MCD – Series de Tiempo

Trabajo final

Integrantes:

Laura Alejandra Guerrero

Pablo Alonzo León

Pablo Antonio Guevara

Ricardo García Ramírez

Tomás Hansen Ibáñez

Julio 2025

Contenido

[Introducción 2](#_Toc1238728955)

[Información del dataset 3](#_Toc1849339233)

[Objetivos 4](#_Toc1139593958)

[Resultados 4](#_Toc346260429)

[a. Análisis exploratorio 5](#_Toc1431613920)

[b. Modelado y selección de modelo 8](#_Toc851180657)

[c. Predicción y visualización 12](#_Toc390578191)

[Conclusión 15](#_Toc648096487)

[Referencias 16](#_Toc45229560)

# **Introducción**

El oro ha sido históricamente un activo financiero clave a nivel global, valorado tanto por su función como reserva de valor como por su capacidad para ofrecer cobertura frente a la inflación y la volatilidad económica, especialmente en contextos de incertidumbre (World Gold Council, 2023). En los últimos años, su cotización ha mostrado una creciente sensibilidad ante factores como la incertidumbre global, los movimientos inflacionarios y las decisiones de política monetaria de las principales economías.

Comprender la dinámica del precio del oro es esencial para inversores, analistas económicos y responsables de políticas públicas. Su análisis permite no solo mejorar la toma de decisiones financieras, sino también anticipar escenarios de riesgo y entender con mayor profundidad la lógica de los mercados en contextos volátiles.

En este contexto, el presente trabajo tiene como objetivo analizar el comportamiento mensual del precio del oro (expresado en USD por onza troy) mediante modelos de series de tiempo. Estos modelos, ampliamente utilizados en finanzas, permiten capturar patrones, tendencias y estacionalidades que explican la evolución del precio a lo largo del tiempo (Box et al., 2015).

# **Información del *dataset***

La calidad y representatividad de los datos son elementos fundamentales en cualquier análisis de series de tiempo, ya que determinan en gran medida la validez de los resultados obtenidos. Para este estudio, se utilizó una base de datos histórica de precios del oro, que proporciona una cobertura temporal amplia y una granularidad adecuada para los objetivos del análisis.

Los datos de los precios del oro se obtuvieron del repositorio público de DataHub (https://datahub.io/core/gold-prices), una plataforma reconocida por ofrecer conjuntos de datos abiertos y confiables. La base de datos original abarca precios mensuales desde el año 1833 hasta la actualidad, lo que garantiza tanto la fiabilidad como la trazabilidad de la información utilizada.

En este trabajo se empleó un subconjunto de dicha base de datos, seleccionando las observaciones correspondientes al periodo comprendido entre enero de 2000 y junio de 2025. Esta ventana temporal fue elegida con el propósito de capturar las dinámicas recientes y relevantes del mercado del oro. La frecuencia de los datos es mensual, y cada punto representa el precio de cierre del oro en dólares estadounidenses (USD).

Es importante señalar que se utiliza el precio de cierre, una métrica estándar en análisis financiero, ya que refleja el valor final de negociación de cada mes. Esta variable es especialmente útil para identificar tendencias y variaciones significativas a lo largo del tiempo.

# **Objetivos**

Este proyecto tiene como propósito desarrollar y aplicar metodologías de series de tiempo para analizar y predecir el comportamiento del precio del oro. Dada su relevancia como activo refugio y su sensibilidad a las condiciones macroeconómicas, la capacidad de anticipar sus movimientos resulta especialmente valiosa en contextos de incertidumbre.

En este sentido, el horizonte de predicción requerido es de 12 meses hacia adelante, lo que permite evaluar el desempeño de los modelos y generar información relevante para la toma de decisiones en escenarios de incertidumbre. Para esto, los objetivos específicos son:

1. Comprender y describir las propiedades de la serie de precios del oro, identificando tendencias, patrones estacionalidades y componentes irregulares.
2. Evaluar la estacionariedad de la serie y, de ser necesario aplicar las transformaciones estadísticas pertinentes para cumplir con los supuestos de los modelos de series de tiempo.
3. Diseñar, ajustar y comparar modelos estadísticos de series de tiempo, específicamente modelos de la familia ARMA y ARIMA, que capturen adecuadamente la estructura temporal de los datos, seleccionando el más adecuado de acuerdo con criterios estadísticos y métricas de error de pronóstico.
4. Realizar predicciones del precio del oro para los próximos 12 meses, y presentando estas estimaciones junto con sus respectivos intervalos de confianza, como una medida de la incertidumbre asociada a las proyecciones.

