.18)$$

Donde α es constante, β un coeficiente de una tendencia temporal y p el orden de retraso de un coeficiente autorregresivo. Tiene como hipótesis $H_0 : \gamma = 0$ y $H_1 : \gamma < 0$, con un estadístico de prueba:

$$df_\tau = \frac{\hat{\gamma}}{\hat{s}_\gamma} \quad (2.19)$$

En donde el estadístico de prueba se compara con valores críticos particulares, particularmente con un nivel de confianza del 95 %, se comparará con un valor crítico $-1,95$ si el modelo no tiene constante y $-2,89$ si tiene.

2.5. Estimación de los parámetros de los modelos

La parte de estimación de los parámetros de los distintos modelos propuestos se hará mediante el método de máxima verosimilitud, donde si se cumplen los supuestos residuales, se puede describir la función de densidad conjunta de la muestra para un proceso ARIMA general como:

$$f(W_T) = \prod_{t=1}^T \sigma^{-1} v_{t|t-1}^{-1/2} (2\pi)^{-1/2} \exp \left(\frac{-1}{2\sigma^2} \sum_{t=1}^T \frac{(W_t - W_{t|t-1})^2}{v_{t|t-1}} \right) \quad (2.20)$$

Con $\sigma^2 v_{t|t-1}$ la varianza de los errores de predicción. Si se toman logaritmos, llamando $\beta = (\mu, \sigma^2, \alpha_1, \dots, \alpha_p, \theta_1, \dots, \theta_q)$ el vector de parámetros, la función de verosimilitud es:

$$L(\mathbf{B}) = -\frac{T}{2} \ln \sigma^2 - \frac{1}{2} \sum_{t=1}^T \ln v_{t|t-1} - \frac{1}{2\sigma^2} \sum_{t=1}^T \frac{e_t^2}{v_{t|t-1}} \quad (2.21)$$

Donde al maximizar esta función se consiguen las estimaciones para los parámetros.

2.6. Medidas de rendimiento de los modelos

Para poder evaluar y elegir los modelos más aptos para la serie presentada, se ocupan distintas métricas tradicionales que se ocupan en el análisis de series temporales.

La primera es el criterio de información de Akaike, AIC por sus siglas en inglés, presentado por Akaike (1974). Este criterio se basa en la estimación relativa de la información perdida cuando se utiliza un modelo determinado para representar el proceso que genera los datos. Este resultado no da información acerca de la calidad del modelo en sentido absoluto, es decir, si todos los modelos candidatos encajan mal, AIC no dará ningún aviso de ello.

Así, el AIC en el caso general está dado por:

$$AIC = 2k - 2\ln(L) \quad (2.22)$$

donde k es el número de parámetros en el modelo estadístico, y L es el máximo valor de la función de verosimilitud para el modelo estimado. Entonces, AIC penaliza con una función creciente la cantidad de parámetros que tiene el modelo en cuestión.

Entonces, dado un conjunto de modelos candidatos para los datos, el modelo preferido es el que tiene el valor mínimo en el AIC.

Por otro lado, también se consideran métricas de precisión, tales como la raíz del error cuadrático medio y el error absoluto medio, RMSE y MAE por sus siglas en inglés, respectivamente.

En cuanto al RMSE, representa la raíz cuadrada del segundo momento de la muestra de las diferencias entre los valores predichos y los valores observados. Entonces, RMSE es una medida de precisión, para comparar errores de

pronóstico de diferentes modelos para un conjunto de datos en particular. Su cálculo está dado por:

$$RMSE = \sqrt{\frac{\sum_{t=1}^T (\hat{y}_t - y_t)^2}{T}} \quad (2.23)$$

Donde \hat{y}_t es el valor predicho por el modelo y y_t el valor real. T es la cantidad de datos.

También, se considera la métrica MAE, la cual es un promedio aritmético de la diferencia absoluta entre los valores predichos y reales. Su fórmula está dada por:

$$MAE = \frac{\sum_{t=1}^T |\hat{y}_t - y_t|}{T} \quad (2.24)$$

Capítulo 3

Metodología

3.1. Metodología a usar

La metodología empleada en este estudio es una de las más conocidas en el ámbito de series de tiempo, la llamada metodología de Box & Jenkins. En ella, se definen 4 procesos fundamentales en un análisis de series temporales, los cuales se presentan en la siguiente imagen con el supuesto de que el ajuste es un ARIMA:

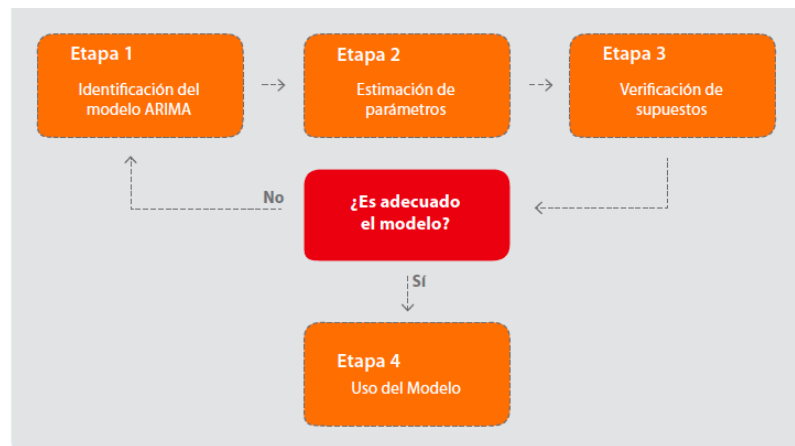


Figura 3.1: Metodología Box & Jenkins

Aquí se destaca lo siguiente:

1. **Identificación del modelo:** En esta etapa se busca uno o varios modelos aptos para poder ajustar a la serie en cuestión, en donde se usan

herramientas variadas, de las cuales algunas fueron explicadas en el marco teórico.

2. **Estimación de parámetros:** Luego de identificar el o los modelos posibles, se busca estimar los parámetros de estos ya sea mediante distintos métodos de optimización.
3. **Verificación de supuestos:** En consiguiente, se busca verificar cada uno de los supuestos propuestos del modelo en cuestión, de los cuales para nuestro interés han sido desglosados en el marco teórico.
4. **Uso del modelo:** Si el modelo resulta ser adecuado, se procede a emplear el modelo para los fines que se buscan, donde generalmente es el de predecir.

3.2. Identificación y elección de posibles modelos ARIMA

Se presenta el conjunto de datos utilizado el cual cuenta con 174 observaciones:

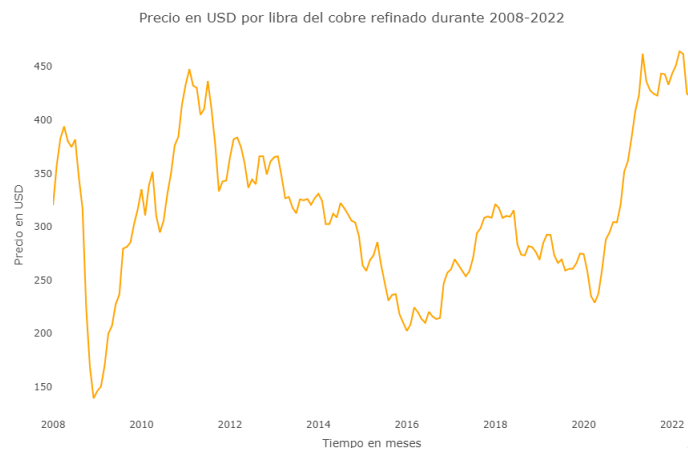


Figura 3.2: Precio mensual en dólares por libra de cobre refinado 2008-2022.

