

UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO
FACULTAD DE CIENCIAS QUÍMICAS, FÍSICAS Y MATEMÁTICAS
ESCUELA PROFESIONAL DE MATEMÁTICA



TESIS

**PRONÓSTICO DE LA PRODUCCIÓN MENSUAL DE COBRE EN
EL PERÚ MEDIANTE MODELOS DE SERIES TEMPORALES
(1999-2024)**

PRESENTADO POR:

Br. ALEXANDRA CONDORI CHOCCE

**PARA OPTAR AL TÍTULO PROFESIONAL DE
LICENCIADA EN MATEMÁTICA MENCIÓN
ESTADÍSTICA**

ASESOR:

Dr. ALFREDO VALENCIA TOLEDO

CUSCO - PERÚ

2025



Universidad Nacional de San Antonio Abad del Cusco

INFORME DE SIMILITUD

(Aprobado por Resolución Nro.CU-321-2025-UNSAAC)

El que suscribe, el Asesor Alfredo Valencia Toledo
..... quien aplica el software de detección de similitud al
trabajo de investigación/tesis titulada: PRONÓSTICO DE LA PRODUCCIÓN
MENSUAL DE COBRE EN EL PERÚ MEDIANTE MODELOS DE
SERIES TEMPORALES (1999 - 2024)

Presentado por: ALEXANDRA CONDORI CHOCCE DNI N° 75710889 ;
presentado por: DNI N°:
Para optar el título Profesional/Grado Académico de LICENCIADA EN MATEMÁTICA
MENCION ESTADÍSTICA

Informo que el trabajo de investigación ha sido sometido a revisión por 2 veces, mediante el
Software de Similitud, conforme al Art. 6° del **Reglamento para Uso del Sistema Detección de**
Similitud en la UNSAAC y de la evaluación de originalidad se tiene un porcentaje de 5 %.

Evaluación y acciones del reporte de coincidencia para trabajos de investigación conducentes a grado académico o título profesional, tesis

Porcentaje	Evaluación y Acciones	Marque con una (X)
Del 1 al 10%	No sobrepasa el porcentaje aceptado de similitud.	X
Del 11 al 30 %	Devolver al usuario para las subsanaciones.	
Mayor a 31%	El responsable de la revisión del documento emite un informe al inmediato jerárquico, conforme al reglamento, quien a su vez eleva el informe al Vicerrectorado de Investigación para que tome las acciones correspondientes; Sin perjuicio de las sanciones administrativas que correspondan de acuerdo a Ley.	

Por tanto, en mi condición de Asesor, firmo el presente informe en señal de conformidad y **adjunto**
las primeras páginas del reporte del Sistema de Detección de Similitud.

Cusco, 20 de noviembre de 2025

Alfredo Valencia Toledo
Firma

Post firma Alfredo Valencia Toledo

Nro. de DNI 43162177

ORCID del Asesor 0000-0001-6505-9634

Se adjunta:

- Reporte generado por el Sistema Antiplagio.
- Enlace del Reporte Generado por el Sistema de Detección de Similitud: oid: 27259:495675984

ALEXANDRA CONDORI CHOCCE

PRONÓSTICO DE LA PRODUCCIÓN MENSUAL DE COBRE EN EL PERÚ MEDIANTE MODELOS DE SERIES TEMPORALES (1999...

 Universidad Nacional San Antonio Abad del Cusco

Detalles del documento

Identificador de la entrega

trn:oid:::27259:495675984

Fecha de entrega

10 sep 2025, 8:03 a.m. GMT-5

Fecha de descarga

10 sep 2025, 8:25 a.m. GMT-5

Nombre del archivo

PRONÓSTICO DE LA PRODUCCIÓN MENSUAL DE COBRE EN EL PERÚ MEDIANTE MODELOS DE SERI....pdf

Tamaño del archivo

2.2 MB

101 páginas

17.416 palabras

88.005 caracteres

5% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para ca...




Filtrado desde el informe

- Bibliografía
- Texto citado
- Texto mencionado
- Coincidencias menores (menos de 10 palabras)

Exclusiones


- N.º de coincidencias excluidas

Fuentes principales

- 4%  Fuentes de Internet
- 1%  Publicaciones
- 4%  Trabajos entregados (trabajos del estudiante)

Marcas de integridad

N.º de alerta de integridad para revisión

-  **Caracteres reemplazados**
335 caracteres sospechosos en N.º de páginas
Las letras son intercambiadas por caracteres similares de otro alfabeto.

Los algoritmos de nuestro sistema analizan un documento en profundidad para buscar inconsistencias que permitirían distinguirlo de una entrega normal. Si advertimos algo extraño, lo marcamos como una alerta para que pueda revisarlo.

Una marca de alerta no es necesariamente un indicador de problemas. Sin embargo, recomendamos que preste atención y la revise.

PRESENTACIÓN

SEÑOR DECANO DE LA FACULTAD DE CIENCIAS QUÍMICAS, FÍSICAS Y MATEMÁTICAS,

SEÑOR DIRECTOR DE LA ESCUELA PROFESIONAL DE MATEMÁTICA,

SEÑORES DOCENTES MIEMBROS DEL JURADO.

Conforme a los lineamientos establecidos por el reglamento de grados y títulos de la Escuela Profesional de Matemática mención Estadística, remito para su consideración y revisión el trabajo de tesis titulado: “PRONÓSTICO DE LA PRODUCCIÓN MENSUAL DE COBRE EN EL PERÚ MEDIANTE MODELOS DE SERIES TEMPORALES (1999-2024)” con el fin de optar al Título Profesional de Licenciado en Matemática mención Estadística.

La presente investigación empleará métodos estadísticos de series temporales SARIMA usando la metodología de Box Jenkins; los cuales permitirán modelar las tendencias estacionales en los datos históricos de producción de cobre. El estudio se basa en datos proporcionados por el Ministerio de Energía y Minas. Para el análisis, se utilizó el software R, seleccionando el modelo que mejor pronostique la producción mensual de cobre en Perú 2024. Este enfoque busca mejorar la precisión de las predicciones y ofrecer herramientas que faciliten una planificación más eficaz y una toma de decisiones estratégicas en el sector minero del país.

Atentamente,

Br. Alexandra Condori Chocce.

DEDICATORIA

A mis queridos abuelos FRANCISCO Y ANICETA, quienes, aunque ya no estén físicamente, siguen viviendo en mi corazón y en cada uno de mis logros. Su amor, enseñanzas y ejemplo han sido una inspiración constante en mi vida. Este trabajo es también un homenaje a su memoria y al legado que dejaron en mí.

Y, finalmente, me dedico esta tesis a mí misma, por cada esfuerzo, por cada desafío superado y por no rendirme a pesar de las dificultades. Por la dedicación y la perseverancia que me han llevado hasta aquí. Este logro es prueba de mi fortaleza y crecimiento, y me lo dedico con orgullo.

Alexandra.

AGRADECIMIENTO

En primer lugar, agradezco a Dios, quien me ha dado la fortaleza, paciencia y sabiduría para recorrer este camino. Sin su guía y bendición, este logro no habría sido posible.

A mi padre, Simón Condori, por su constante insistencia, por recordarme siempre la importancia de concluir lo que se empieza y por ser un pilar fundamental en mi formación. A mi madre, Erlinda Chocce, por su apoyo incondicional a lo largo de este proceso, por su comprensión y por estar siempre presente en cada momento de desafío.

A mi hermano Zhaid, quien ha sido una inspiración en este camino, con la esperanza de que este logro lo motive a seguir sus propios sueños académicos. En especial, a mi hermana Edith, por su confianza inquebrantable en mí y por sus palabras de aliento, que fueron un sostén en los momentos más difíciles.

Expreso mi más sincero agradecimiento al Dr. Alfredo Valencia Toledo, mi asesor de tesis, por su guía, paciencia y valiosos consejos a lo largo de este trabajo. Su compromiso y dedicación fueron fundamentales para el desarrollo de esta investigación, y su conocimiento me permitió crecer académicamente y mejorar cada aspecto de este estudio.

A los docentes del Departamento Académico de Matemática y Estadística de la Universidad Nacional San Antonio Abad del Cusco, quienes, con su enseñanza, contribuyeron significativamente a mi formación. También agradezco a mis familiares, quienes, con su compañía y motivación, me brindaron el ánimo necesario para seguir adelante en los momentos más desafiantes.

Finalmente, a mis queridas mascotas, por ser una fuente inagotable de compañía, alegría y consuelo en los momentos de estrés. Su amor incondicional y su presencia han sido un refugio en este largo camino.

A cada uno de mis amigos y a todas las personas que estuvieron presentes en este recorrido, brindándome su apoyo de una u otra manera, les expreso mi más sincera gratitud.

Alexandra.

ÍNDICE

PRESENTACIÓN	I
DEDICATORIA	II
AGRADECIMIENTO	III
ÍNDICE	IV
ÍNDICE DE FIGURAS	VII
ÍNDICE DE TABLAS	IX
RESUMEN	X
RESUMO	XI
INTRODUCCIÓN	1
CAPÍTULO I	
PLANTEAMIENTO DEL PROBLEMA	3
1.1. Descripción de la situación problemática	3
1.2. Formulación del problema	4
1.2.1. Problema general	4
1.2.2. Problemas específicos	4
1.3. Justificación de la investigación	5
1.4. Objetivos de la investigación	6
1.4.1. Objetivo general	6
1.4.2. Objetivos específicos	6

CAPÍTULO II

MARCO TEÓRICO CONCEPTUAL	7
2.1. Antecedentes	7
2.1.1. Antecedentes internacionales	7
2.1.2. Antecedentes nacionales	8
2.1.3. Antecedentes locales	9
2.2. Bases teóricas	10
2.2.1. Experimento aleatorio	10
2.2.2. Espacio muestral	10
2.2.3. Sigma - álgebra	10
2.2.4. Medida de probabilidad	10
2.2.5. Espacio de probabilidad	11
2.2.6. Variable aleatoria	11
2.2.7. Proceso estocástico	11
2.2.8. Serie de tiempo	13
2.2.9. Metodología de Box Jenkins	41
2.3. Marco conceptual	47

CAPÍTULO III

METODOLOGÍA	49
3.1. Tipo de investigación	49
3.2. Alcance de la investigación	49
3.3. Diseño de investigación	49
3.4. Población de estudio	49
3.5. Técnicas e instrumentos de recolección de datos	50
3.6. Procesamiento de datos	50
3.7. Operacionalización de variable	51

CAPÍTULO IV

RESULTADOS	52
4.1. Exploración de la Producción mensual del cobre en el Perú.	52
4.2. Identificación de las características de la producción total mensual de cobre Perú (1999-2023).	54
4.3. Identificación de modelos de series temporales para la serie de producción mensual de cobre en Perú (1999-2023).	60
4.4. Validación del modelo de la producción mensual de cobre para el año 2024	70
4.5. Pronóstico de la producción mensual de cobre en Perú (Septiembre 2024 - Agosto 2025)	74
DISCUSIÓN	76
CONCLUSIONES	78
RECOMENDACIONES	79
REFERENCIAS	80
ANEXOS	86

ÍNDICE DE FIGURAS

Figura 1:	Característica de tendencia en la serie de tiempo	14
Figura 2:	Característica estacional en la serie de tiempo	14
Figura 3:	Característica aleatoria en la serie de tiempo	15
Figura 4:	Característica cíclica en la serie de tiempo	15
Figura 5:	Característica de heterocedasticidad en la serie de tiempo	16
Figura 6:	Ejemplos de series temporales aditivas	17
Figura 7:	Ejemplos de series temporales multiplicativas	18
Figura 8:	Ruido blanco	20
Figura 9:	Paseo aleatorio	21
Figura 10:	Simulación de series AR(1) y sus funciones ACF y PACF con $\phi = \pm 0,7$. . .	28
Figura 11:	Simulación de series AR(2) y sus funciones ACF y PACF con $\phi_1 = \pm 0,5$ y $\phi_2 = \pm 0,3$	30
Figura 12:	Simulación de series MA(1) y sus funciones ACF y PACF con $\theta = \pm 0,8$. . .	33
Figura 13:	Simulación de series MA(2) y sus funciones ACF y PACF con $\theta_1 = \pm 0,3$ y $\theta_2 = \pm 0,6$	35
Figura 14:	Simulación de series ARMA(1,1) y sus funciones ACF y PACF con $\phi_1 =$ $\pm 0,1$ y $\theta_1 = \pm 0,4$	38
Figura 15:	Simulación de series ARIMA(1,1,1) y sus funciones ACF y PACF con $\phi_1 = 0,5$ y $\theta_1 = 0,3$	39
Figura 16:	Etapas de la metodología de Box Jenkins	42
Figura 17:	Serie Original de la producción total mensual del cobre en Perú 1999-2024 .	52
Figura 18:	Diagrama de cajas de la Producción total mensual de cobre en Perú (1999- 2023)	53
Figura 19:	Diagrama de cajas de la producción total de cobre en Perú (1999-2023) men- sual.	54