# **Resultados**

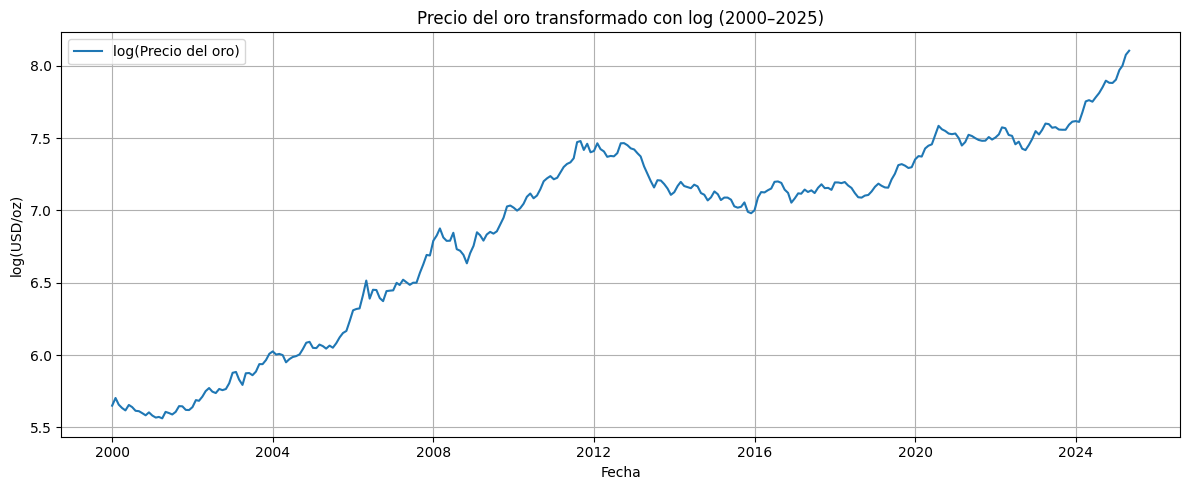
### **a. Análisis exploratorio**

Se observó que el precio del oro presenta una tendencia creciente no lineal, con aceleraciones en períodos de crisis (2008, 2020).



Para estabilizar la varianza de la serie, se aplicó una transformación logarítmica natural. Esta decisión se justifica por tres razones principales:

1. Estabilización de la varianza: La gráfica del precio del oro en escala natural muestra una clara tendencia creciente, acompañada de un aumento en la amplitud de las fluctuaciones. Este patrón sugiere la presencia de heterocedasticidad, es decir, una varianza que crece con el nivel del precio. Al aplicar el logaritmo natural, se estabiliza la varianza relativa, lo cual es deseable para cumplir con los supuestos de modelos lineales como ARIMA o SARIMA.
2. Interpretación económica: Modelar el logaritmo del precio permite interpretar las diferencias como tasas de crecimiento aproximadas. Esto es particularmente útil en series financieras y económicas donde se priorizan los cambios porcentuales más que los absolutos.
3. Linealización de relaciones: En muchos casos, una serie con crecimiento exponencial o multiplicativo puede ser aproximada como aditiva después de una transformación logarítmica, facilitando el modelado con técnicas lineales.