Se aprecia que la serie tiene un comportamiento no estacionario, ya que la media y varianza no son constantes, además no se observa que exista estacionariedad. Además, se advierte una fuerte caída entre mediados de 2008 y comienzos de 2009, seguramente producto de la baja compra de cobre refinado por países importadores, pues en esos años existía una incertidumbre importante respecto a la crisis europea. Luego de lo anterior, desde mediados del año 2009, el cobre

refinado tuvo un alza positiva notable, llegando hasta el valor máximo de 450 dólares por libra. También se evidencia una caída sistemática desde finales del 2011 hasta mediados del año 2016, para luego volver a subir hasta que ocurrió el estallido social en Chile en octubre del año 2019 y luego el comienzo de la pandemia de COVID-19, sin embargo, vuelve a aumentar desde mediados del 2020 seguramente porque se reactivó la economía de diversos países importadores de cobre, llegando incluso a un nuevo máximo valor del cobre.

Para comprobar si esta serie es estacionaria o no, primero se grafica la ACF muestral:

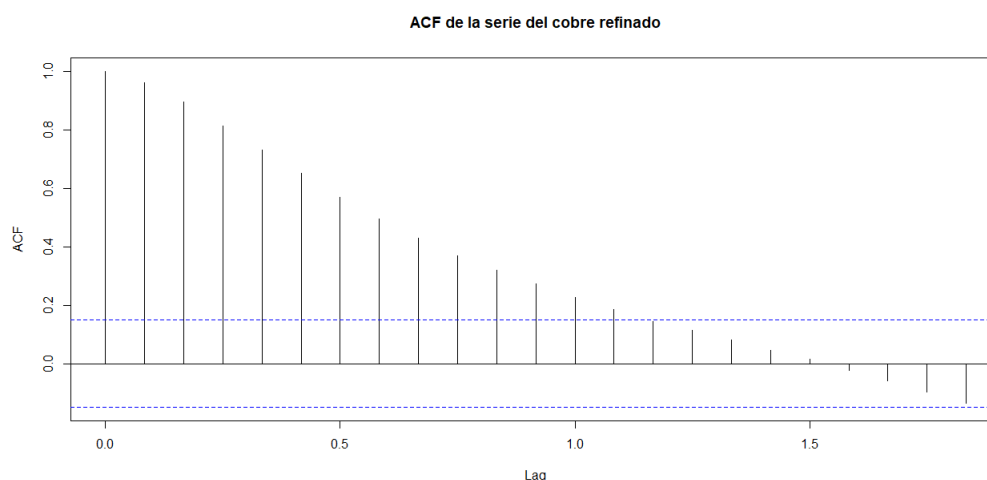


Figura 3.3: ACF de la serie del precio del cobre refinado entre 2008-2022

Se logra apreciar que la ACF empieza con muchas autocorrelaciones positivas, con decrecimiento lento y lineal, lo cual es una evidencia importante de que es una serie no estacionaria.

En consiguiente, se utiliza el contraste de Dickey-Fuller aumentado, donde los resultados obtenidos se presentan en la siguiente tabla:

Dickey-Fuller Aumentado	Orden de lag	Valor p
-1.39	0	0.832
-2.51	1	0.359
-2.65	2	0.303
-2.39	3	0.410
-2.44	4	0.390

Cuadro 3.1: Resultados test Dickey-Fuller Aumentado

Al obtener un valor p bastante altos, se concluye que no existe evidencia suficiente para rechazar la hipótesis nula que indica que la serie es no estacionaria, de esta manera, se obtiene evidencia suficiente de que hay que diferenciar la serie al menos una vez.

Luego, se diferencia la serie 1 vez, obteniendo el siguiente gráfico:

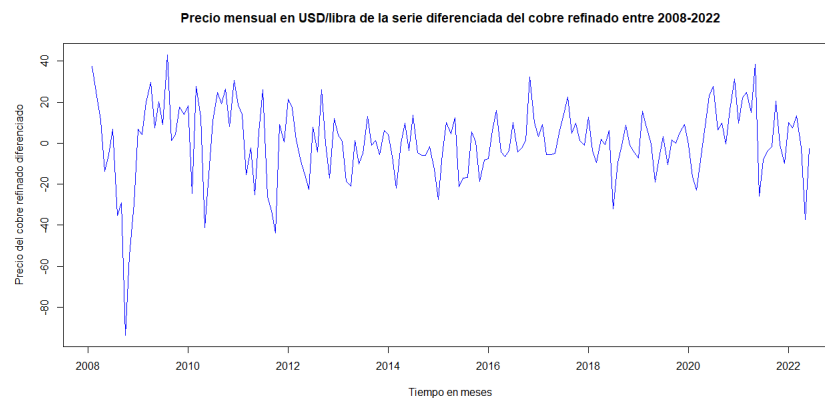


Figura 3.4: Serie diferenciada del precio del cobre refinado entre 2008-2022

Se logra apreciar que pareciera ser una serie más estacionaria a excepción de los datos entre 2008 y mediados del 2009, pues la media y varianza entre 2010 y 2022 parecieran ser constantes. Para comprobar si se logró una serie estacionaria, se presenta la ACF de esta serie diferenciada:

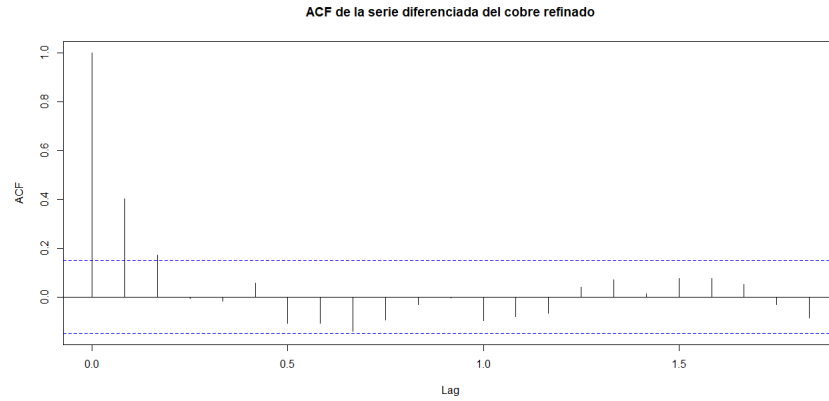


Figura 3.5: ACF de la serie diferenciada del precio del cobre refinado entre 2008-2022

Se logra apreciar que los lags ya no tienen un decrecimiento lento y lineal, además no empieza con tantas autocorrelaciones positivas, lo cual es una evidencia importante de que al diferenciar la serie se pudo haber conseguido una estacionariedad.

Complementando lo anterior, se realiza nuevamente el cálculo del test de Dickey-Fuller aumentado, obteniendo los siguientes resultados:

Dickey-Fuller aumentado	Orden de lag	Valor p
-8.66	0	$\leq 0,01$
-7.15	1	$\leq 0,01$
-6.88	2	$\leq 0,01$
-5.86	3	$\leq 0,01$
-4.85	4	$\leq 0,01$

Cuadro 3.2: Resultados test Dickey-Fuller aumentado para la serie diferenciada

Como al diferenciar la serie en el test de Dickey-Fuller aumentado se obtiene un valor p menor que 0.01 se infiere que existe evidencia suficiente para rechazar la hipótesis nula de no estacionariedad. Entonces, se ha logrado obtener una serie estacionaria con solo una diferenciación. En adelante, para obtener los coeficientes p y q del modelo ARIMA se trabajará con esta serie diferenciada.