Figura 20: Descomposición de la serie de producción de cobre Perú (1999-2023)	56
Figura 21: Correlogramas de la producción total de cobre Perú (1999-2023)	57
Figura 22: Serie logarítmica diferenciada (d=1) y correlogramas de la producción total mensual de cobre Perú (1999-2023)	58
Figura 23: Serie logarítmica diferenciada (D=1) estacional de la producción de cobre Perú (1999-2023)	59
Figura 24: Correlogramas de los residuales del modelo SARIMA2 (0,1,0)(2,1,0)₁₂ . . .	63
Figura 25: Ljung-Box y Gráfico Q-Q de los residuales del modelo SARIMA2 (0,1,0)(2,1,0)₁₂	64
Figura 26: Correlogramas de los residuales del modelo SARIMA5 (1,1,2)(0,1,1)₁₂ . . .	65
Figura 27: Ljung-Box y Gráfico Q-Q de los residuales del modelo SARIMA5 (1,1,2)(0,1,1)₁₂	66
Figura 28: Correlogramas de los residuales del modelo SARIMA8 (2,1,0)(0,1,1)₁₂ . . .	67
Figura 29: Ljung-Box y Gráfico Q-Q de los residuales del modelo SARIMA8 (2,1,0)(0,1,1)₁₂	68
Figura 30: Correlogramas de los residuales del modelo SARIMA10 (2,1,0)(2,1,0)₁₂ . .	69
Figura 31: Ljung-Box y Gráfico Q-Q de los residuales del modelo SARIMA10 (2,1,0)(2,1,0)₁₂	70
Figura 32: Valores pronosticados del modelo 5, ₁₂ de la producción mensual de cobre en Perú, periodo Enero 2024 a Agosto 2024.	71
Figura 33: Valores pronosticados del modelo 8, SARIMA(2,1,0)(0,1,1)₁₂ de la producción mensual de cobre en Perú, periodo Enero 2024 a Agosto 2024.	72
Figura 34: Valores pronosticados del modelo 10, SARIMA(2,1,0)(2,1,0)₁₂ de la producción mensual de cobre en Perú, periodo Enero 2024 a Agosto 2024.	72

ÍNDICE DE TABLAS

Tabla 1:	Operacionalización de la variable	51
Tabla 2:	Estadísticos descriptivos de la producción total de cobre en Perú (1999-2023)	53
Tabla 3:	Pruebas de características de la serie de producción de cobre en el Perú (1999-2023)	55
Tabla 4:	Prueba estadística de tendencia y estacionariedad de la serie logarítmica di- ferenciada	58
Tabla 5:	Modelos propuestos para la serie logarítmica diferenciada (d=1, D=1, s=12)	60
Tabla 6:	Pruebas de validación para los residuos del modelo SARIMA2 (0,1,0)(2,1,0)₁₂	63
Tabla 7:	Pruebas de validación para los residuos del modelo SARIMA5(1,1,2)(0,1,1)₁₂	65
Tabla 8:	Pruebas de validación para los residuos del modelo SARIMA8 (2,1,0)(0,1,1)₁₂	67
Tabla 9:	Pruebas de validación para los residuos del modelo SARIMA10 (2,1,0)(2,1,0)₁₂	69
Tabla 10:	Valores pronosticados para la producción total mensual de cobre para el año 2024.	71
Tabla 11:	Comparación de la evaluación de los errores de pronósticos de los modelos SARIMA propuestos.	73
Tabla 12:	Pronóstico de la producción de cobre según el modelo SARIMA(2,1,0)(2,1,0)₁₂	75

RESUMEN

El objetivo de esta investigación fue identificar el mejor modelo de series de tiempo para pronosticar la producción mensual de cobre en Perú durante el periodo 1999-2024, utilizando datos proporcionados por el Ministerio de Energía y Minas (MINEM). Se adoptó un enfoque tanto teórico como aplicado, con un diseño no experimental, longitudinal y retrospectivo. Se empleó la metodología de Box-Jenkins y se estableció un nivel de confianza del 95 % en las pruebas de hipótesis.

El análisis de la serie temporal mostró una tendencia creciente, estacionalidad anual (de orden 12) y heterocedasticidad en la variabilidad anual. De los doce modelos SARIMA evaluados, tres superaron la prueba de Ljung-Box, lo que sugiere una adecuada independencia de los residuos. Sin embargo, la prueba de Jarque-Bera indicó que los residuos no seguían una distribución normal, lo que podría afectar la precisión de los pronósticos.

De los modelos evaluados, el SARIMA(2,1,0)(2,1,0)₁₂ se destacó como el mejor, con un error porcentual absoluto medio (MAPE) de 5.14 %, lo que lo convierte en la mejor opción dentro del enfoque Box-Jenkins. Este modelo proporciona una herramienta robusta y confiable para la predicción de la producción mensual de cobre en Perú, contribuyendo a la toma de decisiones en el sector minero.

Palabras Claves: SARIMA, Box-Jenkins, Series de tiempo, Producción de cobre

RESUMO

O objetivo desta pesquisa foi identificar o melhor modelo de séries temporais para prever a produção mensal de cobre no Peru durante o período de 1999 a 2024, utilizando dados fornecidos pelo Ministério de Energia e Minas (MINEM). Adotou-se uma abordagem tanto teórica quanto aplicada, com um delineamento não experimental, longitudinal e retrospectivo. Empregou-se a metodologia Box-Jenkins e estabeleceu-se um nível de confiança de 95 % nas provas de hipóteses.

A análise da série temporal revelou uma tendência crescente, sazonalidade anual (de ordem 12) e heterocedasticidade na variabilidade anual. Dos doze modelos SARIMA avaliados, três passaram no teste de Ljung-Box, indicando independência adequada dos resíduos. No entanto, o teste de Jarque-Bera mostrou que os resíduos não seguiam uma distribuição normal, o que pode afetar a precisão das previsões.

Entre os modelos analisados, o SARIMA(2,1,0)(2,1,0)₁₂ apresentou o melhor desempenho, com um erro percentual absoluto médio (MAPE) de 5,14%, consolidando-se como a melhor opção dentro da abordagem Box-Jenkins. Esse modelo oferece uma ferramenta robusta e confiável para a previsão da produção mensal de cobre no Peru, contribuindo para a tomada de decisões no setor minerador.

Palavras-chave: SARIMA, Box-Jenkins, Séries temporais, Produção de cobre

INTRODUCCIÓN

El sector minero juega un papel crucial como motor del crecimiento del Producto Bruto Interno (PBI) del país (Instituto Peruano de Economía, 2024), entre el 2017 y 2021 el sector minero aportó el 16 % del PBI nacional (Ministerio de energía y minas, 2024). En particular, el cobre se destaca como el principal producto minero de exportación de Perú, y sus exportaciones han sido fundamentales para el crecimiento económico del país en los últimos años (Montalvan & Silva, 2020), en Enero del 2024 las exportaciones peruanas crecieron 12.6 % y alcanzaron los US\$ 5,263.4 millones, así lo reportó el Centro de investigación de economía y negocios globales (2024). Por ejemplo: en 2023, Perú alcanzó un récord en la producción de cobre, con 2,755,066 toneladas métricas finas (TMF), marcando un aumento del 12.7 % en comparación con el año anterior. Esto consolidó al país como el segundo mayor productor y exportador de cobre en el mundo (Instituto de Ingenieros de Minas del Perú, 2024).

Los pronósticos son herramientas valiosas que proporcionan estimaciones cuantitativas sobre la probabilidad de eventos futuros, permitiendo una mejor planificación y toma de decisiones basada en información predictiva (Contreras et al., 2016). Los modelos propuestos por Box y Jenkins (1976) son ampliamente utilizados en el análisis de series temporales estacionarias, ya que permiten identificar y capturar patrones como tendencias y estacionalidades. Estas metodologías proporcionan herramientas robustas para analizar patrones en los datos y realizar estimaciones confiables en contextos diversos.

El objetivo principal de este estudio es determinar el modelo de serie temporal que mejor se ajuste para el pronóstico de la producción mensual de cobre en Perú. Para ello, se empleará una metodología que abarcará desde el análisis exploratorio de los datos hasta la aplicación de modelos estadísticos de series temporales, como la metodología de Box-Jenkins. Este análisis permitirá identificar patrones y tendencias en la producción, mejorando la precisión de las estimaciones y proporcionando herramientas útiles para la planificación estratégica en el sector minero.

El presente trabajo de investigación está estructurado de la siguiente manera:

CAPÍTULO I: Presenta la situación problemática, la justificación del estudio, así como los pro-

blemas de investigación y los objetivos generales y específicos que guían este trabajo.

CAPÍTULO II: Exposición de los antecedentes relevantes para la investigación, bases teóricas que sustentan el estudio y el marco conceptual necesario para comprender los términos y enfoques empleados.

CAPÍTULO III: Descripción de la metodología empleada en el desarrollo de la investigación, incluyendo el diseño del estudio, población de estudio, el tipo de investigación y la operacionalización de las variables.

CAPÍTULO IV: Presentación de los resultados obtenidos a partir del análisis de los datos recopilado en relación con los objetivos establecidos en el estudio.

Finalmente, se presentan la discusión de los hallazgos obtenidos, las conclusiones derivadas del estudio y las recomendaciones orientadas a futuras investigaciones o aplicaciones prácticas.

CAPÍTULO I

PLANTEAMIENTO DEL PROBLEMA

1.1. Descripción de la situación problemática

El sector minero es un pilar fundamental de la economía peruana, a pesar que el cobre se destaca como uno de los principales minerales exportados en el país, no se han desarrollado estudios específicos que aborden el pronóstico de la producción mensual de cobre en el país. Esta falta de análisis limita la capacidad del sector minero y del gobierno para anticipar fluctuaciones en la producción, dificultando la planificación efectiva y la respuesta estratégica frente a cambios en el mercado global.

El Perú es reconocido mundialmente como uno de los principales productores de cobre, siendo este mineral un pilar fundamental en su economía debido a su abundancia y demanda global (Ministerio de Energía y Minas del Perú, 2022; U.S. Geological Survey, 2023). Desde enero de 1999, la producción de cobre ha experimentado un crecimiento sostenido impulsado por la expansión de grandes minas como Cerro Verde, Antamina y Las Bambas, así como por la inversión en tecnologías y mejoras en la eficiencia productiva (Instituto Nacional de Estadística e Informática, 2021; Suárez, 2020).

Durante las primeras dos décadas del siglo XXI, la producción mostró fluctuaciones relacionadas con la volatilidad de los precios internacionales del cobre, el contexto político y social interno, y eventos globales como la crisis financiera de 2008 (Banco Mundial, 2019). Sin embargo, en la última década, el país ha consolidado su posición mediante la ampliación de proyectos mineros y la mejora en la infraestructura, alcanzando niveles récords de extracción y exportación hacia 2023 (Banco Central de Reserva del Perú, 2023; Ministerio de Energía y Minas del Perú, 2023).

Este crecimiento ha sido clave para la economía peruana, representando una significativa fuente de ingresos fiscales y divisas, además de influir en el desarrollo regional de las zonas mineras (Organización para la Cooperación y el Desarrollo Económicos, 2018). No obstante, la producción también enfrenta retos como la sostenibilidad ambiental y la gestión social en las comunidades afectadas (Gonzales, 2019).

Según la revisión literaria, se han llevado a cabo estudios relacionados con series temporales en el sector minero, pero ninguna se enfoca directamente en la producción mensual de cobre en Perú. Como el análisis del estudio realizado por Contreras Cerpa (2021) pronosticó la producción de cobre en Chile mediante modelos de redes neuronales, el estudio realizado por Cotrina Teatino et al. (2022) realizó pronósticos para el precio del cobre mediante pos promedio simples y alisado exponencial, otro estudio realizado por Henríquez et al. (2022) aplicó la metodología de Box-Jenkins para el pronóstico de la exportación minera de los principales metales en Perú. Estos estudios no abordan la dinámica particular de la producción de cobre en Perú, lo que evidencia una brecha de conocimiento en esta área.

El pronóstico preciso de las series de producción mensual no solo permitiría una mejor planificación operativa por parte de las empresas mineras, sino que también facilitaría la toma de decisiones estratégicas a nivel gubernamental.

1.2. Formulación del problema

1.2.1. Problema general

- ¿Cuál es el modelo de series de tiempo que mejor se ajusta para pronosticar la producción mensual de cobre en Perú con datos históricos (1999-2023) y validar con datos de Enero a Agosto del 2024?

1.2.2. Problemas específicos

- ¿Cuáles son las características presentes en la serie de producción mensual de cobre en Perú (1999-2023)?
- ¿Cuáles son los modelos de series temporales que tienen mejor ajuste para la serie de producción mensual de cobre en Perú desde 1999 a Diciembre del 2023?
- ¿Cuál es el modelo que tiene mejor capacidad de validación de la producción mensual de cobre para los meses del último año (Enero - Agosto 2024)?

1.3. Justificación de la investigación

Desde un punto de vista teórico, este estudio busca aportar al campo de la industria minera, mediante la aplicación de diferentes modelos de pronóstico y técnicas de suavizamiento para predecir la producción mensual de cobre. Aunque existen investigaciones previas en este sector relacionadas con la economía, los estudios específicos sobre la producción de cobre en Perú son escasos. Con este estudio esperamos encontrar el modelo que pronostique mejor la producción mensual de cobre y poder brindar un marco teórico que aborde aspectos que puedan ser replicados en otros contextos dentro del sector minero.

Desde una perspectiva económica, este estudio es importante debido a que la minería representa una de las principales fuentes de ingresos para el país. Pronosticar con precisión la producción de cobre tiene un impacto directo en la economía nacional, ya que permite a las empresas mineras, inversores y al gobierno tomar decisiones informadas basadas en pronósticos confiables. Al determinar el modelo predictivo más adecuado, este estudio ofrecerá herramientas valiosas para optimizar la gestión de recursos y mejorar la eficiencia en la toma de decisiones económicas. Esto podría llevar a una mejor planificación de la producción y a un aumento en la competitividad del país en el mercado global del cobre.