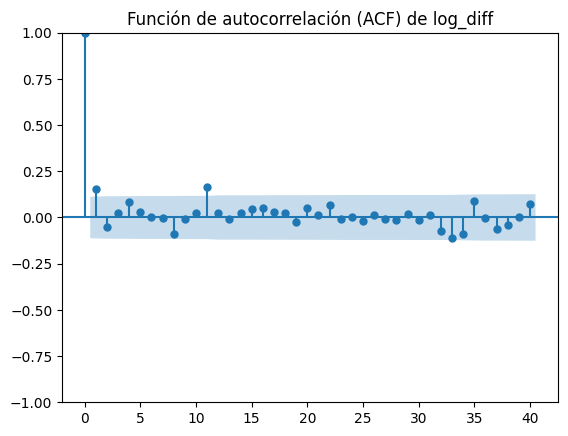
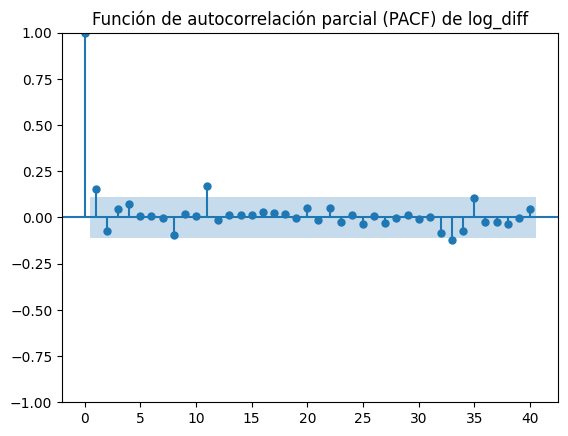
La no estacionariedad de la serie logarítmica del precio del oro fue confirmada mediante la prueba de Dickey-Fuller aumentada (ADF), la cual arrojó un valor p > 0.9. En este contexto, la aplicación de la transformación logarítmica se considera una decisión adecuada y justificada tanto desde una perspectiva estadística, al contribuir a la estabilización de la varianza, como interpretativa, al facilitar el análisis en términos de tasas de crecimiento.

Adicionalmente, se evaluó la posibilidad de emplear modelos SARIMA que incorporan componentes estacionales. Sin embargo, el análisis exploratorio no evidenció patrones estacionales significativos ni persistentes en la serie, tanto en niveles como en log-retornos. Esta conclusión se basa en la inspección de las funciones ACF y PACF, así como en la descomposición estacional. Por consiguiente, se descartó el uso de modelos SARIMA, optando por enfoques más sencillos y acordes a la estructura temporal observada, específicamente modelos de las familias ARMA y ARIMA.

Sin embargo, la serie resultante aún presenta no estacionariedad, por lo que se recomienda aplicar una diferenciación sobre log(Gold\_Price) antes de proceder con el análisis de autocorrelaciones y el modelado.

**Transformación adicional: primera diferencia de log(Gold\_Price)**

Dado que la serie log(Gold\_Price) no es estacionaria según la prueba de Dickey-Fuller, se procede a aplicar una primera diferencia. Esto equivale a modelar los retornos logarítmicos, que en términos económicos representan las tasas de crecimiento relativas mensuales del precio del oro.

Tras aplicar la primera diferencia a log(Gold\_Price), se analizaron las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) sobre la serie log\_diff. Los gráficos resultantes se interpretan a continuación.

**ACF (Función de autocorrelación)**

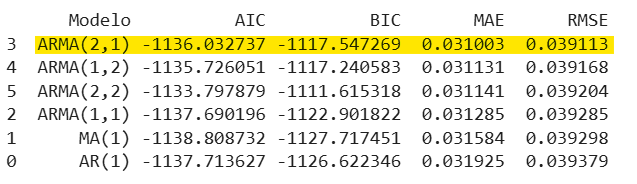
* Se observa un pico significativo en el lag 1, y posteriormente todas las autocorrelaciones caen dentro del intervalo de confianza.
* Este patrón es característico de una serie que puede ser bien modelada con una componente MA(1).

**PACF (Función de autocorrelación parcial)**

* También presenta un solo pico significativo en el lag 1, con el resto de los valores no significativamente diferentes de cero.
* Este comportamiento sugiere la presencia de una estructura autoregresiva de primer orden, es decir, AR(1).

Ambas funciones ACF y PACF indican que la serie de log-retornos del precio del oro posee una estructura temporal simple, que puede ser bien capturada por un modelo ARMA(1,1).

La serie diferenciada log\_diff resultó estacionaria. ACF y PACF indicaron una estructura ARMA(2,1) para los log-retornos, lo cual fue confirmado mediante backtesting con evaluación de RMSE y MAE.