En consiguiente, teniendo la serie diferenciada, se obtiene un posible modelo heurístico según su ACF y PACF, donde se presentan a continuación:

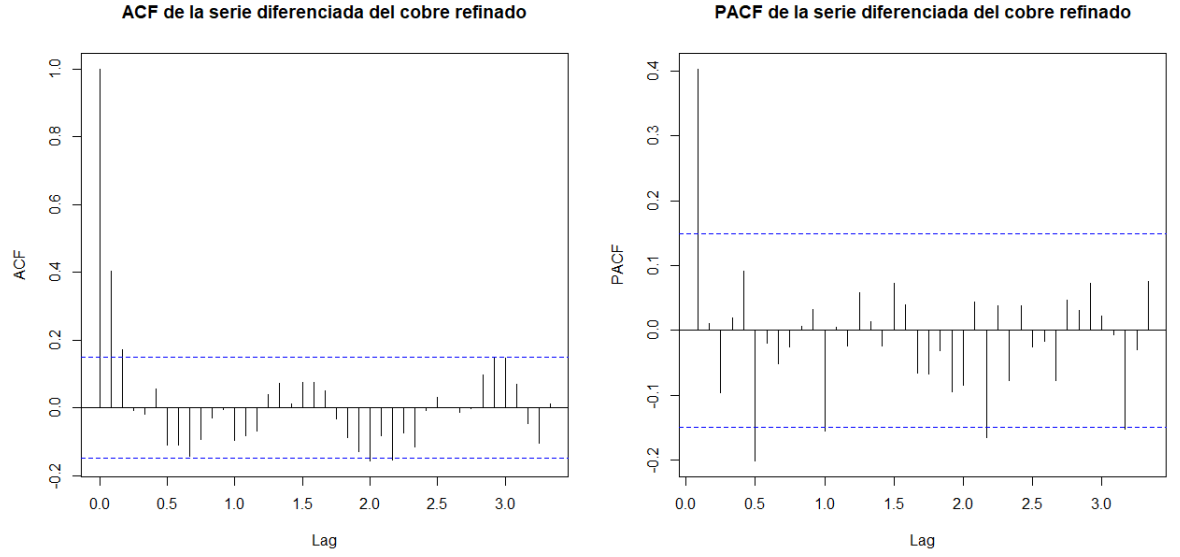


Figura 3.6: ACF y PACF de la serie diferenciada del precio en dólares del cobre refinado entre 2008-2022

Se observa en la ACF que el primer rezago es significativo, habiendo otro más cerca de la banda de Bartlett, se decide tomar en cuenta un modelo con 1 coeficiente MA y otro modelo con 2 coeficientes MA. Además, se observa un comportamiento estacional en los lags, sin embargo, estos no son significativos, por lo tanto, tomando en cuenta la ACF solo se considera uno y dos coeficientes para el proceso MA. Por otro lado, según la PACF se observa un rezago significativo, además, si se observa un comportamiento estacional significativo, por lo cual se considera un AR(1) y un estacional AR(1), por tanto, los modelos propuestos heurísticamente son:

$$ARIMA(1, 1, 1)x(1, 0, 0)[12]$$

$$ARIMA(1, 1, 2)x(1, 0, 0)[12]$$

En consiguiente, se procede a buscar los 2 mejores modelos ARIMA según su AIC, mediante la función `auto.arima` de R. Los 2 mejores modelos encontrados junto a los modelos heurísticos son los siguientes:

Modelo ARIMA	AIC
ARIMA(3,1,0)x(2,0,0)[12]	1465.93
ARIMA(2,1,1)x(2,0,0)[12]	1467.26
ARIMA(1,1,1)x(1,0,0)[12]	1466.54
ARIMA(1,1,2)x(1,0,0)[12]	1472.82

Cuadro 3.3: Los 2 mejores modelos ARIMA según su AIC y los dos modelos heurísticos

Se observa que el modelo heurístico ARIMA(1,1,2)X(1,0,0)[12] tiene un mejor AIC que el segundo mejor modelo entregado por la función `auto.arima` de R, por otro lado, el segundo modelo heurístico tiene un AIC peor que los demás.

3.3. Ajuste de los modelos ARIMA

En consiguiente, se ajustan los distintos modelos ARIMA propuestos, donde los parámetros estimados son:

Modelo ARIMA	ar1	ar2	ar3	sar1	sar2
ARIMA(3,1,0)x(2,0,0)[12]	0.3918	0.0182	-0.0879	-0.1165	-0.1695

Cuadro 3.4: Parámetros estimados del modelo ARIMA(3,1,0)x(2,0,0)[12]

Se observa que el segundo y tercer parámetro estimado del AR son bastante cercanos a 0, también el primer parámetro estimado del AR estacional. Entonces, podría ser un factor para reducir la cantidad de parámetros en este modelo, ya que se podrían considerar no significativos. En consiguiente, se presenta el gráfico de este modelo contrastado por la serie del cobre refinado, en conjunto de un IC del 95 % asumiendo normalidad, es la siguiente:

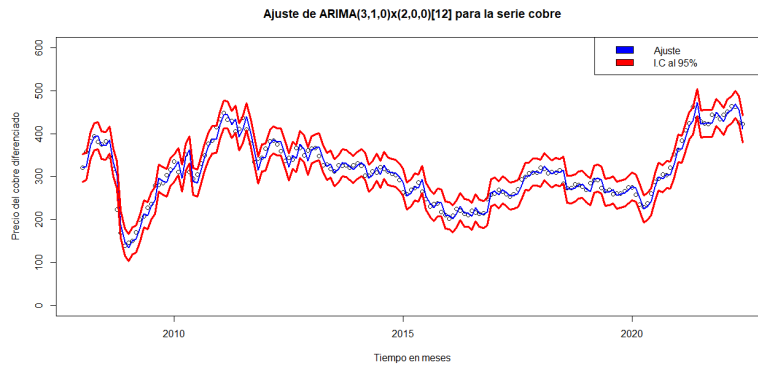


Figura 3.7: Ajuste del modelo ARIMA(3,1,0)x(2,0,0)[12] para la serie del cobre

Se aprecia que la mayoría de datos está dentro del intervalo de confianza del 95 %, lo cual es bueno.

Ahora se presentan los parámetros estimados del modelo ARIMA(2,1,1)x(2,0,0)[12]:

Modelo ARIMA	ar1	ar2	ma1	sar1	sar2
ARIMA(2,1,1)x(2,0,0)[12]	0.1791	0.0825	0.208	-0.1229	-0.1686

Cuadro 3.5: Parámetros estimados del modelo ARIMA(2,1,1)x(2,0,0)[12]

Se destaca que el segundo parámetro del AR es el más pequeño, lo cual si se quisiera rebajar la cantidad de parámetros podría eliminarse este, ahora se presenta el mismo gráfico pero para este modelo ARIMA(2,1,1)x(2,0,0)[12]:

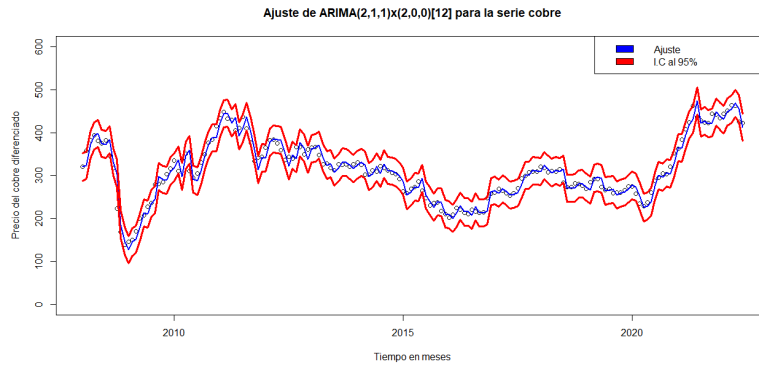


Figura 3.8: Ajuste del modelo ARIMA(2,1,1)x(2,0,0)[12] para la serie del cobre

También logra contener dentro del intervalo de confianza al 95 %, es bastante parecido el comportamiento que el anterior modelo.