Desde el punto de vista social, este estudio tiene una relevancia significativa, dado que la minería es uno de los pilares económicos de Perú, y el cobre es uno de los minerales mas relevantes en términos de producción y exportación, donde las comunidades mineras dependen directamente de esta actividad. Los resultados de este estudio podrían contribuir a una mejor planificación, gestión de políticas públicas y estrategias empresariales en el sector, minimizando los impactos sociales negativos asociados con la explotación minera, como los problemas de salud pública y los conflictos sociales. Pronosticar con precisión la producción de cobre, puede ayudar a contribuir a la estabilidad social permitiendo a los actores económicos y gubernamentales anticipar y prepararse mejor para los cambios en la producción minera.

En ese sentido, existe la necesidad de estudiar el pronóstico de las series de producción mensual del cobre, con este trabajo pretendemos llenar este vacío de conocimiento, donde pretendemos estudiar

las series de producción mensual de cobre aplicando modelos de series temporales, mejorando así la precisión en el pronóstico y poder facilitar una planificación más efectiva y una toma de decisión estratégica en el sector minero nacional.

1.4. Objetivos de la investigación

1.4.1. Objetivo general

- Determinar el modelo de series de tiempo que mejor se ajusta para pronosticar la producción mensual de cobre en Perú con datos históricos (1999-2023) y validación con datos de Enero y Agosto del 2024.

1.4.2. Objetivos específicos

- Identificar los características presentes en la serie de producción mensual de cobre en Perú (1999-2023).
- Analizar los modelos de series temporales que ofrecen el mejor ajuste para la serie de producción mensual de cobre en Perú desde 1999 a Diciembre del 2023.
- Determinar el modelo con la mejor capacidad de validación de la producción mensual de cobre para los meses del último año (Enero - Agosto 2024).

CAPÍTULO II

MARCO TEÓRICO CONCEPTUAL

2.1. Antecedentes

2.1.1. Antecedentes internacionales

Un estudio reciente de Mutele y Carranza (2024) analizó las tendencias en la producción de oro en Sudáfrica y a nivel mundial desde 1990 hasta 2022, utilizando modelos ARIMA, VAR y ARNN. El modelo ARIMA (1,1,0) demostró ser el más eficaz según las métricas de precisión RME, MAE y MAPE; sus proyecciones sugirieron un incremento anual del 0.6 % en Sudáfrica y del 2.2 % a nivel mundial.

Por otro lado, Contreras Cerpa (2021) aplicó modelos de redes neuronales recurrentes (RNN), GRU y LSTM para prever la producción de cobre, considerando el consumo de agua, energía y las emisiones de gases de efecto invernadero. El modelo LSTM ofreció el mejor desempeño y un ajuste más preciso, seguido por los modelos GRU y RNN.

Astudillo et al. (2020) exploraron la regresión del vector de soporte (SVR) para pronosticar los precios en la Bolsa de Metales de Londres a plazos de 5, 10, 15, 20 y 30 días. Utilizando validación cruzada y búsqueda de cuadrícula, el modelo SVR mostró ser robusto y preciso, con un error cuadrático medio (RMSE) igual o inferior al 2.2 % para predicciones a 5 y 10 días.

El estudio realizado por Oktaviani et al. (2021) pronosticó las anomalías de las temperaturas de la superficie del mar (SSTA) con patrones estacionales, donde emplearon el modelo SARIMA combinado con GARCH, utilizando datos diarios de julio de 2010 a julio de 2018. El mejor modelo identificado fue SARIMA (2,1,1)(0,1,1) - GARCH(1,1), con una precisión destacada (MAPE de 3.67 %), demostrando su efectividad para pronosticar SSTA y contribuir a mitigar los efectos del cambio climático.

2.1.2. Antecedentes nacionales

Aycaya-Paco et al. (2023) desarrollaron una herramienta en R-Studio para estudiar y predecir la extracción de minerales en los distintos departamentos de Perú. Utilizaron el modelo ARIMA, proyectaron la extracción de minerales hasta 2027, estimando cifras como 2,694,957 TMF de cobre, etc.

Según la investigación de Cotrina Teatino et al. (2022), se realizó un estudio para predecir el precio del cobre en libras (lb). Donde aplicaron diversos modelos de predicción de series de tiempo, para evaluar el ajuste de estos modelos, se empleó el coeficiente de determinación (R^2) y el error porcentual absoluto medio (MAPE) como medidas de precisión. Los resultados de este análisis revelaron que el método de suavizado exponencial simple fue el más adecuado, prediciendo un precio del cobre de 3.60 lb para septiembre de 2022, con un MAPE de 4,4 % y un (R^2) de 0,86.

En el estudio realizado por Henríquez et al. (2022), estableció un modelo de pronóstico para la exportación minera en Perú, analizando los cambios ocasionados por la pandemia de COVID-19. Utilizaron la metodología de Box-Jenkins para construir un modelo autoregresivo integrado de medias móviles (ARIMA). Se verificó la estacionariedad con la prueba de raíz unitaria de Dickey-Fuller ($p = 0.027$), la normalidad de los residuos con la prueba de Jarque-Bera ($p = 0.229$), la no autocorrelación con el test de Durbin-Watson ($d = 2.01$) y la homocedasticidad de los residuos con Q-Stat de Ljung-Box. Finalmente, concluyeron que el modelo autoregresivo de primer orden (1, 0, 0) con la ecuación $y_t = 2270,176 + 0,749596y_{t-1} + t$ se ajusta adecuadamente a la exportación minera en Perú durante el período estudiado.

De acuerdo con el estudio realizado por Chafloque Céspedes et al. (2018), se enfocaron en modelar, estimar y analizar las exportaciones mineras en Perú desde 1985 hasta 2017. El principal objetivo del estudio fue aplicar metodologías de predicción y suavizamiento para pronosticar las exportaciones mineras a corto plazo. Compararon los métodos predictivos ARIMA y SARIMA. Los resultados revelaron que el modelo SARIMA demostró ser el más adecuado para la predicción de las exportaciones mineras, superando al modelo ARIMA en términos de precisión predictiva.

El objetivo del trabajo realizado por Orihuela et al. (2023) fue el de analizar la relación entre las

exportaciones de materias primas de oro y cobre y la balanza de pagos en Perú (BP). Se estimó un Modelo de Vectores Autorregresivos (VAR). Para este modelo, se aplicaron las pruebas de Raíz Unitaria de Dickey Fuller Aumentada, Cointegración de Johansen y Causalidad de Granger. Los resultados indicaron que: Las exportaciones de oro y cobre y la Balanza de Pagos mantienen una relación de equilibrio a largo plazo, estando cointegradas.

2.1.3. Antecedentes locales

En la investigación llevada a cabo por Mamani Quispe (2019), se analizó el impacto de la exportación de cobre en el crecimiento económico de Perú durante el período de 2010 a 2018. Los hallazgos demostraron que las exportaciones de cobre tienen un efecto significativo en el crecimiento del PBI del país: un incremento del 10 % en la producción de cobre resultaría en un aumento del 0.38 % en el PBI. Asimismo, se observó que un aumento del 10 % en el precio del cobre hace 12 meses llevaría a un incremento del 0.27 % en el PBI del mes actual.

2.2. Bases teóricas

2.2.1. Experimento aleatorio

Un experimento que puede proporcionar diferentes resultados, aunque sea repetido todas las veces de la misma manera, es llamado un *experimento aleatorio* (Montgomery & Runger, 2012).

2.2.2. Espacio muestral

El espacio muestral, denotado por Ω , es la colección o totalidad de todos los resultados posibles de un experimento aleatorio (Mood et al., 1974).

2.2.3. Sigma - álgebra

Según Grimmett y Stirzaker (2001), \mathcal{F} es una colección de subconjuntos de Ω , es decir $\mathcal{F} \subseteq \mathcal{P}(\Omega)$, que es llamado sigma álgebra, denotado por σ -álgebra, si satisface las siguientes propiedades:

1. $\emptyset \in \mathcal{F}$,
2. Cerradura bajo complementación: Si $A \in \mathcal{F}$ entonces $A^c \in \mathcal{F}$,
3. Cerradura bajo uniones numerables: Si $A_1, A_2, \dots \in \mathcal{F}$ entonces $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

2.2.4. Medida de probabilidad

Una *medida de probabilidad* \mathbb{P} en (Ω, \mathcal{F}) es una función $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ satisfaciendo (Grimmett & Stirzaker, 2001):

1. $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$;
2. Si A_1, A_2, \dots es una colección de conjuntos disjuntos de \mathcal{F} en que $A_i \cap A_j = \emptyset$ para todos los pares i, j satisfaciendo $i \neq j$, entonces:

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

2.2.5. Espacio de probabilidad

Según Grimmett y Stirzaker (2001), la terna $(\Omega, \mathcal{F}, \mathbb{P})$, de un conjunto Ω , un σ -álgebra \mathcal{F} del subconjunto de Ω , y una medida de probabilidad \mathbb{P} en (Ω, \mathcal{F}) , es llamada un *espacio de probabilidad*.

2.2.6. Variable aleatoria

Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad. Una **variable aleatoria** es una función

$$X : \Omega \longrightarrow \mathbb{R}$$

tal que, para todo número real x , el este conjunto es un evento medible.

$$[\omega \in \Omega : X(\omega) = x] \in \mathcal{F} \text{ para cada } x \in \mathbb{R}$$

(Billingsley, 1995).

La variable aleatoria se clasifica en:

- **Variable aleatoria discreta:** Si el conjunto de resultados posibles, R_X , es finito ó infinito numerable, decimos que X es una variable aleatoria discreta.
- **Variable aleatoria continua:** Una variable aleatoria es continua cuando el conjunto de sus valores posibles son todos los valores de un intervalo o de una unión de intervalos de números reales.

2.2.7. Proceso estocástico

Un **proceso estocástico** es una colección de variables aleatorias indexadas en el tiempo, donde cada variable describe el estado de un sistema en un momento específico. Este proceso modela la evolución de un fenómeno que cambia de manera aleatoria a lo largo del tiempo (Billingsley,

1995).

Según Morettin y Toloi (2006) sea T un conjunto arbitrario. Un **proceso estocástico** es una familia $\{X_t, t \in T\}$, tal que, para cada $t \in T$, X_t es una variable aleatoria. Un **proceso estocástico** es una familia de variables aleatorias, que podemos definir en un mismo espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$. Sea $t = 1, 2, \dots, n$ elementos de T se considera las distribuciones finito-dimensionales, para todo $n \geq 1$:

$$F(x_1, \dots, x_n, 1, \dots, n) = P\{X_1 \leq x_1, \dots, X_n \leq x_n\} \quad (1)$$

Si $n = 1$ se conoce las distribuciones unidimensionales de variables aleatorias X_1 , y para $n = 2$ distribuciones bidimensionales de variables aleatorias (X_1, X_2) , la función de distribución (1) debe satisfacer las condiciones:

1. Simetría para cualquier permutación j_1, \dots, j_n de los índices $1, 2, \dots, n$ tenemos:

$$F(x_{j_1}, \dots, x_{j_n}, t_{j_1}, \dots, t_{j_n}) = F(x_1, \dots, x_n, 1, \dots, n)$$

2. Condición de compatibilidad. para $m < n$

$$F(X_1, \dots, x_m, +\infty, \dots, +\infty; 1, \dots, m, m+1, \dots, n) = F(x_1, \dots, x_m; 1, \dots, m)$$

- Un proceso estocástico $\{X_t, t \in T\}$ se dice que es *estrictamente estacionario* si todas las distribuciones finito-dimensionales (1) permanecen en las mismas sub traslaciones del tiempo; es decir que para el conjunto de $t = 1, 2, \dots, n$ y cualquier desplazamiento k :

$$(X_1, X_2, \dots, X_n) \stackrel{d}{=} (X_{1+k}, X_{2+k}, \dots, X_{n+k})$$

las distribuciones unidimensionales son invariantes en el tiempo, con media y varianza constante.

$$\mu_t = 0 \quad \text{Var}(X_t) = \sigma^2$$

- Un proceso estocástico $\{X_t, t \in T\}$ se dice que es *débilmente estacionario* o estacionario de segundo orden si y solamente si cumple:

i) $E[X_t] = \mu_t = \mu$ constante para todo $t \in T$.

ii) $\text{Var}(X_t) < \infty$ para todo $t \in T$.

iii) $\gamma(t_1, t_2) = \text{Cov}(X_{(t_1)}, X_{(t_2)})$, es una función de $|t_1 - t_2|$

El cual para este estudio será denominado *procesos estacionarios*

2.2.8. Serie de tiempo

De acuerdo con Brockwell y Davis (2016), una *serie de tiempo* es un conjunto de observaciones X_t , donde cada una se registra en un momento específico t . Una serie de tiempo discreto es aquella en la que el conjunto de tiempos T_0 en los que se hacen observaciones es un conjunto discreto. Una serie de tiempo continua es obtenida cuando las observaciones se registran continuamente durante intervalo de tiempo.

Shumway y Stoffer (2016), define una *serie de tiempo* como una colección de variables aleatorias según el orden que se encuentren en el tiempo t .