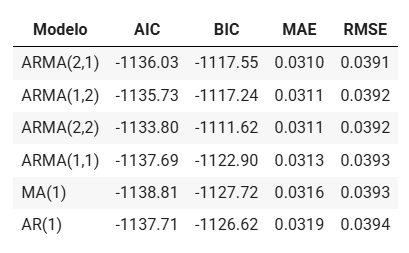
Adicionalmente al enfoque sobre log-retornos, se exploró la posibilidad de modelar la serie de precios del oro en nivel mediante modelos ARIMA con mayor grado de diferenciación, dado que, en la serie original, tal como se evidenció anteriormente, persistían indicios de no estacionariedad incluso después de la transformación logarítmica y la primera diferencia.

Por tal motivo, se evaluaron diferentes especificaciones de modelos ARIMA (p,d,q) con dos diferenciaciones, contrastando su capacidad predictiva y sus métricas de ajuste con los modelos seleccionados sobre log-retornos. Esto permitió comparar de manera robusta qué enfoque resultaba más adecuado para anticipar el comportamiento futuro del precio del oro, considerando la naturaleza de los datos y los objetivos del análisis.

Los resultados y detalles de estos enfoques se presentan a profundidad en la siguiente sección.

### **b. Modelado y selección de modelo**

Se compararon modelos AR(1), MA(1), ARMA(1,1), ARMA(2,1), ARMA(2,2) y otros. El modelo ARMA(2,1) fue seleccionado por tener el mejor equilibrio entre capacidad predictiva (menor MAE y RMSE) y complejidad.



**Análisis**

* Precisión predictiva (MAE y RMSE): El modelo ARMA(2,1) presenta el menor error absoluto medio (MAE = 0.0310) y el menor error cuadrático medio (RMSE = 0.0391), lo cual indica mejor desempeño en la predicción a corto plazo.
* Criterios de información (AIC y BIC): Aunque el modelo MA(1) tiene el mejor AIC (-1138.81), sus métricas de predicción son ligeramente peores. Asimismo, el modelo ARMA(1,1) presenta el segundo mejor AIC, pero no mejora en precisión predictiva respecto a ARMA(2,1).
* Complejidad del modelo: El modelo ARMA(2,1) representa un equilibrio razonable entre complejidad y rendimiento. Modelos más complejos como ARMA(2,2) o ARMA(1,2) no aportan mejoras significativas en desempeño y sí aumentan el número de parámetros.

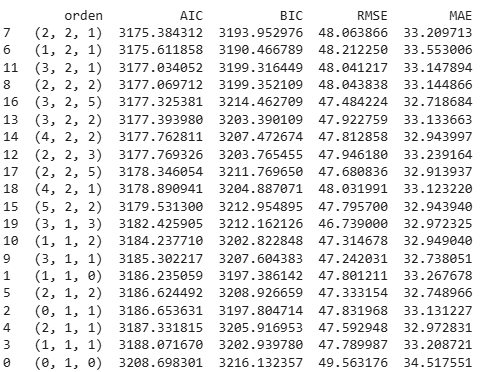
El modelo ARMA(2,1) es el más adecuado para continuar el análisis de los modelos que hemos comparado hasta el momento. Ofrece el mejor rendimiento en predicción a 6 meses y mantiene un nivel moderado de complejidad. Será utilizado para reconstruir el precio del oro en escala real y generar la proyección futura.

**Modelo ARIMA:**

Ahora bien, además del enfoque explorado anteriormente, se implementó un modelo ARIMA directamente sobre la serie mensual de precios del oro. Para identificar la estructura temporal más adecuada, se compararon diez configuraciones ARIMA, evaluando el desempeño de cada modelo principalmente a través de los criterios de información de Akaike (AIC) y Bayesiano (BIC), y considerando adicionalmente los indicadores de error RMSE y MAE para asegurar un análisis más robusto.