Continuando con lo anterior, se presentan los parámetros estimados del modelo heurístico ARIMA(1,1,1)x(1,0,0)[12]:

Modelo ARIMA	ar1	ma1	sar1
ARIMA(1,1,1)x(1,0,0)[12]	0.4256	-0.0192	-0.1138

Cuadro 3.6: Parámetros estimados del modelo ARIMA(1,1,1)x(1,0,0)[12]

Se aprecia que el parámetro estimado del MA y del AR estacionario son bastante cercanos a 0, por lo cual es posible que se pueda estimar un AR(1) en este caso. Ahora, se presenta el respectivo gráfico del ajuste:

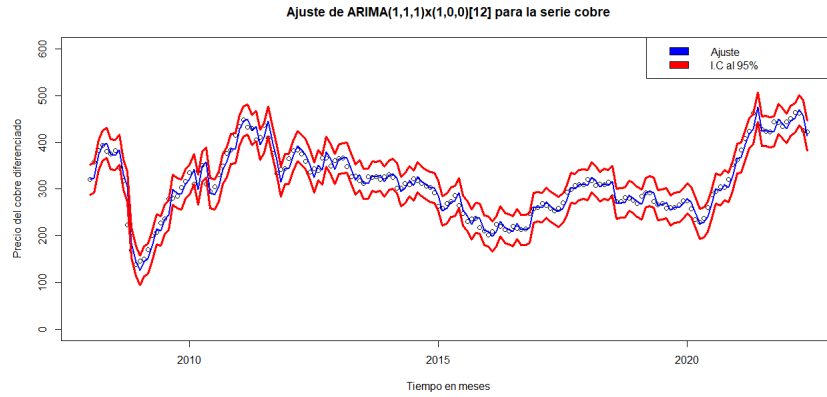


Figura 3.9: Ajuste del modelo $ARIMA(1,1,1)x(1,0,0)$ para la serie del cobre

Se consiguen las mismas conclusiones que los anteriores.

Finalizando, se presentan los parámetros estimados del modelo heurístico $ARIMA(1,1,2)x(1,0,0)[12]$.

Modelo ARIMA	ar1	ma1	ma2	sar1
$ARIMA(1,1,2)x(1,0,0)[12]$	-0.0420	0.4496	0.2797	-0.1032

Cuadro 3.7: Parámetros estimados del modelo $ARIMA(1,1,2)x(1,0,0)[12]$

Se aprecia que el parámetro estimado del AR es bastante cercano a 0, seguido por el AR estacionario. Lo anterior puede ser evidencia para poder querer reducir el número de parámetros estimados. En consiguiente, se presenta el gráfico del ajuste de este modelo:

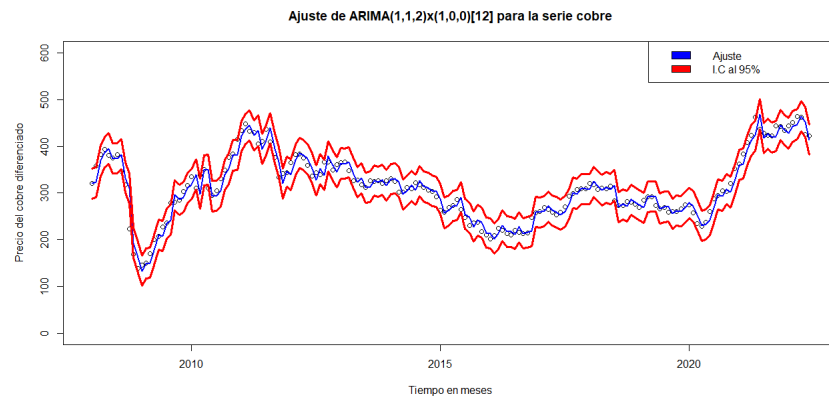


Figura 3.10: Ajuste del modelo $ARIMA(1,1,2) \times (1,0,0)$ para la serie del cobre

Obteniendo las mismas conclusiones que los otros 3 modelos.

3.4. Diagnóstico de los modelos ARIMA

Comenzando, se analizará si se cumplen los supuestos residuales. Primero, se grafican los residuos de los 4 modelos:

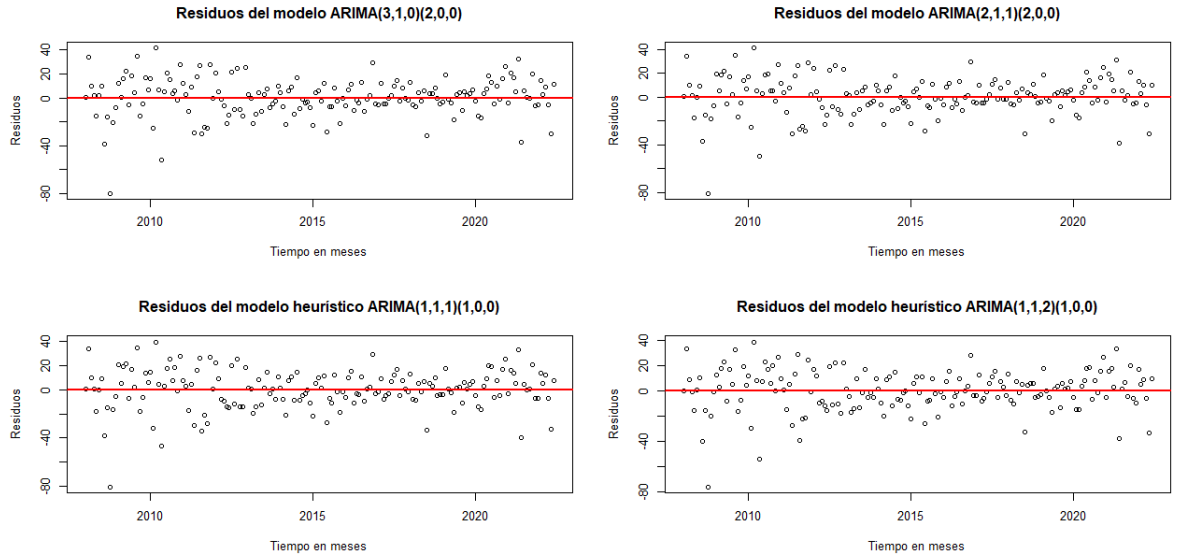


Figura 3.11: Residuos de los modelos propuestos

Se observa que en los 4 gráficos tienen un comportamiento distinto según el tiempo de la serie entre 2008 y 2009, donde se aprecia una mayor dispersión de los datos, también se aprecia lo mismo desde el año 2020 en adelante. Lo anterior, puede ocasionar que la media de los residuos sea significativamente distinta de 0, además de que los residuos tengan correlación según un rezago en particular, además de una varianza no constante. Por otro lado, se observa un comportamiento aleatorio a lo largo del resto de la serie en los 4 modelos.

Continuando con lo anterior, se presenta la ACF de los residuos de los 4 modelos, para chequear si los residuos son no correlacionados:

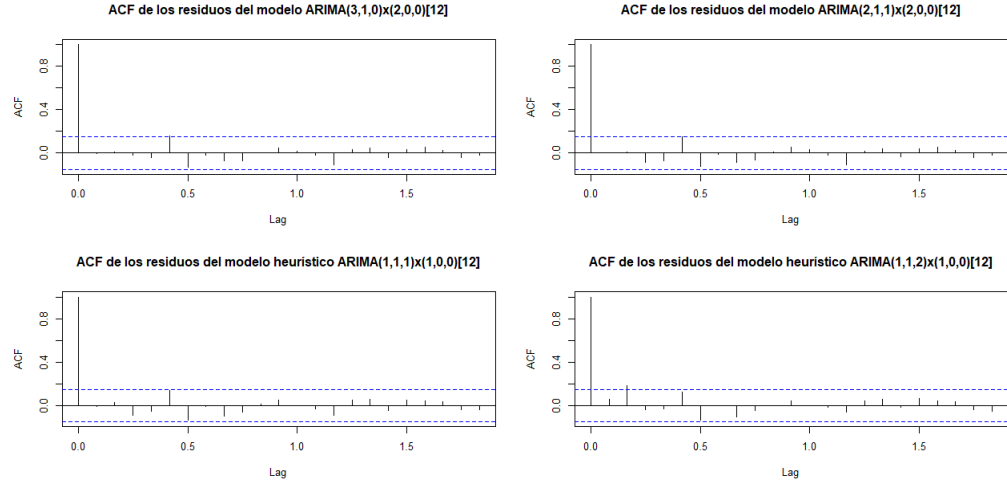


Figura 3.12: ACF de los residuos de los 4 modelos propuestos

Es notorio ver que en los 4 modelos existe una evidencia significativa de que los residuos son no correlacionados, sin embargo, en el último modelo se aprecia una cercanía de rezagos a la banda de Bartlett más notoria.