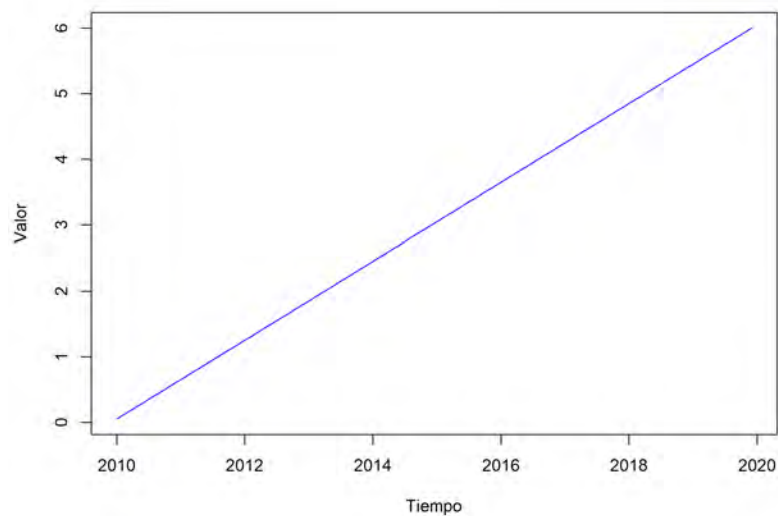
Una *serie de tiempo* es una secuencia de observaciones ordenadas cronológicamente lo largo del tiempo. Desde un punto de vista probabilístico, cada observación puede considerarse como una realización de un *proceso estocástico* (Box et al., 2016).

2.2.8.1. Características de la serie de tiempo. Ríos (2008) define las características como componentes, donde indica que una serie de tiempo puede descomponerse en cuatro componentes que no son directamente observables:

1. **Tendencia(T):** Representa el comportamiento predominante de la serie, es definida como el cambio de la media a lo largo de un extenso periodo de tiempo.

Figura 1

Característica de tendencia en la serie de tiempo

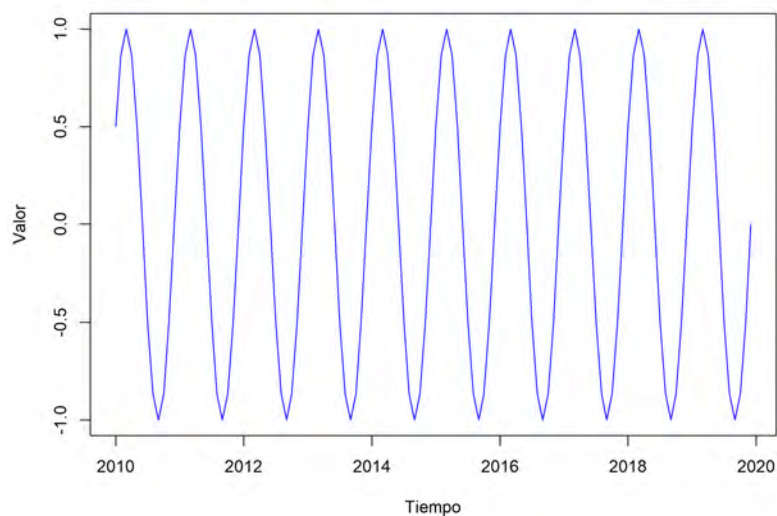


Nota: Elaboración propia con datos simulados en R-Studio.

2. **Estacionalidad (E):** Movimiento periódico que se producen dentro de un periodo corto y conocido, son oscilaciones estrictamente periódicas, donde el período es igual o inferior al año.

Figura 2

Característica estacional en la serie de tiempo

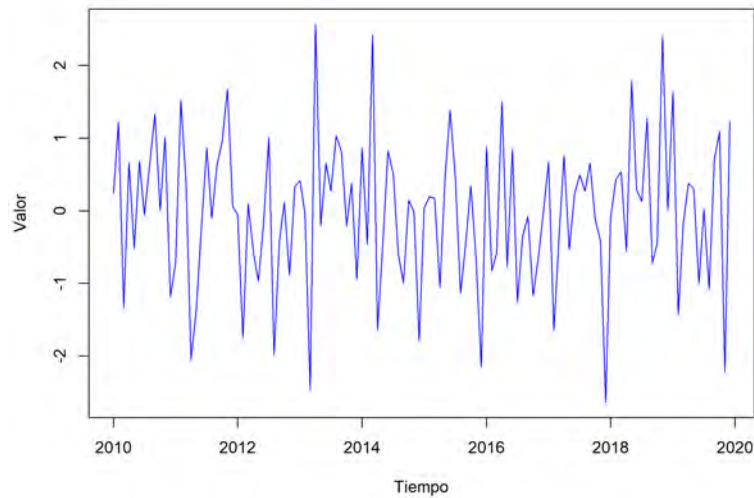


Nota: Elaboración propia con datos simulados en R-Studio.

3. **Aleatorio (A):** Son movimientos erráticos que no siguen un patrón específico, este componente es prácticamente impredecible.

Figura 3

Característica aleatoria en la serie de tiempo

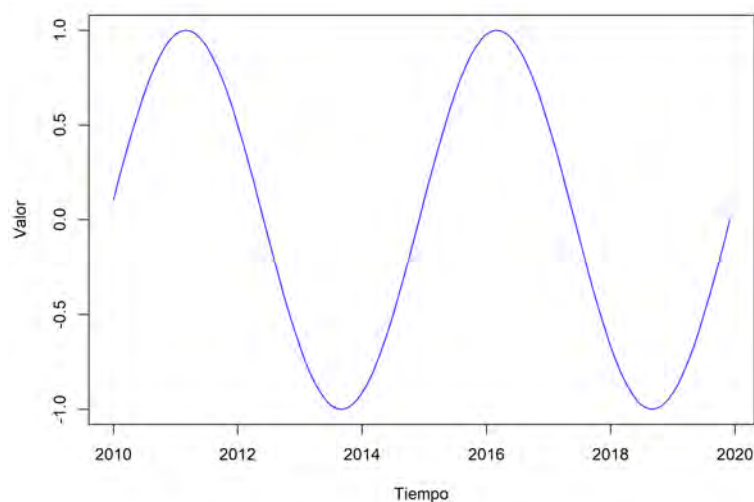


Nota: Elaboración propia con datos simulados en R-Studio.

4. **Ciclo (C):** Fluctuaciones recurrentes de duración variable, asociadas a fenómenos económicos, climáticos o sociales, que carecen de periodicidad fija y pueden variar en amplitud (Box et al., 2016).

Figura 4

Característica cíclica en la serie de tiempo



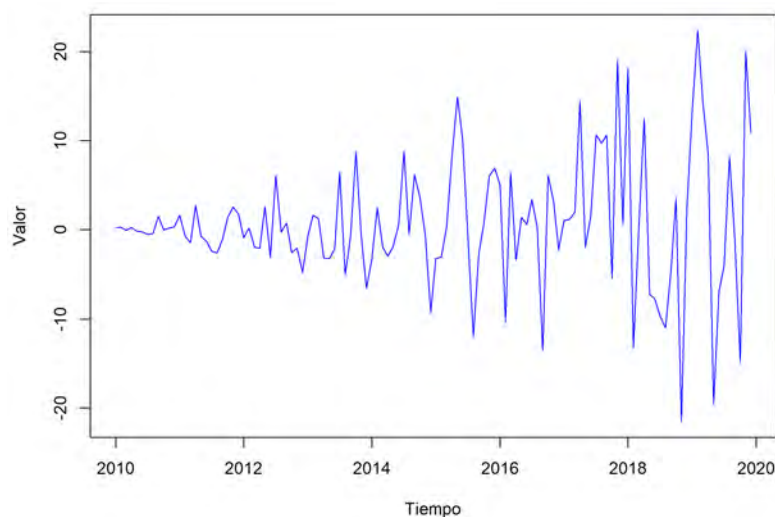
Nota: Elaboración propia con datos simulados en R-Studio.

Hamilton (1994) explica la heterocedasticidad en el contexto de series de tiempo sobre todo en el área financiera.

- 5) **Heterocedasticidad:** Se refiere a la variación no constante de la varianza de los errores de predicción a lo largo del tiempo.

Figura 5

Característica de heterocedasticidad en la serie de tiempo



Nota: Elaboración propia con datos simulados en R-Studio.

2.2.8.2. Transformaciones de Box Cox. Según Morettin y Toloi (2006) existen dos razones por las cuales se debe transformar los datos originales: estabilizar la varianza y conseguir el efecto estacional aditivo, en estos casos la transformación logarítmica puede ser la adecuada. En el caso de series económicas y financieras indica que puede ser necesario aplicar alguna transformación no lineal, como logarítmica o la transformación de la forma:

$$X_t^{(\lambda)} = \begin{cases} \frac{X_t^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \ln(X_t), & \lambda = 0 \end{cases}$$

llamada transformación de Box y Cox (1964), un método para estabilizar la varianza y aproximar los datos a una distribución normal. Box y Cox (1964) propusieron determinar el valor óptimo de

λ mediante un análisis basado en máxima verosimilitud, optimizando el ajuste de los datos a un modelo estadístico con supuestos de homocedasticidad y normalidad.

2.2.8.3. Descomposición de una serie temporal. Según Montgomery et al. (2008) existe un enfoque clásico para la descomposición de una serie temporal en componentes de tendencia y estacionalidad.

El modelo matemático general para esta descomposición es:

$$X_t = f(E_t, T_t, \varepsilon_t)$$

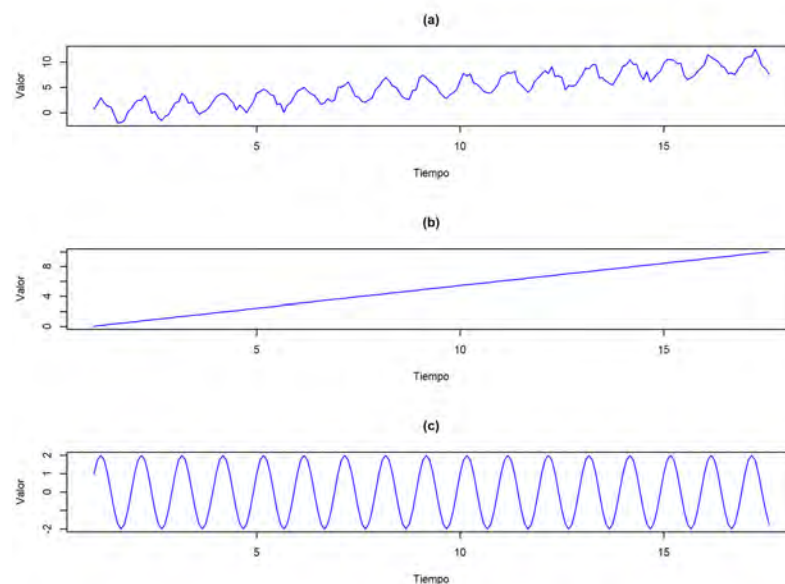
donde: E_t es el componente estacional, T_t es el componente de tendencia y ε_t es el componente de error aleatorio.

Generalmente existen dos formas para la función f , una es el modelo aditivo

$$X_t = E_t + T_t + \varepsilon_t$$

Figura 6

Ejemplos de series temporales aditivas



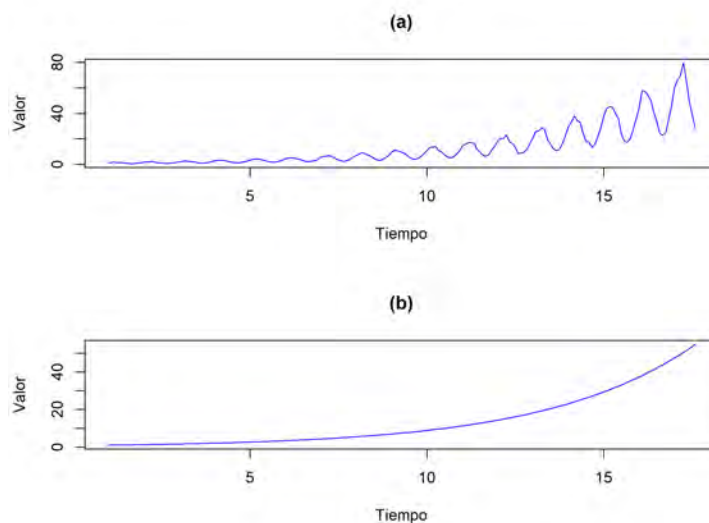
(a) Serie temporal con componentes de tendencia y estacionalidad. (b) Serie temporal con una tendencia creciente lineal, sin estacionalidad. (c) Serie temporal con estacionalidad y sin tendencia.

Nota: Elaboración propia con datos simulados en R-Studio.

y la otra modelo multiplicativo

Figura 7

Ejemplos de series temporales multiplicativas



(a) *Serie temporal con tendencia y estacionalidad. Las amplitudes de las fluctuaciones aumentan con el tiempo, indicando la presencia de heterocedasticidad.* (b) *Serie temporal con una tendencia creciente exponencial, sin estacionalidad.*

Nota: Elaboración propia con datos simulados en R-Studio.

El modelo aditivo Figura (6) es apropiado si la magnitud (amplitud) de la variación estacional no varía con el nivel de la serie, mientras que la versión multiplicativa es más apropiada si la amplitud de las fluctuaciones estacionales aumenta o disminuye con el nivel promedio de la serie temporal.

2.2.8.4. Media y varianza. Según Cryer y Chan (2008) para un proceso estocástico $X_t : t = 0, \pm 1, \pm 2, \pm 3, \dots$ la *función media* esta definida por:

$$\mu = E(X_t) = \int_{-\infty}^{\infty} xf(x) dx$$

para $t = 0, \pm 1, \pm 2, \pm 3, \dots$

Es decir, μ es simplemente el valor esperado del proceso en el tiempo t . En general, μ puede ser diferente en cada momento t .