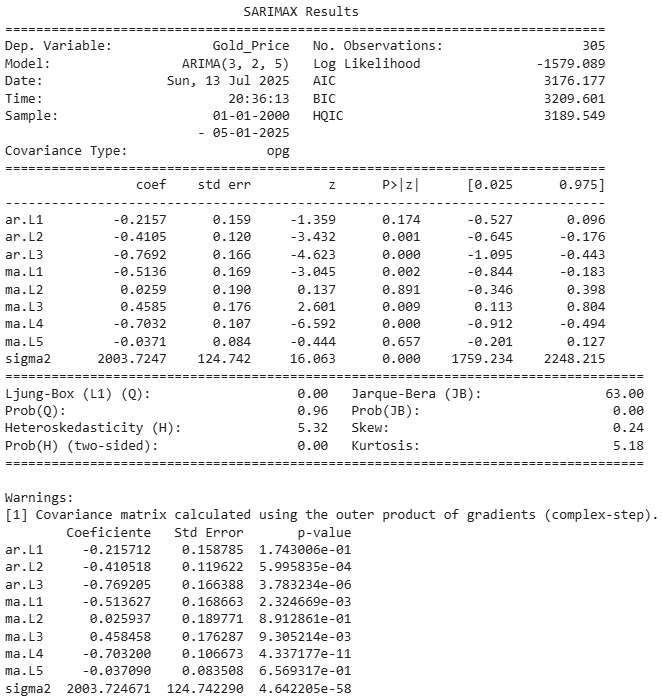
Inicialmente, se probaron modelos básicos como Random Walk (ARIMA(0,1,0)), modelos autorregresivos y de medias móviles simples (ARIMA(1,1,0), ARIMA(0,1,1)), así como una configuración combinada (ARIMA(1,1,1)). Posteriormente, para afinar la selección y garantizar mayor robustez en el análisis, se amplió el set de modelos evaluados incluyendo configuraciones más complejas (por ejemplo, ARIMA(3,2,5), ARIMA(4,2,2), ARIMA(5,2,2)), utilizando una búsqueda sistemática con auto\_arima para identificar el mejor ajuste posible.

La siguiente tabla muestra los resultados obtenidos para cada configuración evaluada, facilitando la comparación objetiva entre alternativas en términos de AIC, BIC, RMSE y MAE:



De acuerdo con estos resultados, el modelo ARIMA(3,2,5) destacó por presentar los menores valores de AIC y BIC dentro del grupo de modelos mejor rankeados, y mostró un desempeño consistente en los indicadores de error RMSE y MAE. Aunque algunas configuraciones alternativas presentaron valores cercanos de AIC o BIC, ARIMA(3,2,5) fue el que obtuvo los errores de predicción (RMSE y MAE) más bajos dentro de ese grupo, lo que evidencia un ajuste más eficiente a la estructura temporal de la serie de precios del oro. Esta selección fue el resultado de comparar veinte combinaciones diferentes de parámetros, lo que asegura un análisis robusto y una mayor confianza en el equilibrio entre calidad de ajuste y precisión de pronóstico.

Dado que el modelo ARIMA(3,2,5) fue el que presentó el mejor desempeño general según los criterios de AIC, BIC, RMSE y MAE, se realizó el análisis detallado de sus resultados. El resumen estadístico muestra que la mayoría de los coeficientes estimados son estadísticamente significativos (valor-p < 0.05), particularmente AR2, AR3, MA1, MA3 y MA4, lo que indica que estos términos aportan información relevante para explicar la dinámica del precio del oro en la serie diferenciada. Sin embargo, algunos coeficientes (AR1, MA2, MA5) no resultaron significativos, aunque se mantienen en el modelo por su contribución al ajuste global.



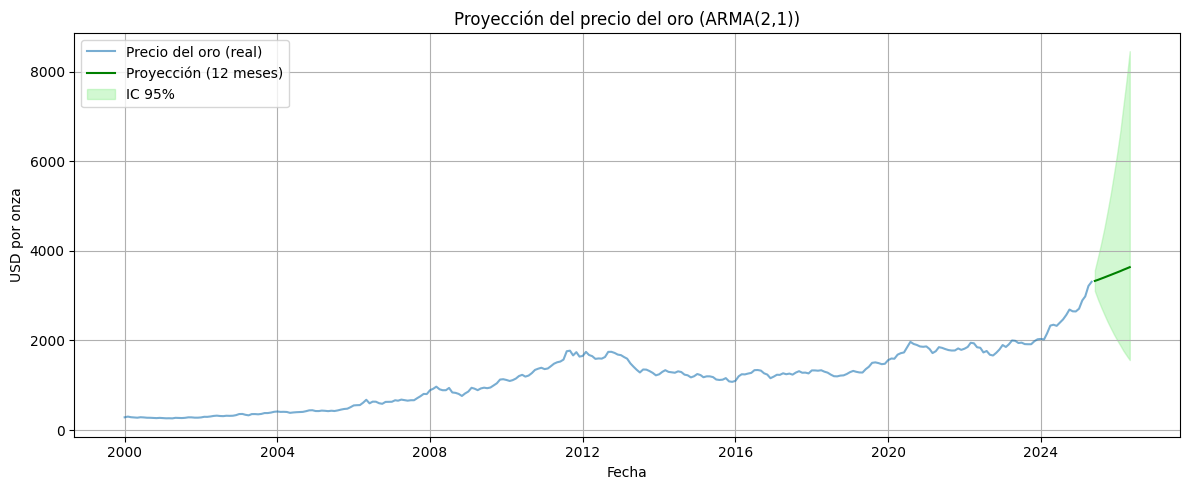
En cuanto a los residuos, aunque el test de Jarque-Bera señala desviaciones importantes respecto a la normalidad (p=0.00), algo habitual en series financieras, y se detecta cierta heterocedasticidad (H=5.32), el análisis de autocorrelación de residuos mediante la prueba de Ljung-Box arroja una probabilidad muy alta (p=0.96), indicando ausencia de autocorrelación significativa. Esta ausencia de autocorrelación refuerza la confianza en las proyecciones, ya que sugiere que el modelo logra captar adecuadamente la dependencia temporal de la serie.