En consiguiente, se realiza el contraste Ljung-Box que tiene como hipótesis nula de que los residuos son no correlacionados, consiguiendo los siguientes gráficos de los p-valores del test según distintos lags:

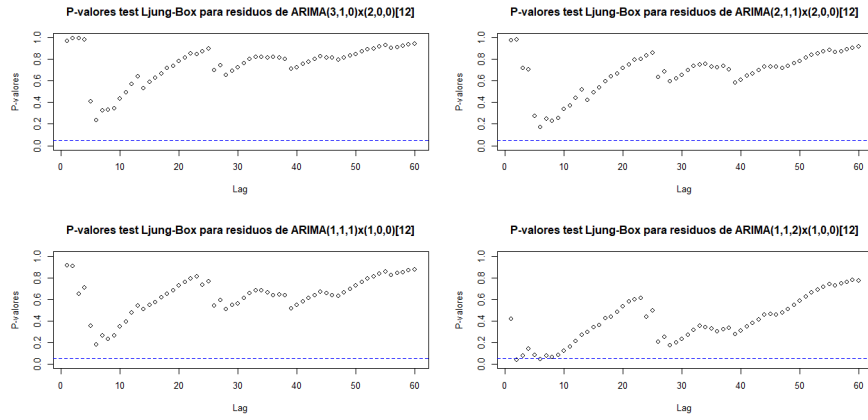


Figura 3.13: Contraste Ljung-Box para los residuos

Se observa en el modelo 1,2 y 3 se obtiene un comportamiento similar de los p-valores del test de Ljung-Box, desde el lag 1 al 60, con unos p-valores en su mayoría bastante alejados del 0.05, además se destaca el p-valor va aumentando generalmente según el orden del rezago crece. Por otro lado, reafirmando lo visto según la ACF residual, se aprecia que los p-valores del modelo 4 al comienzo son bastante cercanos a la banda del 0.05, incluso algunos son menores a 0.05 como el segundo rezago, por lo cual en este modelo se estaría al límite de rechazar la hipótesis nula de que los residuos son no correlacionados a un rezago entre 2 y 10, lo cual es una desventaja importante respecto a los otros modelos.

Continuando con el análisis residual, se presenta el histograma de los residuos de los 4 modelos:

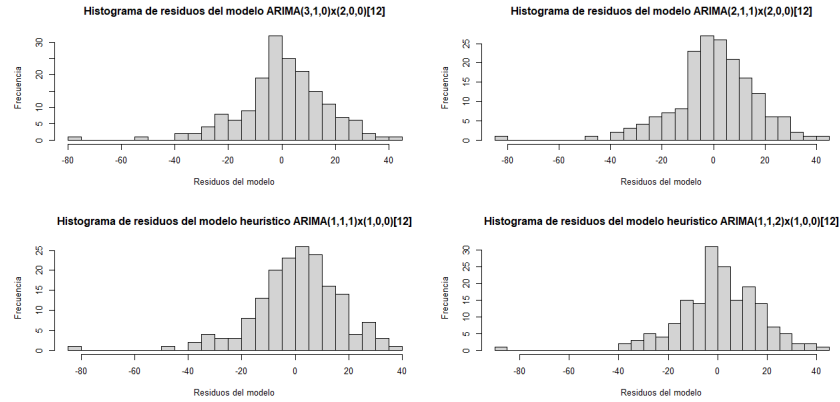


Figura 3.14: Histogramas de los residuos de los 4 modelos

Se aprecia que los 4 modelos tienen una distribución de residuos un poco asimétrica a la derecha, además de tener colas ligeras. No es total evidente el rechazo de la hipótesis de que sean normales, para comprobar lo anterior, se realiza un test Kolgomorov-Smirnov en el caso normal que tiene como hipótesis nula de que los datos siguen una distribución normal. Se presentan los resultados:

Modelo ARIMA	P-valor test Kolgomorov-Smirnov
ARIMA(3,1,0)x(2,0,0)[12]	0.1
ARIMA(2,1,1)x(2,0,0)[12]	0.09
ARIMA(1,1,1)x(1,0,0)[12]	0.32
ARIMA(2,1,1)x(1,0,0)[12]	0.19

Cuadro 3.8: P-valores del test Kolgomorov-Smirnov para los residuos de los modelos

Se concluye, con un nivel de confianza de al menos 95 % que no existe evidencia suficiente para rechazar la hipótesis nula de normalidad residual en los 4 modelos, por lo cual, es bastante bueno este resultado según los supuestos establecidos.

Luego, se realiza un test de hipótesis t para contrastar la hipótesis nula de que la media de los residuos es 0. Notar que este test es apto debido a que se ha diagnosticado que los residuos son no correlacionados y además siguen una distribución normal:

Modelo ARIMA	P-valor test t
ARIMA(3,1,0)x(2,0,0)[12]	0.7192
ARIMA(2,1,1)x(2,0,0)[12]	0.7516
ARIMA(1,1,1)x(1,0,0)[12]	0.7682
ARIMA(1,1,2)x(1,0,0)[12]	0.6248

Cuadro 3.9: P-valores del test t de media =0

Se observa que no existe evidencia suficiente para rechazar la hipótesis nula de que la media de los residuos es 0 en ninguno de los 4 modelos, lo cual es bastante bueno.

Luego, se realiza un análisis de igualdad de varianzas, para el tramo de la serie entre 2008-2010 y el resto de la serie, ya que se observa en el gráfico residual a través del tiempo que pareciera tener un cambio de varianza. Para ello, se realiza el test F para el cociente de varianzas, obteniendo los siguientes resultados:

Modelo ARIMA	Valor F	P-valor
ARIMA(3,1,0)x(2,0,0)[12]	3.4353	< 0.01
ARIMA(2,1,1)x(2,0,0)[12]	3.3946	< 0.01
ARIMA(1,1,1)x(1,0,0)[12]	3.4254	< 0.01
ARIMA(2,1,1)x(1,0,0)[12]	3.4046	< 0.01

Cuadro 3.10: Valores de test de varianza con 48 y 126 grados de libertad

Es notorio que existe evidencia suficiente para el rechazo de la hipótesis nula con un nivel de confianza de al menos 99 %, por tanto, se concluye que los residuos no tienen una varianza constante al menos en estos 2 tramos propuestos, lo cual es malo para los supuestos.