La varianza esta dada por:

$$Var(X_t) = \gamma_0 = E(X_t - \mu)^2 = \int_{-\infty}^{\infty} (X - \mu)^2 f(x) dx = \sigma^2$$

2.2.8.5. Función de autocovarianza y autocorrelación. Montgomery et al. (2008) indica que si una serie de tiempo es estacionaria, significa que la distribución de probabilidad conjunta de dos observaciones cualesquiera, sea X_t y X_{t+k} es el mismo para dos periodos de tiempo t y $t + k$, que son separados por el mismo intervalo k , llamado *lag*.

La covarianza entre X_t y su valor con otro periodo de tiempo X_{t+k} es llamado la *autocovarianza de lag k* definido por:

$$\gamma_k = \text{Cov}(X_t, X_{t+k}) = E[(X_t - \mu)(X_{t+k} - \mu)]$$

la colección de los valores de γ_k , $k = 0, 1, 2, 3, \dots$ es llamado la *función de autocovarianza*, la autocovarianza de *lag k* = 0, es justo la varianza de la serie, $\gamma_0 = \sigma^2$

La *función de autocorrelación* (ACF) en *lagk* es la colección de los valores de ρ_k , $k = 0, 1, 2, 3, \dots$; para $k = 0$ entonces $\rho_0 = 1$, lo cual indica que la variables esta correlacionada consigo misma.

$$\rho_k = \text{Corr}(X_t, X_{t+k}) = \frac{\text{Cov}(X_t, X_{t+k})}{\text{Var}(X_t)} = \frac{\gamma_k}{\gamma_0}$$

La *función de autocorrelación parcial* (PACF) de un proceso estacionario X_t , denotado por α_k , para $k = 0, 1, 2, 3, \dots$, proporciona la relación directa que existe entre dos observaciones separadas por k retardos, sin tomar en cuenta los retardos intermedios (Cotrina Araujo, 2020), es definida como:

$$\alpha_k = \text{Corr}(X_t, X_{t-k}/X_{t-1}, X_{t-2}, \dots, X_{t-k+1})$$

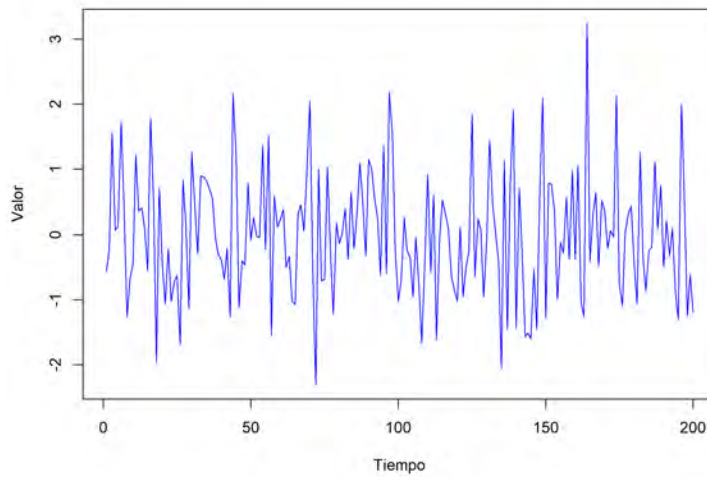
$$\alpha_k = \frac{\text{Cov}(X_t - \hat{X}_t, X_{t-k} - \hat{X}_{t-k})}{\sqrt{\text{Var}(X_t - \hat{X}_t)\text{Var}(X_{t-k} - \hat{X}_{t-k})}}$$

2.2.8.6. Ruido blanco. Box et al. (2016) expresa que un proceso estacionario es una secuencia de variables aleatorias independientes e idénticamente distribuidas (idd), denotadas como a_1, \dots, a_t, \dots , que también asumimos que tienen media cero y varianza σ_a^2 . Este proceso es estrictamente estacionario y se denomina proceso de *ruido blanco*. Debido a que la independencia implica que los a_t no están correlacionados; es decir incorreladas, donde su función de autocovarianza es simplemente:

$$\gamma_k = E[a_t a_{t+k}] = \begin{cases} \sigma_a^2 & \text{si } k = 0 \\ 0 & \text{si } k \neq 0 \end{cases}$$

Figura 8

Ruido blanco

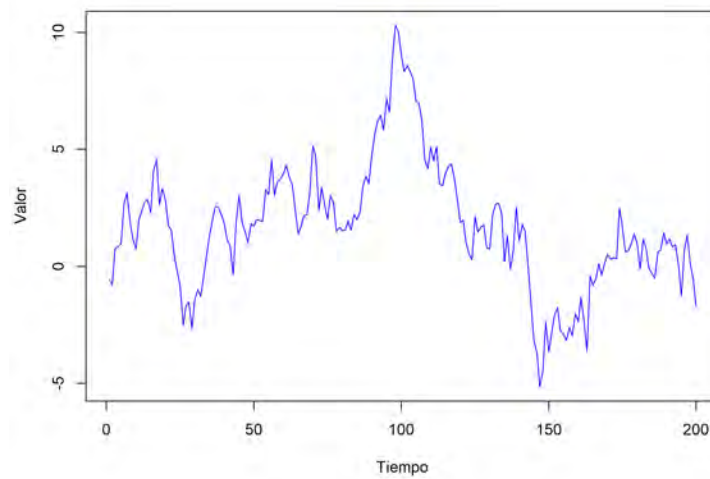


Nota: Elaboración propia con datos simulados en R-Studio.

2.2.8.7. Paseo aleatorio. Conforme a Brockwell y Davis (2016), el paseo aleatorio $S_t, t = 0, 1, 2, \dots$, que comienza en cero, se forma al sumar de manera acumulativa variables aleatorias independientes e idénticamente distribuidas (iid). De este modo, un paseo aleatorio con media cero se define estableciendo $S_0 = 0$ y

$$S_t = a_1 + a_2 + \dots + a_t,$$

para $t = 1, 2, \dots$, donde a_t representa el ruido idénticamente independiente.

Figura 9*Paseo aleatorio**Nota: Elaboración propia con datos simulados en R-Studio.*

2.2.8.8. Pruebas de hipótesis para las características de una serie de tiempo. Según Morrettin y Toloi (2006) existen algunas pruebas no paramétricas que son útiles para probar las características de una serie de tiempo: para la tendencia, tenemos la prueba de Mann-Kendall; para la estacionalidad, Kruskal-Wallis; para la estacionariedad, la prueba de Dickey-Fuller; y para la heterocedasticidad, la prueba de Levene. Estos en general se basan en hipótesis que pueden no estar verificadas para el caso de una serie temporal.

1. Pruebas de hipótesis de tendencia

■ Prueba de Mann-Kendall

H_0 : No existe tendencia

H_1 : Existe tendencia

Según Teixeira-Gandra et al. (2014) utiliza toda la serie, donde compara todos los pares posibles de observaciones a lo largo del tiempo.

Sea

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sign}(x_j - x_i)$$

Donde la función signo $\text{sign}(x_j - x_i)$ se define como:

$$\text{sign}(x_j - x_i) = \begin{cases} 1 & \text{si } x_j - x_i > 0, \text{ tendencia creciente} \\ 0 & \text{si } x_j - x_i = 0, \text{ sin tendencia} \\ -1 & \text{si } x_j - x_i < 0, \text{ tendencia decreciente} \end{cases}$$

2. Pruebas de Hipótesis de Estacionalidad

■ Prueba de Kruskal-Wallis

H_0 : No existe estacionalidad

H_1 : Existe estacionalidad

Esta prueba consiste en dividir la serie en grupos basados en estaciones del año (mensuales, trimestrales, etc). Dichas observaciones X_t son sustituidas por sus puntos R_i , obtenidas ordenándose todas las N observaciones.

Sea

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1)$$

donde:

k : Número de grupos.

R_i : Suma de rangos en el grupo i .

n_i : Número de observaciones en el grupo i .

3. Pruebas de hipótesis de estacionariedad

■ Prueba de hipótesis de Dickey Fuller (prueba de raíz unitaria)

$$H_0 : \rho = 1; \text{ (es no estacionaria)}$$

$$H_1 : \rho < 1; \text{ (es estacionaria)}$$

La prueba de Dickey-Fuller es un procedimiento estadístico utilizado para verificar la presencia de una raíz unitaria en una serie temporal, lo que indica no estacionariedad.

Parte del modelo autorregresivo de primer orden:

$$Y_t = \rho Y_{t-1} + a_t,$$

donde a_t es un término de error con media cero y varianza constante. Para contrastar la hipótesis de raíz unitaria, el modelo se reescribe como:

$$\Delta Y_t = \alpha Y_{t-1} + u_t,$$

donde: $\Delta Y_t = Y_t - Y_{t-1}$ representa la primera diferencia de la serie y $\alpha = \rho - 1$. (Dickey & Fuller, 1979)

■ Prueba de Hipótesis de Heterocedasticidad

• Prueba de Levene

$$H_0 : \text{Existe homocedasticidad}$$

$$H_1 : \text{Existe heterocedasticidad}$$

El estadístico de prueba de Levene se define como: (Correa et al., 2006)

$$W = \frac{(N - k) \sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2}{(k - 1) \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2}$$

donde:

k = número de grupo.

n_i = número de observaciones en el grupo i .

N = número total de las observaciones,

\bar{X} = es la media general.

X_{ij} = es la j -ésima observación en el grupo i .

\bar{X}_i = es la media en el grupo i .

2.2.8.9. Procesos lineales estacionarios. Un proceso lineal general $\{X_t\}$, es aquel que puede representarse como una combinación lineal ponderada de términos de ruido blanco presentes y pasados, de la siguiente manera:

$$X_t = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots$$

Shumway y Stoffer (2016), define al operador de rezago como:

$$BX_t = X_{t-1}$$

que pueden ser extendidas a potencias $B^2 X_t = B(BX_t) = BX_{t-1} = X_{t-2}$, por lo tanto de manera general el operador de rezago esta dado por:

$$B^k X_t = X_{t-k}$$

Las diferencias de orden d , son definidas como:

$$\nabla^d = (1 - B)^d$$

a) Proceso autoregresivo AR(p)

Los procesos autoregresivo son, como su nombre indica, regresiones sobre sí mismos.

Específicamente, un proceso autoregresivo de orden p , X_t satisface la ecuación: (Shumway & Stoffer, 2016)

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \phi_2(X_{t-2} - \mu) + \dots + \phi_{t-p}(X_{t-p} - \mu) + a_t \quad (2)$$

el valor actual de la serie X_t , es una combinación lineal de los p valores pasados mas recientes de sí misma mas un termino a_t (error o ruido blanco), cuando $\mu \neq 0$

$$X_t = \delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + a_t \quad (3)$$

donde: $\delta = \mu(1 - \phi_1 - \phi_2 - \dots - \phi_p)$

La ecuación (2) también puede ser definido en términos de su desviación de la siguiente manera:

$$\bar{X}_t = \phi_1 \bar{X}_{t-1} + \phi_2 \bar{X}_{t-2} + \dots + \phi_p \bar{X}_{t-p} + a_t$$

El modelo (3) puede ser representado utilizando un polinomio en el operador de rezago B (también conocido como operador de retardo), de la siguiente manera:

$$\phi(B)X_t = a_t + \delta$$

donde:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

■ Media

$$E[X_t] = E[\delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + X_{t-p} + a_t]$$

$$E[X_t] = \delta + \mu(\phi_1 + \phi_2 + \dots + \phi_p)$$

$$\text{donde: } \mu = \frac{\delta}{1 - (\phi_1 + \phi_2 + \dots + \phi_p)}, \text{ con } \phi_1 + \phi_2 + \dots + \phi_p \neq 1$$

■ Varianza

$$Var[X_t] = E[(\bar{X}_t)^2] = E[(\phi_1 \bar{X}_{t-1} + \phi_2 \bar{X}_{t-2} + \cdots + \phi_p \bar{X}_{t-p} + a_t) \bar{X}_t]$$

$$Var[X_t] = \gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \cdots + \phi_p \gamma_p + \sigma_a^2$$

■ Autocovarianzas

$$\gamma_1 = E[\bar{X}_t \bar{X}_{t-1}] = \phi_1 \gamma_0 + \phi_2 \gamma_1 + \cdots + \phi_p \gamma_{p-1},$$

$$\gamma_2 = E[\bar{X}_t \bar{X}_{t-2}] = \phi_1 \gamma_1 + \phi_2 \gamma_0 + \cdots + \phi_p \gamma_{p-2},$$

$$\vdots$$

$$\gamma_p = E[\bar{X}_t \bar{X}_{t-p}] = \phi_1 \gamma_{p-1} + \phi_2 \gamma_{p-2} + \cdots + \phi_p \gamma_0,$$

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \cdots + \phi_p \gamma_{k-p}, \text{ para todo } k \geq p+1$$

■ Autocorrelaciones

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \phi_1 + \phi_2 \rho_1 + \cdots + \phi_p \rho_{p-1},$$

$$\rho_2 = \frac{\gamma_2}{\gamma_0} = \phi_1 \rho_1 + \phi_2 + \cdots + \phi_p \rho_{p-2},$$

$$\vdots$$

$$\rho_p = \frac{\gamma_p}{\gamma_0} = \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \cdots + \phi_p,$$

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p}, \text{ para todo } k \geq p+1.$$

1) Proceso autoregresivo de primer orden AR(1)