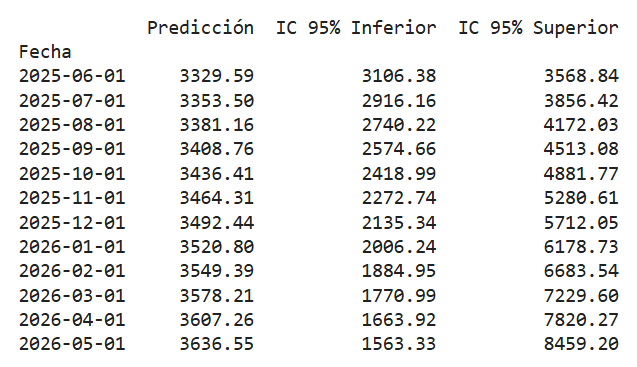
Adicionalmente, el valor del AIC (3176.18) confirma que, dentro del conjunto de modelos evaluados, esta especificación logra el mejor equilibrio entre calidad de ajuste y complejidad, evitando tanto el sobreajuste como un modelo demasiado simple. No obstante, se debe considerar que la presencia de heterocedasticidad y la curtosis observada en los residuos son características habituales en datos financieros y pueden influir en la precisión de los intervalos de confianza del pronóstico.

En conjunto, estos resultados validan el uso del modelo ARIMA(3,2,5) para realizar proyecciones a 12 meses, aunque es importante reconocer las limitaciones propias del contexto financiero y considerar estos aspectos al interpretar las predicciones y la incertidumbre asociada.

### **c. Predicción y visualización**

El modelo ARMA(2,1) fue utilizado para generar una proyección a 12 meses, con intervalos de confianza del 95% calculados a partir de la varianza de los retornos acumulados.





**Interpretación Proyección a 12 meses con intervalos de confianza**

* Proyección puntual: La línea verde representa la predicción puntual del precio del oro para los próximos 12 meses. El modelo proyecta un crecimiento moderado, sin anticipar un cambio de tendencia brusco.
* Intervalos de confianza: El área sombreada en verde claro representa el intervalo de confianza al 95%, que se ensancha de forma significativa a medida que avanza el horizonte temporal. Esto es característico en modelos ARMA, ya que la incertidumbre se acumula con el paso del tiempo.
* Interpretación práctica: Aunque la proyección central parece plausible, la amplitud de los intervalos sugiere alta incertidumbre en el pronóstico. El rango de valores proyectados va desde niveles ligeramente inferiores al valor actual hasta valores que duplican o triplican el precio.
* Conclusión: El modelo proporciona una visión razonable en el corto plazo (3–4 meses), pero su utilidad decrece más allá de ese horizonte. Para predicciones más robustas, podrían considerarse modelos más complejos o incorporar técnicas como SARIMA o modelos basados en redes neuronales recurrentes.

Los resultados muestran que ambos modelos capturan la tendencia creciente del oro, pero con diferentes niveles de suavidad y certeza.