Finalizando, se presenta una tabla con distintas métricas importantes para evaluar los modelos ARIMA:

Modelo ARIMA	AIC	Log-EMV	RMSE	MAE	σ^2
ARIMA(3,1,0)x(2,0,0)[12]	1465.93	-726.97	16.12335	11.70268	260
ARIMA(2,1,1)x(2,0,0)[12]	1467.26	-727.63	16.18605	11.78073	262
ARIMA(1,1,1)x(1,0,0)[12]	1466.54	-729.27	16.37097	12.06829	268
ARIMA(1,1,2)x(1,0,0)[12]	1472.82	-731.41	16.51157	12.04673	272.6

Cuadro 3.11: Métricas importantes para comparar los modelos

Se observa que el modelo ARIMA(3,1,0)x(2,0,0)[12] fue el que tuvo menor AIC, RMSE, MAPE, varianza y mayor Log-EMV, por lo cual es el mejor modelo evaluado en esta sección. Seguido por el modelo ARIMA(2,1,1)x(2,0,0)[12] y finalizando con el modelo ARIMA(1,1,1)x(1,0,0)[12] que se aleja de los otros 2 en las métricas RMSE y MAE, pero no tanto del Log-EMV y la varianza, además tiene un AIC un poco menor que el modelo ARIMA(2,1,1)x(1,0,0)[12]. Por último, el modelo ARIMA(1,1,2)x(1,0,0)[12] si obtuvo peores resultados que los otros 3 modelos, incluso también en el análisis residual, por lo cual este modelo es el que menor se logró ajustar a los datos. Luego, como los modelos 1 y 2 estiman la misma cantidad de parámetros, es posible decir que estos modelos son los más óptimos para poder modelar la serie de tiempo del precio del cobre refinado. Por otro lado, también se puede indicar que es óptimo el modelo heurístico 1, pues este modelo estima solo 3 parámetros, lo cual es una diferencia de 2 con los modelos 1 y 2 de R.

3.5. Predicción de los modelos ARIMA

Se presenta los 4 gráficos del ajuste de los modelos junto con la predicción hasta Diciembre de 2022, es decir, a 6 pasos:

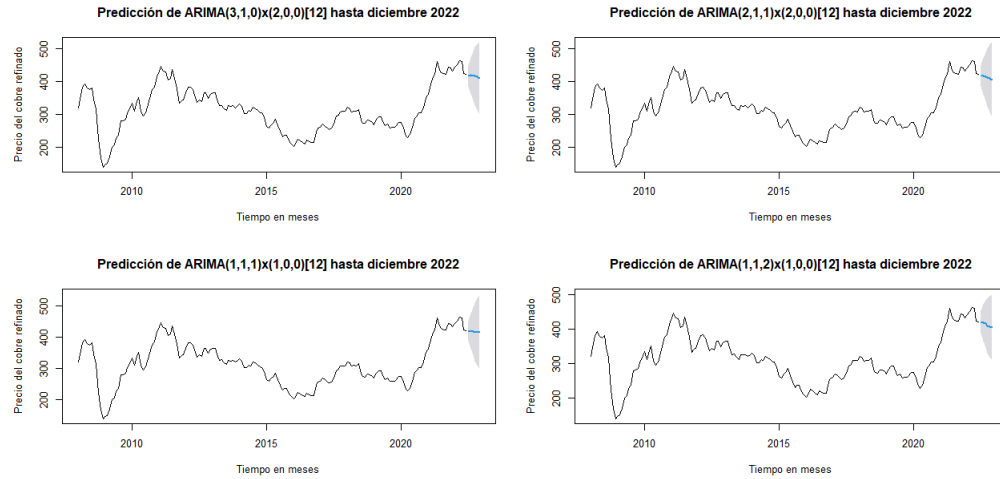


Figura 3.15: Predicción de 6 tiempos para los modelos

Se aprecia que los modelos $ARIMA(3,1,0) \times (2,0,0)[12]$ y $ARIMA(2,1,1)(2,0,0)[12]$ pareciera que tienen predicciones a la baja, mientras que el modelo heurístico $ARIMA(1,1,1) \times (1,0,0)[12]$ tiene una predicción más plana y lineal. En cuanto al último modelo, que fue el que peores resultados tuvo, tiene una predicción a la baja mucho más fuerte que los 2 primeros modelos.

En consiguiente, se presentan distintas tablas de cada modelo en donde indican la predicción de cada mes, el límite inferior y superior al 95 % de confianza y el rango de estos límites:

Mes	Predicción	LI 95 %	LS 95 %	Rango IC
Jul 2022	417.1640	385.5628	448.7652	63.2024
Ago 2022	418.7528	364.5954	472.9102	108.3148
Sep 2022	418.2995	344.9917	491.6073	146.6156
Oct 2022	416.5157	328.3509	504.6804	176.3295
Nov 2022	413.8426	313.5429	514.1423	200.5994
Dic 2022	409.5926	298.8757	520.3094	221.4337

Cuadro 3.12: Predicción de modelo ARIMA(3,1,0)x(2,0,0)[12] hasta diciembre 2022.

Mes	Predicción	LI 95 %	LS 95 %	Rango IC
Jul 2022	417.5338	385.8097	449.2579	63.4482
Ago 2022	416.5589	362.3120	470.8059	108.4939
Sep 2022	414.9711	341.9911	487.9512	145.9601
Oct 2022	412.4647	323.6049	501.3246	177.7197
Nov 2022	409.7604	307.0914	512.4294	205.338
Dic 2022	405.6615	290.7037	520.6192	229.9155

Cuadro 3.13: Predicción de modelo ARIMA(2,1,1)x(2,0,0)[12] hasta diciembre 2022.

Mes	Predicción	LI 95 %	LS 95 %	Rango IC
Jul 2022	420.4146	388.3281	452.5011	64.173
Ago 2022	419.7280	364.3543	475.1017	110.7474
Sep 2022	419.4386	344.3713	494.5060	150.1347
Oct 2022	416.8963	324.9761	508.8164	183.8403
Nov 2022	416.8800	310.2432	523.5167	213.2735
Dic 2022	417.9554	298.2056	537.7052	239.4996

Cuadro 3.14: Predicción de modelo heurístico ARIMA(1,1,1)x(2,0,0)[12] hasta diciembre 2022.

Mes	Predicción	LI 95 %	LS 95 %	Rango IC
Jul 2022	422.1458	389.7141	454.5775	64.8634
Ago 2022	419.0042	365.1932	472.8153	107.6221
Sep 2022	415.7607	347.6581	483.8632	136.2051
Oct 2022	410.1991	330.9227	489.4755	158.5528
Nov 2022	407.0718	318.5192	495.6245	177.1053
Dic 2022	405.0544	308.5470	501.5617	193.0147

Cuadro 3.15: Predicción de modelo heurístico ARIMA(1,1,2)x(2,0,0)[12] hasta diciembre 2022.

Se aprecia, con asombro, de que el rango del peor modelo evaluado hasta el momento, el ARIMA(1,1,1)x(2,0,0)[12] tuvo un menor rango de los límites de confianza, lo cual es llamativo. Además, se aprecia un comportamiento bastante similar en los 2 modelos obtenidos de la función `auto.arima` de R. En cuanto al otro modelo heurístico ARIMA(2,1,1)x(1,0,0)[12] se confirma una tendencia lineal plana.

Capítulo 4

Conclusiones

Según la naturalidad de la serie, es posible que ajustar un modelo ARIMA no sea lo suficientemente sofisticado, pudiendo haber realizado un análisis ARCH o GARCH, incluso ARMA-GARCH, u otro modelo más acorde a este tipo de series de tiempo.

Además, no se ha considerado factores externos que afectan a la evolución del precio del cobre refinado a través del tiempo, tales como la alza histórica que ha tenido el dólar, la reciente pandemia por el SARS-CoV-2 e incluso la recesión económica que han tenido distintos países en el último periodo que son principales responsables de la producción de cobre a nivel mundial, donde uno de ellos es Chile.

Sin embargo, ha sido posible ajustar distintos modelos ARIMA obteniendo resultados en su mayoría buenos, logrando poder elegir y comparar entre distintos modelos cuál es el mejor.