El proceso de primer orden es:

$$X_t = \delta + \phi_1 X_{t-1} + a_t$$

en términos de su desviación se define de la siguiente manera:

$$\bar{X}_t = \phi_1 \bar{X}_{t-1} + a_t$$

■ Media

$$E[X_t] = E[\delta + \phi_1 X_{t-1} + a_t] = \delta + \phi_1 E[X_{t-1}] + E[a_t] = \delta + \phi_1 \mu$$

$$\text{donde: } \mu = \frac{\delta}{1 - \phi_1}, \text{ con } \phi_1 \neq 1$$

■ Varianza

$$Var[X_t] = E[(X_t - \mu)^2] = E[(\bar{X}_t)^2] = E[(\phi_1 \bar{X}_{t-1} + a_t)^2] = \phi_1^2 \gamma_0 + \sigma_a^2$$

$$Var[X_t] = \gamma_0 = \frac{\sigma_a^2}{1 - \phi_1^2}, \text{ con } \phi_1 \neq \pm 1$$

■ Autocovarianzas

$$\gamma_1 = E[\bar{X}_t \bar{X}_{t-1}] = E[(\phi_1 \bar{X}_{t-1} + a_t) \bar{X}_{t-1}] = \phi_1 \gamma_0,$$

$$\gamma_2 = E[\bar{X}_t \bar{X}_{t-2}] = E[(\phi_1 \bar{X}_{t-1} + a_t) \bar{X}_{t-2}] = \phi_1 \gamma_1 = \phi_1^2 \gamma_0,$$

$$\vdots$$

$$\gamma_k = \phi_1 \gamma_{k-1} = \phi_1^k \gamma_0.$$

■ Autocorrelaciones

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \phi_1,$$

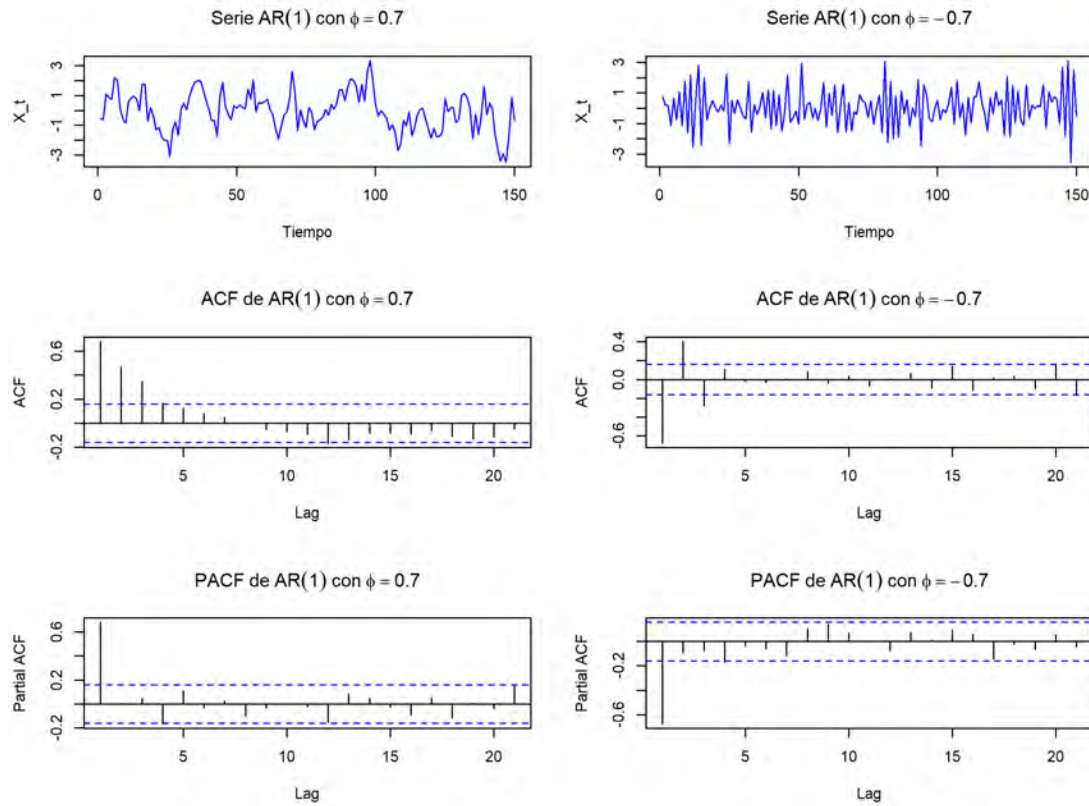
$$\rho_2 = \frac{\gamma_2}{\gamma_0} = \phi_1^2,$$

$$\vdots$$

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi_1^k.$$

Figura 10

Simulación de series AR(1) y sus funciones ACF y PACF con $\phi = \pm 0,7$



Nota: Elaboración propia con datos simulados en R-Studio.

2) Proceso autoregresivo de segundo orden AR(2)

El proceso de segundo orden esta definido como:

$$X_t = \delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + a_t$$

en términos de su desviación se define:

$$\bar{X}_t = \phi_1 \bar{X}_{t-1} + \phi_2 \bar{X}_{t-2} + a_t$$

■ Media

$$E[X_t] = E[\delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + a_t] = \delta + \mu(\phi_1 + \phi_2)$$

donde: $\mu = \frac{\delta}{1 - (\phi_1 + \phi_2)}$, con $\phi_1 + \phi_2 \neq 1$

■ **Varianza**

$$Var[X_t] = E[(X_t - \mu)^2] = E[(\bar{X}_t)^2] = E[(\phi_1 \bar{X}_{t-1} + \phi_2 \bar{X}_{t-2} + a_t)^2]$$

$$Var[X_t] = \gamma_0 = \phi_1^2 \gamma_0 + \phi_2^2 \gamma_0 + \sigma_a^2 + 2\phi_1 \phi_2 \gamma_1$$

■ **Autocovarianzas**

$$\gamma_1 = E[\bar{X}_t \bar{X}_{t-1}] = \phi_1 \gamma_0 + \phi_2 \gamma_1 = \frac{\phi_1}{1 - \phi_2} \gamma_0,$$

$$\gamma_2 = E[\bar{X}_t \bar{X}_{t-2}] = \phi_1 \gamma_1 + \phi_2 \gamma_0,$$

\vdots

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}, \text{ para todo } k \geq 2$$

■ **Autocorrelaciones**

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\phi_1}{1 - \phi_2},$$

$$\rho_2 = \frac{\gamma_2}{\gamma_0} = \frac{\phi_1 \gamma_1 + \phi_2 \gamma_0}{\gamma_0} = \phi_1 \rho_1 + \phi_2,$$

\vdots

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, \text{ para todo } k > 2.$$

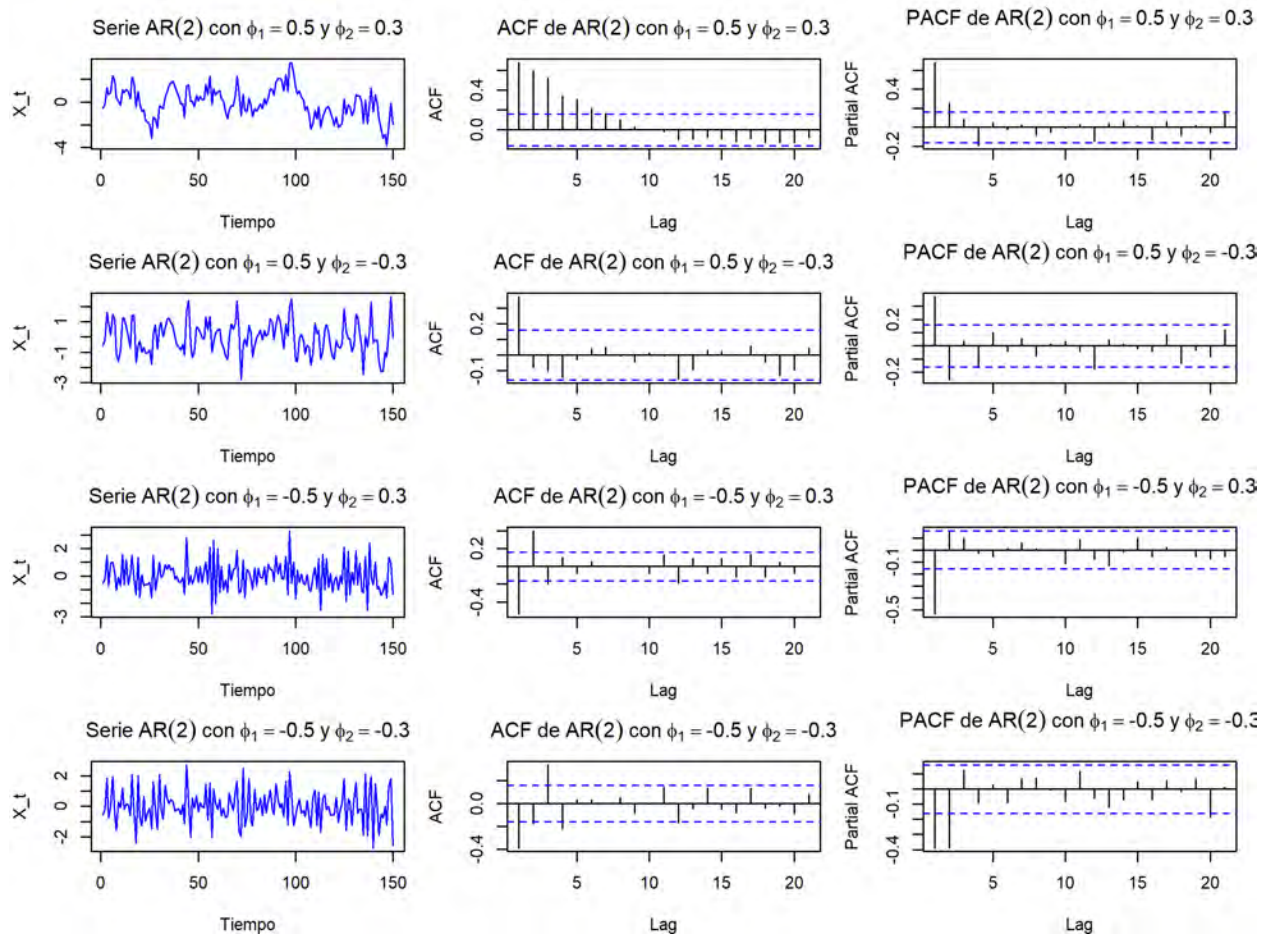
b) Proceso de media móvil MA(q)

Los procesos de media móvil utiliza los valores pasados de los errores, llamados también residuos. El proceso de media móvil de orden q, X_t estará definida como:

$$X_t = \mu + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q} + a_t$$

Figura 11

Simulación de series AR(2) y sus funciones ACF y PACF con $\phi_1 = \pm 0,5$ y $\phi_2 = \pm 0,3$



Nota: Elaboración propia con datos simulados en R-Studio.

Cuando $\mu \neq 0$, su representación polinomial es:

$$X_t = \theta(B)a_t + \mu$$

donde:

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$$

■ Media

$$E[X_t] = E[\mu + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q} + a_t] = \mu$$

■ **Varianza**

$$\text{Var}[X_t] = E[(X_t - \mu)^2],$$

$$\text{Var}[X_t] = E[(\theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q} + a_t)^2]$$

$$\text{Var}[X_t] = \gamma_0 = (\theta_1^2 + \theta_2^2 + \dots + \theta_q^2 + 1)\sigma_a^2$$

■ **Autocovarianzas**

$$\gamma_1 = E[\bar{X}_t \bar{X}_{t-1}] = (\theta_1 + \theta_1 \theta_2 + \dots + \theta_{q-1} \theta_q) \sigma_a^2,$$

$$\gamma_2 = E[\bar{X}_t \bar{X}_{t-2}] = (\theta_2 + \theta_1 \theta_3 + \dots + \theta_{q-2} \theta_q) \sigma_a^2$$

$$\vdots$$

$$\gamma_k = E[\bar{X}_t \bar{X}_{t-k}] = (\theta_k + \theta_1 \theta_{k+1} + \dots + \theta_{q-k} \theta_q) \sigma_a^2, \text{ para todo } k < q$$

$$\gamma_q = \theta_q \sigma_a^2.$$

■ **Autocorrelaciones**

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\theta_1 + \theta_1 \theta_2 + \dots + \theta_{q-1} \theta_q}{\theta_1^2 + \theta_2^2 + \dots + \theta_q^2 + 1},$$

$$\rho_2 = \frac{\gamma_2}{\gamma_0} = \frac{\theta_2 + \theta_1 \theta_3 + \dots + \theta_{q-2} \theta_q}{\theta_1^2 + \theta_2^2 + \dots + \theta_q^2 + 1},$$

$$\vdots$$

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\theta_k + \theta_1 \theta_{k+1} + \dots + \theta_{q-k} \theta_q}{\theta_1^2 + \theta_2^2 + \dots + \theta_q^2 + 1}, \text{ para todo } k < q,$$

$$\rho_q = \frac{\theta_q}{\theta_1^2 + \theta_2^2 + \dots + \theta_q^2 + 1}.$$

1) Proceso de media móvil de primer orden MA(1)

El proceso de primer orden es:

$$X_t = \mu + \theta_1 a_{t-1} + a_t$$

■ Media

$$E[X_t] = E[\mu + \theta_1 a_{t-1} + a_t] = \mu + \theta_1 E[a_{t-1}] + E[a_t] = \mu$$

■ Varianza

$$Var[X_t] = E[(X_t - \mu)^2] = E[(\theta_1 a_{t-1} + a_t)^2] = \theta_1^2 \sigma_a^2 + \sigma_a^2$$

$$Var[X_t] = \gamma_0 = \sigma_a^2(\theta_1^2 + 1)$$

■ Autocovarianzas

$$\gamma_1 = E[(X_t - \mu)(X_{t-1} - \mu)] = E[(\theta_1 a_{t-1} + a_t)(\theta_1 a_{t-2} + a_{t-1})] = \theta_1 \sigma_a^2,$$

$$\gamma_2 = E[(X_t - \mu)(X_{t-2} - \mu)] = E[(\theta_1 a_{t-1} + a_t)(\theta_1 a_{t-3} + a_{t-2})] = 0,$$