**Modelo ARIMA:**

Una vez validado el ajuste del modelo, se generó el pronóstico para los próximos 12 meses, presentando los resultados tanto en formato tabular como gráfico para facilitar su interpretación. Los valores proyectados mantienen la tendencia creciente registrada en los últimos años. Los intervalos de confianza al 95% permiten dimensionar el rango dentro del cual podría fluctuar el precio, con una mayor amplitud hacia el final del periodo.



La figura posterior ilustra de manera clara cómo el pronóstico sigue la línea alcista reciente del oro, mostrando además la magnitud de la incertidumbre asociada a cada estimación. Este comportamiento es relevante porque, aunque el modelo incorpora la información histórica para proyectar, siempre existe un margen de variabilidad, especialmente en mercados sujetos a shocks externos.

Es importante resaltar que la fuerte dinámica alcista observada en el precio reciente del oro también se refleja en las proyecciones, aunque siempre se debe tener presente que eventos inesperados pueden modificar la tendencia observada. Por ello, la combinación de tabla y gráfica permite comparar fácilmente los valores estimados con el comportamiento histórico, resaltando tanto la tendencia central como la dispersión de los posibles escenarios futuros.



# **Conclusión**

El análisis realizado permitió modelar de manera adecuada la evolución del precio mensual del oro utilizando técnicas de series de tiempo. A partir del análisis exploratorio, se identificó una tendencia creciente no lineal y se justificó el uso de transformaciones logarítmicas y diferenciación para estabilizar la varianza y alcanzar la estacionariedad.

Se evaluaron distintas especificaciones de modelos ARMA y ARIMA, y se seleccionaron aquellos con mejor desempeño en términos de ajuste y capacidad predictiva. El modelo ARMA(2,1) fue el más eficaz para capturar la dinámica de los retornos logarítmicos, mientras que el modelo ARIMA(3,2,5) mostró un ajuste robusto para la serie en niveles. Ambos modelos fueron utilizados para realizar predicciones a 12 meses, presentando sus respectivos intervalos de confianza.

Nuestro análisis mostró que el modelo ARMA(2,1) es una buena herramienta para entender cómo cambian los log-retornos del precio del oro y para hacer predicciones a corto plazo. Este modelo es bueno capturando los movimientos diarios o semanales del precio, lo que vimos en las pruebas que hicimos con datos pasados (conocido como backtesting). El modelo predijo bastante bien el comportamiento del oro en periodos normales.

Sin embargo, notamos una limitación importante: el modelo no es tan preciso cuando el precio del oro sube o baja muy rápido y de forma inesperada. En estas situaciones de cambios bruscos, el ARMA(2,1) tiende a fallar en sus predicciones. Esto significa que, aunque funciona bien la mayor parte del tiempo, necesitamos considerar otros enfoques para esos momentos de gran agitación en el mercado.

En cuanto al análisis directo de la serie original de precios, el modelo ARIMA(3,2,5) resultó el más adecuado según los criterios AIC, BIC, RMSE y MAE, superando al resto de configuraciones probadas. Este modelo fue capaz de capturar la tendencia alcista más reciente y permitió proyectar el comportamiento a 12 meses, incorporando intervalos de confianza para reflejar la incertidumbre. No obstante, entre sus limitaciones se identificó cierta heterocedasticidad y desviaciones a la normalidad en los residuos, lo que sugiere que las proyecciones podrían tener mayor incertidumbre en escenarios extremos o ante shocks inesperados.

En resumen, el ARMA(2,1) es útil para analizar cambios moderados en el precio del oro, pero puede quedarse corto si el mercado se mueve bruscamente. Por otro lado, el ARIMA(3,2,5) permitió prever cómo podría comportarse el precio en el próximo año y nos da una mejor idea de la tendencia general, aunque tampoco puede anticipar cambios drásticos inesperados. Al final, usar ambos modelos da una visión más completa del mercado, pero siempre es clave tener en cuenta que cualquier predicción tiene cierto grado de incertidumbre, sobre todo cuando el mercado es muy volátil.

# **Referencias**

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). Wiley.

World Gold Council. (2023). *Gold as an inflation hedge: What you need to know*. <https://www.gold.org/goldhub/research/gold-inflation-hedge>

Pontificia Universidad Católica de Chile. (2025). *Material de clase para Series de Tiempo*. Material de curso, Maestría en Ciencia de Datos, Pontificia Universidad Católica de Chile.