Dado lo anterior, al ajustar un modelo ARIMA se ha logrado obtener resultados bastante decentes, en donde el modelo $ARIMA(3,1,0) \times (2,0,0)$ [12] ha sido el modelo con mejores resultados, logrando satisfacer los supuestos residuales de media igual a 0, residuos no correlacionados y normales y obteniendo el mejor puntaje en AIC, RMSE, MAE, Log-EMV.

Capítulo 5

Anexo: Códigos en R.

```
library(forecast)
library(readxl)
library(stats)
library(car)

data=read_excel("preciocobre.xlsx")
cobre=ts(data$precio,frequency=12,start=c(2008,1))
plot(cobre,xlab="Tiempo en meses",ylab="Precio del cobre refinado",main="Precio
mensual en USD/libra del cobre refinado entre 2008-2022")

#### ACF
acf(cobre,main="ACF de la serie del cobre refinado")

#### TEST DICKEY FULLER AUMENTADO
adf.test(cobre)
# no es estacionaria

##### serie diferenciada
cobrediff=ts(diff(cobre),frequency=12,start=c(2008,2))
cobrediff
plot(cobrediff,xlab="Tiempo en meses",ylab="Precio del cobre refinado dife-
renciado", main="Precio mensual en USD/libra de la serie diferenciada del
cobre refinado entre 2008-2022",col="blue")

#### ACF de la serie diferenciada
acf(cobrediff,main="ACF de la serie diferenciada del cobre refinado",lag.max=36)

#### ADF test
adf.test(cobrediff)
```

```

##### ACF y PACF serie diferenciada para detectar parámetros AR y MA
par(mfrow=c(1,2))
acf(cobrediff,main="ACF de la serie diferenciada del cobre refinado",lag.max=40)
pacf(cobrediff,main="PACF de la serie diferenciada del cobre refinado",ylab="PACF",lag.max=40)
## ARIMA(1,1,1)x(1,0,0)[12] y ARIMA(1,1,2)x(1,0,0)[12]

##### mejores modelos ARIMA
mejoresmodelos=auto.arima(cobrediff,trace=T)
summary(mejoresmodelos)
#1. ARIMA(3,0,0)(2,0,0)[12]
#2. ARIMA(2,0,1)(2,0,0)[12]

##### modelos propuestos
modelo1=arima(cobre,order=c(3,1,0),seasonal=c(2,0,0),method="ML",include.mean=FALSE)
summary(modelo1)
#AIC:1465.93 LOG-EMV:-726.97 MAE:11.70268 RMSE:16.12335 sigma2:260

modelo2=arima(cobre, order=c(2,1,1),seasonal=c(2,0,0),method="ML",include.mean=FALSE)
summary(modelo2)
#AIC:1467.26 LOG-EMV:-727.63 MAE:11.78073 RMSE:16.18605 sigma2:262

modeloheuristico1=arima(cobre, order=c(1,1,1),seasonal=c(1,0,0),method="ML",include.mean=FALSE)
summary(modeloheuristico1)
#AIC:1466.54 LOG-EMV:-729.27 MAE:12.06829 RMSE:16.37097 sigma2:268

modeloheuristico2=arima(cobre, order=c(1,1,2),seasonal=c(1,0,0),method="ML",include.mean=FALSE)
summary(modeloheuristico2)
#AIC:1472.82 Log-EMV:-731.41 MAE:12.04673 RMSE:16.51157 sigma2:272.6

##### residuos modelos
par(mfrow=c(2,2))
plot(residuals(modelo1),xlab="Tiempo en meses",main="Residuos del modelo
ARIMA(3,1,0)(2,0,0)", ylab="Residuos")
abline(h=0,col="red",lwd=2)

plot(residuals(modelo2),xlab="Tiempo en meses",main="Residuos del mo-
delo ARIMA(2,1,1)(2,0,0)", ylab="Residuos")
abline(h=0,col="red",lwd=2)

plot(residuals(modeloheuristico1),xlab="Tiempo en meses",main="Residuos
del modelo heurístico ARIMA(1,1,1)(1,0,0)", ylab="Residuos")
abline(h=0,col="red",lwd=2)

plot(residuals(modeloheuristico2),xlab="Tiempo en meses",main="Residuos
del modelo heurístico ARIMA(1,1,2)(1,0,0)", ylab="Residuos") abline(h=0,col="red",lwd=2)

```

```

##### acf de residuos
par(mfrow=c(2,2))
acf(residuals(modelo1),main="ACF de los residuos del modelo ARIMA(3,1,0)x(2,0,0)[12]")
acf(residuals(modelo2),main="ACF de los residuos del modelo ARIMA(2,1,1)x(2,0,0)[12]")
acf(residuals(modeloheuristico1),
main="ACF de los residuos del modelo heurístico ARIMA(1,1,1)x(1,0,0)[12]")
acf(residuals(modeloheuristico2),
main="ACF de los residuos del modelo heurístico ARIMA(1,1,2)x(1,0,0)[12]")

##### LJUNG-BOX #####
BT1 = rep(NA,60)
for(i in 1:60){ BT1[i]=Box.test(residuals(modelo1), lag = i,
type = "Ljung-Box", fitdf = 0)$p.value
}
BT2 = rep(NA,60) for(i in 1:60){
BT2[i]=Box.test(residuals(modelo2), lag = i,
type = "Ljung-Box", fitdf = 0)$p.value
}
BT3 = rep(NA,60)
for(i in 1:60){
BT3[i]=Box.test(residuals(modeloheuristico1), lag = i,
type = "Ljung-Box", fitdf = 0)$p.value
}
BT4 = rep(NA,60)
for(i in 1:60){
BT4[i]=Box.test(residuals(modeloheuristico2), lag = i,
type = "Ljung-Box", fitdf = 0)$p.value
}

par(mfrow=c(2,2))
plot(BT1, ylim=c(0,1),
main="P-valores test Ljung-Box para residuos de ARIMA(3,1,0)x(2,0,0)[12]",
ylab="P-valores",xlab="Lag")
abline(h=0.05, col="blue", lty=2)
plot(BT2, ylim=c(0,1),
main="P-valores test Ljung-Box para residuos de ARIMA(2,1,1)x(2,0,0)[12]",
ylab="P-valores",xlab="Lag")
abline(h=0.05, col="blue", lty=2)
plot(BT3, ylim=c(0,1),
main="P-valores test Ljung-Box para residuos de ARIMA(1,1,1)x(1,0,0)[12]",
ylab="P-valores",xlab="Lag")
abline(h=0.05, col="blue", lty=2)
plot(BT4, ylim=c(0,1),
main="P-valores test Ljung-Box para residuos de ARIMA(1,1,2)x(1,0,0)[12]",
ylab="P-valores",xlab="Lag")
abline(h=0.05, col="blue", lty=2)