\vdots

$$\gamma_k = 0, \text{ para todo } k \geq 2.$$

■ Autocorrelaciones

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\theta_1}{\theta_1^2 + 1},$$

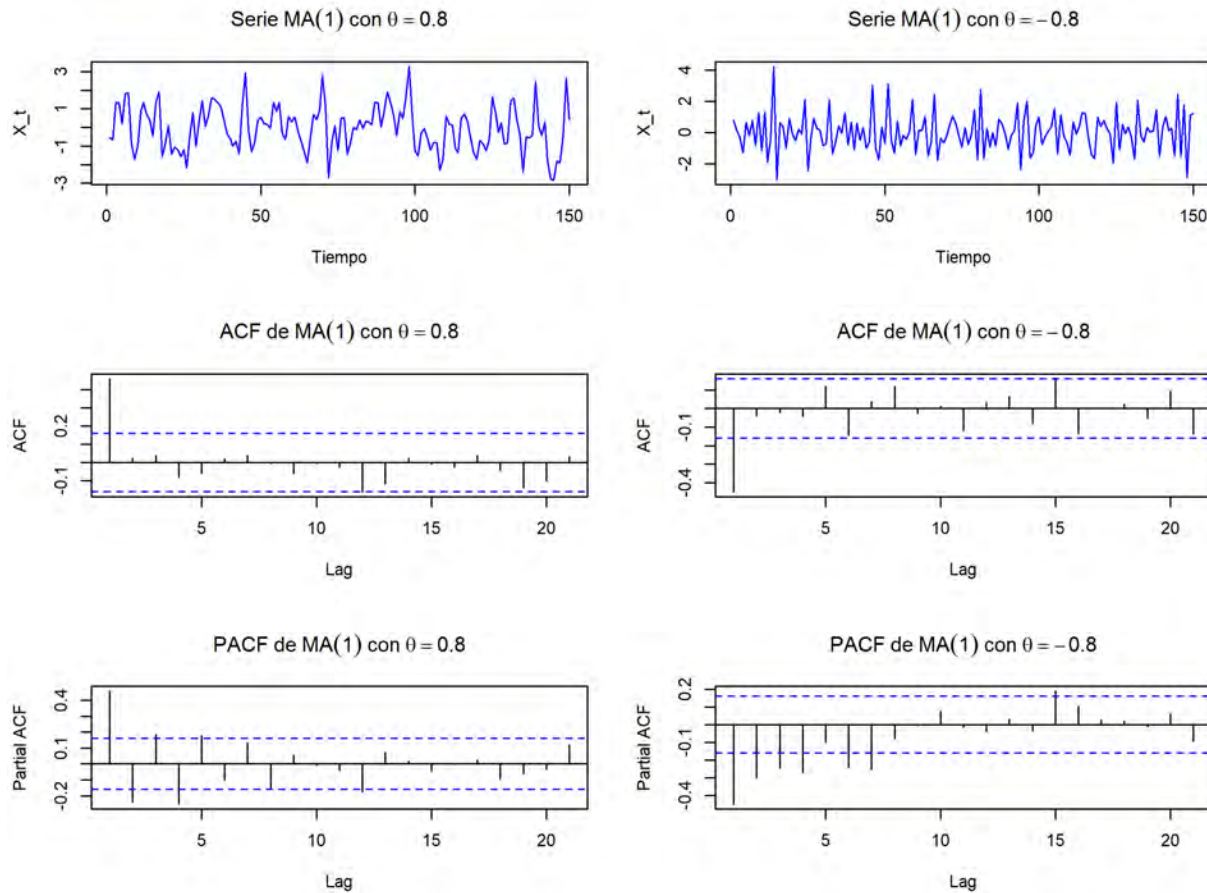
$$\rho_2 = \frac{\gamma_2}{\gamma_0} = 0,$$

\vdots

$$\rho_k = \frac{\gamma_k}{\gamma_0} = 0, \text{ para todo } k \geq 2.$$

Figura 12

Simulación de series MA(1) y sus funciones ACF y PACF con $\theta = \pm 0,8$



Nota: Elaboración propia con datos simulados en R-Studio.

2) Proceso de media móvil de segundo orden MA(2)

El proceso de segundo orden esta definido como:

$$X_t = \mu + \theta_1 a_{t-1} + \theta_2 a_{t-2} + a_t$$

■ Media

$$E[X_t] = E[\mu + \theta_1 a_{t-1} + \theta_2 a_{t-2} + a_t] = \mu$$

■ **Varianza**

$$Var[X_t] = E[(X_t - \mu)^2] = E[(\theta_1 a_{t-1} + \theta_2 a_{t-2} + a_t)^2] = \theta_1^2 \sigma_a^2 + \theta_2^2 \sigma_a^2 + \sigma_a^2$$

$$Var[X_t] = \gamma_0 = \sigma_a^2(\theta_1^2 + \theta_2^2 + 1)$$

■ **Autocovarianzas**

$$\gamma_1 = E[(X_t - \mu)(X_{t-1} - \mu)] = E[(\theta_1 a_{t-1} + \theta_2 a_{t-2} + a_t)(\theta_1 a_{t-2} + \theta_2 a_{t-3} + a_{t-1})],$$

$$\gamma_1 = (\theta_1 + \theta_1 \theta_2) \sigma_a^2$$

$$\gamma_2 = E[(X_t - \mu)(X_{t-2} - \mu)] = E[(\theta_1 a_{t-1} + \theta_2 a_{t-2} + a_t)(\theta_1 a_{t-3} + \theta_2 a_{t-4} + a_{t-2})],$$

$$\gamma_2 = \theta_2 \sigma_a^2$$

$$\gamma_3 = E[(X_t - \mu)(X_{t-3} - \mu)] = E[(\theta_1 a_{t-1} + \theta_2 a_{t-2} + a_t)(\theta_1 a_{t-4} + \theta_2 a_{t-5} + a_{t-3})],$$

$$\gamma_3 = 0,$$

$$\vdots$$

$$\gamma_k = 0, \text{ para todo } k \geq 3.$$

■ **Autocorrelaciones**

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\theta_1 + \theta_1 \theta_2}{\theta_1^2 + \theta_2^2 + 1},$$

$$\rho_2 = \frac{\gamma_2}{\gamma_0} = \frac{\theta_2}{\theta_1^2 + \theta_2^2 + 1},$$

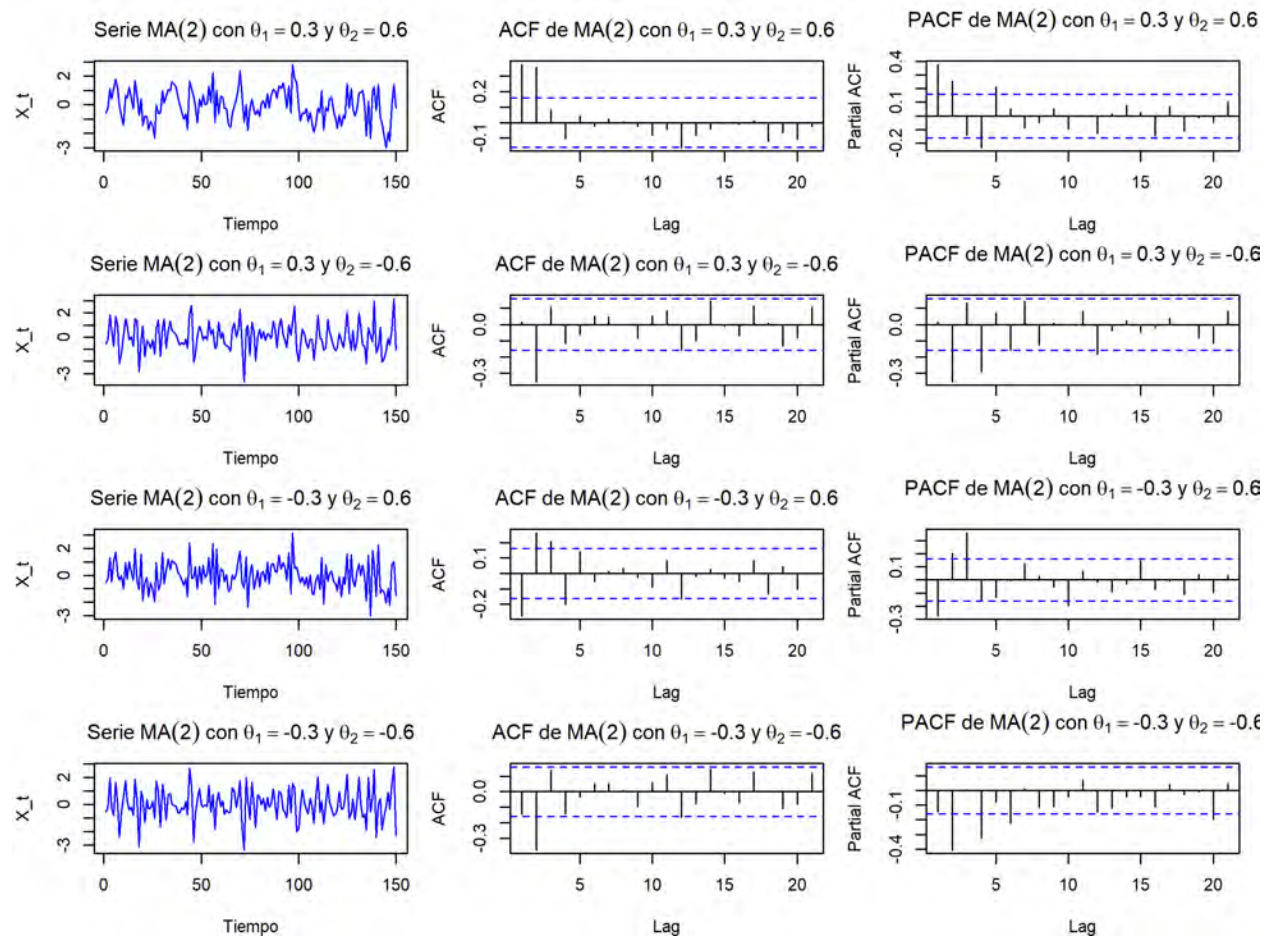
$$\rho_3 = \frac{\gamma_3}{\gamma_0} = 0$$

$$\vdots$$

$$\rho_k = \frac{\gamma_k}{\gamma_0} = 0, \text{ para todo } k \geq 3.$$

Figura 13

Simulación de series MA(2) y sus funciones ACF y PACF con $\theta_1 = \pm 0,3$ y $\theta_2 = \pm 0,6$



Nota: Elaboración propia con datos simulados en R-Studio.

c) Proceso mixto autoregresivo de media móvil ARMA(p,q)

Si asumimos que la serie es en parte autoregresiva y en parte de media móvil, obtenemos un modelo de series temporales bastante general.

$$X_t = \mu + \phi_1(X_{t-1} - \mu) + \phi_2(X_{t-2} - \mu) + \dots + \phi_p(X_{t-p} - \mu) + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q} \quad (4)$$

decimos que X_t , es un proceso mixto autoregresivo de media móvil de ordenes p y q , respectivamente, con $\mu \neq 0$.

$$X_t = \delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q} \quad (5)$$

donde: $\delta = \mu(1 - \phi_1 - \phi_2 - \dots - \phi_q)$

La ecuación (4), puede definirse en términos de sus desviaciones como:

$$\bar{X}_t = \phi_1 \bar{X}_{t-1} + \phi_2 \bar{X}_{t-2} + \dots + \phi_p \bar{X}_{t-p} + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q}$$

La ecuación (5) también se puede escribir como:

$$\phi(B)X_t = \theta(B)a_t + \delta$$

1) Proceso mixto autoregresivo de media móvil ARMA(1,1)

El proceso esta definido como:

$$X_t = \delta + \phi_1 X_{t-1} + a_t + \theta_1 a_{t-1}$$

■ Media

$$E[X_t] = E[\delta + \phi_1 X_{t-1} + \theta_1 a_{t-1} + a_t] = \delta + \phi_1 E[X_{t-1}] + \theta_1 E[a_{t-1}] + E[a_t]$$

$$E[X_t] = \delta + \phi_1 \mu$$

■ Varianza

$$Var[X_t] = E[(X_t - \mu)^2] = E[(\phi_1 X_{t-1} + \theta_1 a_{t-1} + a_t)^2],$$

$$Var[X_t] = \phi_1^2 \gamma_0 + \theta_1^2 \sigma_a^2 + \sigma_a^2 + 2\phi_1 \theta_1 \sigma_a^2$$

$$Var[X_t] = \gamma_0 = \frac{\sigma_a^2(\theta_1^2 + 1 + 2\phi_1 \theta_1)}{1 - \phi_1^2}$$

■ Autocovarianzas

$$\begin{aligned}
 \gamma_1 &= E[\bar{X}_t \bar{X}_{t-1}] = E[(\phi_1 \bar{X}_{t-1} + \theta_1 a_{t-1} + a_t) \bar{X}_{t-1}] = \phi_1 \gamma_0 + \theta_1 \sigma_a^2, \\
 \gamma_2 &= E[\bar{X}_t \bar{X}_{t-2}] = E[(\phi_1 \bar{X}_{t-1} + \theta_1 a_{t-1} + a_t) \bar{X}_{t-2}] = \phi_1^2 \gamma_0 + \phi_1 \theta_1 \sigma_a^2, \\
 &\vdots \\
 \gamma_k &= \phi_1 \gamma_{k-1} = \phi_1^k \gamma_0 + \phi_1^{k-1} \theta_1 \sigma_a^2, \text{ para todo } k \geq 2.
 \end{aligned}$$

■ Autocorrelaciones

$$\begin{aligned}
 \rho_1 &= \frac{\gamma_1}{\gamma_0} = \phi_1 + \frac{\theta_1(1 - \phi_1^2)}{1 + \theta_1^2 + 2\phi_1\theta_1}, \\
 \rho_2 &= \frac{\gamma_2}{\gamma_0} = \phi_1^2 + \phi_1 \frac{\theta_1(1 - \phi_1^2)}{1 + \theta_1^2 + 2\phi_1\theta_1}, \\
 &\vdots \\
 \rho_k &= \frac{\gamma_k}{\gamma_0} = \phi_1 \rho_{k-1}, \text{ para todo } k \geq 2.
 \end{aligned}$$

2.2.8.10. Proceso lineal no estacionario.

a) Proceso autoregresivo integrado de media móvil ARIMA(p,d,q)

Se dice que una serie temporal $\{X_t\}$ sigue un proceso de media móvil autoregresivo integrado si la d-ésima diferencia $W_t = \nabla^d X_t = X_t - X_{t-d}$ es un proceso ARMA estacionario.