```

```

##### contraste de normalidad, histograma:#####
par(mfrow=c(2,2))
hist(residuals(modelo1),breaks=20,xlab="Residuos del modelo", ylab="Frecuencia",
main="Histograma de residuos del modelo ARIMA(3,1,0)x(2,0,0)[12]")
hist(residuals(modelo2),breaks=20,xlab="Residuos del modelo", ylab="Frecuencia",
main="Histograma de residuos del modelo ARIMA(2,1,1)x(2,0,0)[12]")
hist(residuals(modeloheuristico1),breaks=20,xlab="Residuos del modelo", ylab="Frecuencia",
main="Histograma de residuos del modelo heurístico ARIMA(1,1,1)x(1,0,0)[12]")
hist(residuals(modeloheuristico2),breaks=20,xlab="Residuos del modelo", ylab="Frecuencia",
main="Histograma de residuos del modelo heurístico ARIMA(1,1,2)x(1,0,0)[12]")

##### test Kolgomorov-Smirnov #####
ks.test( residuals(modelo1) , "pnorm", mean=mean(residuals(modelo1)), sd=sd(residuals(modelo1)))
ks.test( residuals(modelo2) , "pnorm", mean=mean(residuals(modelo2)) , sd=sd(residuals(modelo2)))
ks.test( residuals(modeloheuristico1) , "pnorm", mean=mean(residuals(modeloheuristico1))
, sd=sd(residuals(modeloheuristico1)))
ks.test( residuals(modeloheuristico2) , "pnorm", mean=mean(residuals(modeloheuristico2))
, sd=sd(residuals(modeloheuristico2)))

##### contraste de media cero #####
t.test(residuals(modelo1),mu=0)
#no se rechaza hipótesis nula de que la media es 0
t.test(residuals(modelo2),mu=0)
#no se rechaza hipótesis nula de que la media es 0
t.test(residuals(modeloheuristico1),mu=0)
#no se rechaza hipótesis nula de que la media es 0
t.test(residuals(modeloheuristico2),mu=0)
#no se rechaza hipótesis nula de que la media es 0

##### Contraste homocedasticidad #####
var.test(residuals(modelo1)[1:48], residuals(modelo1)[49:174],
null.value=1, alternative='two.sided', conf.level=0.95)
var.test(residuals(modelo2)[1:48], residuals(modelo2)[49:174],
null.value=1, alternative='two.sided', conf.level=0.95)
var.test(residuals(modeloheuristico1)[1:48], residuals(modeloheuristico1)[49:174],

```

```

null.value=1, alternative='two.sided', conf.level=0.95)
var.test(residuals(modeloheuristico2)[1:48], residuals(modeloheuristico2)[49:174],
null.value=1, alternative='two.sided', conf.level=0.95)

```

```

##### gráfico ajuste modelo 1 #####
plot(cobre,type="p",xlab="Tiempo en meses",main="Ajuste de ARIMA(3,1,0)x(2,0,0)[12]
para la serie cobre",ylim=c(0,600),ylab="Precio del cobre diferenciado")
lines(fitted.values(modelo1),col="blue",lwd=2)
lines(fitted.values(modelo1)+1.96*sqrt(modelo1$sigma2),col="red",lwd=3)
lines(fitted.values(modelo1)-1.96*sqrt(modelo1$sigma2),col="red",lwd=3)
legend(x="topright",legend=c("Ajuste","I.C al 95 %"),fill=c("blue","red"))

```

```

##### gráfico ajuste modelo 2 #####
plot(cobre,type="p",xlab="Tiempo en meses",main="Ajuste de ARIMA(2,1,1)x(2,0,0)[12]
para la serie cobre",ylim=c(0,600),ylab="Precio del cobre diferenciado")
lines(fitted.values(modelo2),col="blue",lwd=2)
lines(fitted.values(modelo2)+1.96*sqrt(modelo2$sigma2),col="red",lwd=3)
lines(fitted.values(modelo2)-1.96*sqrt(modelo2$sigma2),col="red",lwd=3)
legend(x="topright",legend=c("Ajuste","I.C al 95 %"),fill=c("blue","red"))

```

```

##### gráfico ajuste modelo heuristico 1 #####
plot(cobre,type="p",xlab="Tiempo en meses",main="Ajuste de ARIMA(1,1,1)x(1,0,0)[12]
para la serie cobre",ylim=c(0,600),ylab="Precio del cobre diferenciado")
lines(fitted.values(modeloheuristico1),col="blue",lwd=2)
lines(fitted.values(modeloheuristico1)+1.96*sqrt(modeloheuristico1$sigma2),col="red",lwd=3)
lines(fitted.values(modeloheuristico1)-1.96*sqrt(modeloheuristico1$sigma2),col="red",lwd=3)
legend(x="topright",legend=c("Ajuste","I.C al 95 %"),fill=c("blue","red"))

```

```

##### gráfico ajuste modelo heuristico 2 #####
plot(cobre,type="p",xlab="Tiempo en meses",main="Ajuste de ARIMA(1,1,2)x(1,0,0)[12]
para la serie cobre",ylim=c(0,600),ylab="Precio del cobre diferenciado")
lines(fitted.values(modeloheuristico2),col="blue",lwd=2)
lines(fitted.values(modeloheuristico2)+1.96*sqrt(modeloheuristico2$sigma2),col="red",lwd=3)
lines(fitted.values(modeloheuristico2)-1.96*sqrt(modeloheuristico2$sigma2),col="red",lwd=3)
legend(x="topright",legend=c("Ajuste","I.C al 95 %"),fill=c("blue","red"))

```

```

##### predicción #####
par(mfrow=c(2,2))
plot(forecast(modelo1,h=6,level=c(95)),
main="Predicción de ARIMA(3,1,0)x(2,0,0)[12] hasta diciembre 2022",xlab="Tiempo
en meses",ylab="Precio del cobre refinado")
plot(forecast(modelo2,h=6,level=c(95)),
main="Predicción de ARIMA(2,1,1)x(2,0,0)[12] hasta diciembre 2022",xlab="Tiempo
en meses",ylab="Precio del cobre refinado")
plot(forecast(modeloheuristico1,h=6,level=c(95)),
main="Predicción de ARIMA(1,1,1)x(1,0,0)[12] hasta diciembre 2022",xlab="Tiempo

```

```

en meses", ylab="Precio del cobre refinado")
plot(forecast(modeloheuristico2,h=6,level=c(95)),
main="Predicción de ARIMA(1,1,2)x(1,0,0)[12] hasta diciembre 2022",xlab="Tiempo
en meses", ylab="Precio del cobre refinado")

```

```

##### valores de predicción #####
forecast(modelo1,h=6,level=c(95))
forecast(modelo2,h=6,level=c(95))
forecast(modeloheuristico1,h=6,level=c(95))
forecast(modeloheuristico2,h=6,level=c(95))

```

Capítulo 6

Referencias

- Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE transactions on automatic control*, 19(6), 716-723. Recuperado de: [\[Ver aquí\]](#)
- Box, G., Jenkins, G., Reinsel, G. & Ljung, G. (2016). Time series analysis: Forecasting and control. 5ed. New Jersey: John Wiley & Sons, Inc. Recuperado de: [\[Ver aquí\]](#)
- Diario UChile.(21-06-2022). Trabajadores de Codelco convocan paralización nacional a partir de este miércoles: Se espera movilizar a 55 mil personas. Recuperado de: [\[Ver aquí\]](#)
- Cochilco. (s.f.). Precio de los Metales. Recuperado de: [\[Ver aquí\]](#)
- Contreras, J. & Idrovo, B. (2009). Un Modelo SARIMA para Predecir la Tasa de Desempleo de Chile. *Cámara Chilena de la construcción*. Recuperado de: [\[Ver aquí\]](#)
- Din, M. (2015). ARIMA by Box Jenkins methodology for estimation and forecasting models in higher education. *Conferencia, Romanian American University of Buchares*. Recuperado de: [\[Ver aquí\]](#)
- Donoso, M. (2013). El mercado del cobre a nivel mundial: evolución, riesgos, características y potencialidades futuras. *Ingeniare. Revista chilena*

de ingeniería, 21(2), 248-261. Recuperado de: [[Ver aquí](#)]

- Engel, E. & Valdés, R. (2001). Prediciendo el precio del cobre: ¿Más allá del camino aleatorio?. *Centro de Economía Aplicada, Universidad de Chile*. Recuperado de: [[Ver aquí](#)]
- Gosset, W. (1908). The probable error of a mean. *Biometrika*, 6(1), 1-25. Recuperado de: [[Ver aquí](#)]
- Peña, D. (2005). Análisis de series temporales. 1ed. Madrid: Alianza editorial, S.A. Recuperado de: [[Ver aquí](#)]
- Servicio Nacional de Geología y Minería. (2020). Anuario de la minería de Chile. Recuperado de: [[Ver aquí](#)]
- US Geological Survey. (2021). Copper. Recuperado de: [[Ver aquí](#)]