Si $\{W_t\}$ sigue un proceso ARMA(p, q), decimos que $\{X_t\}$ es un proceso ARIMA(p, d, q).

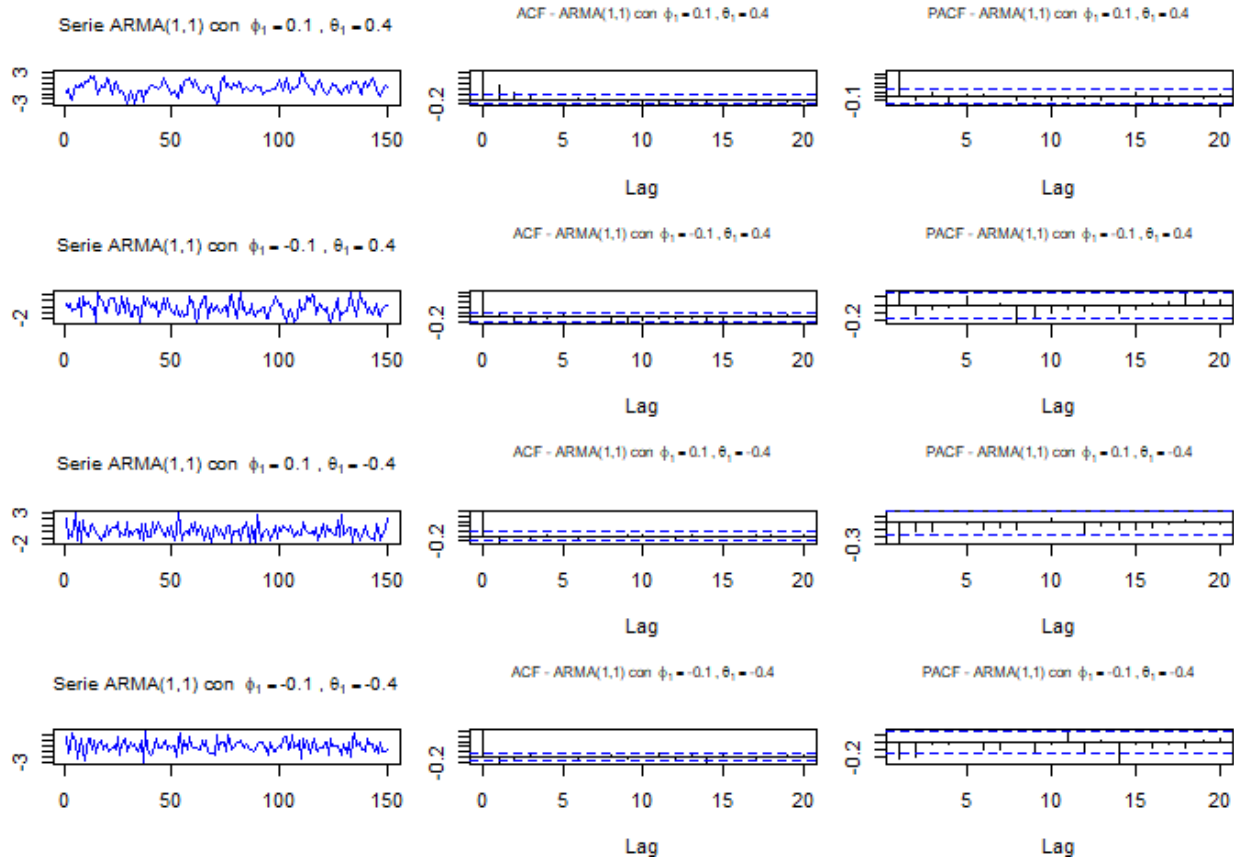
$$\nabla^d X_t = \phi_1 \nabla^d X_{t-1} + \phi_2 \nabla^d X_{t-2} + \dots + \phi_p \nabla^d X_{t-p} + a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q} \quad (6)$$

La ecuación (6), puede expresarse de la siguiente manera en términos de su rezago.

$$\phi(B)(1 - B)^d X_t = \theta(B)a_t$$

Figura 14

Simulación de series ARMA(1,1) y sus funciones ACF y PACF con $\phi_1 = \pm 0,1$ y $\theta_1 = \pm 0,4$



Nota: Elaboración propia con datos simulados en R-Studio.

1) Proceso autoregresivo integrado de media móvil ARIMA(1,1,1)

Se define de la siguiente manera:

$$\theta(B)a_t = \phi(B)(1 - B)X_t$$

$$a_t + \theta_1 a_{t-1} = (1 - \phi_1 B)(X_t - X_{t-1})$$

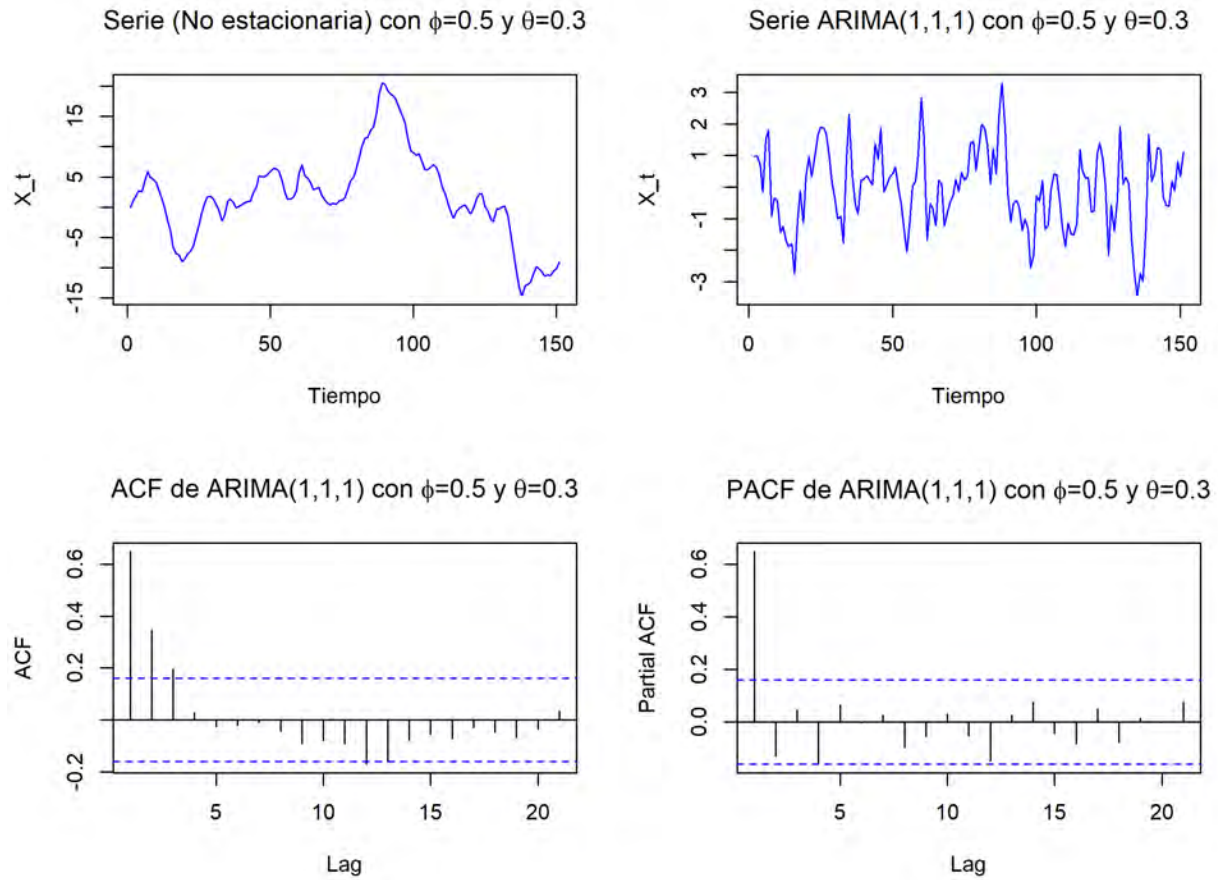
$$a_t + \theta_1 a_{t-1} = X_t - X_{t-1} - \phi_1 (X_{t-1} - X_{t-2})$$

$$X_t = X_{t-1} + \phi_1 (X_{t-1} - X_{t-2}) + a_t + \theta_1 a_{t-1}$$

$$\nabla X_t = \phi_1 \nabla X_{t-1} + a_t + \theta_1 a_{t-1}$$

Figura 15

Simulación de series ARIMA(1,1,1) y sus funciones ACF y PACF con $\phi_1 = 0,5$ y $\theta_1 = 0,3$



Nota: Elaboración propia con datos simulados en R-Studio.

2.2.8.11. Procesos estacionales.

a) Proceso ARMA estacionales multiplicativos SARMA(p, q) \times (P, Q)_s

Shumway y Stoffer (2016), indica al modelo SARMA el cual incluye componentes estacionales, denotado como:

$$X_t = \mu + \sum_{i=1}^P \Phi_i X_{t-i.s} + \sum_{j=1}^Q \Theta_j a_{t-j.s} + \sum_{m=1}^p \phi_m X_{t-m} + \sum_{n=1}^q \theta_n a_{t-n} + a_t$$

y en términos de su rezago definido como:

$$\Phi(B^s) \phi(B) X_t = \Theta(B^s) \theta(B) a_t \quad (7)$$

Donde:

$\Phi(B^s)$: Polinomio autoregresivo estacional de orden P.

$\phi(B)$: Polinomio autoregresivo regular de orden p.

$\Theta(B^s)$: Polinomio de media móvil estacional de orden Q.

$\theta(B)$: Polinomio de media móvil regular de orden q.

a_t : Ruido blanco.

s : Periodo estacional.

Un modelo $SARMA(0, 1) \times (1, 0)_{12}$, según la ecuación (7), está definido como:

$$\Phi(B^{12})X_t = \theta(B)a_t$$

$$(1 - \Phi B^{12})X_t = (1 + \theta B)a_t$$

$$X_t - \Phi B^{12}X_t = a_t + \theta B a_t$$

$$X_t - \Phi X_{t-12} = a_t + \theta a_{t-1}$$

$$X_t = \Phi X_{t-12} + a_t + \theta a_{t-1}$$

b) Proceso ARIMA estacionales no estacionarios $SARIMA(p,d,q) \times (P,D,Q)_s$

Estos modelos son flexibles en el sentido de que especifican estacionalidades estocásticas, tendencias estocásticas y además recogen la posible interacción entre ambos componentes (González Casimiro, 2009).

Brockwell y Davis (2016) este modelo es una extensión del modelo ARIMA que incorpora componentes estacionales, esta definido como:

$$\Phi(B^s) \phi(B)(1 - B)^d(1 - B^s)^D X_t = \Theta(B^s) \theta(B) a_t \quad (8)$$

$\Phi(B^s)$: Polinomio autoregresivo estacional de orden P.

$\phi(B)$: Polinomio autoregresivo regular de orden p.

$\Theta(B^s)$: Polinomio de media móvil estacional de orden Q.

$\theta(B)$: Polinomio de media móvil regular de orden q.

$(1 - B)^d$: Operador de diferencias regular.

$(1 - B^s)^D$: Operador de diferencias estacional.

a_t : Ruido blanco.

s : Es el periodo estacional.

Un modelo SARIMA(1,1,0) \times (0, 1, 1)₁₂, según la ecuación (8) estaría definido como:

$$\begin{aligned}\Theta(B^{12})a_t &= \phi(B)(1 - B^{12})(1 - B)X_t \\ (1 + \Theta_1 B^{12})a_t &= (1 - \phi_1 B)(1 - B^{12})(X_t - X_{t-1}) \\ a_t + \Theta_1 a_{t-12} &= (1 - \phi_1 B)(X_t - X_{t-1} - X_{t-12} + X_{t-13}) \\ a_t + \Theta_1 a_{t-12} &= X_t - X_{t-1} - X_{t-12} + X_{t-13} + \phi_1(X_{t-1} - X_{t-2}) + \phi_1(X_{t-13} - X_{t-14}) \\ X_t &= X_{t-1} + X_{t-12} - X_{t-13} - \phi_1(X_{t-1} - X_{t-2}) - \phi_1(X_{t-13} - X_{t-14}) + a_t + \Theta_1 a_{t-12}\end{aligned}$$

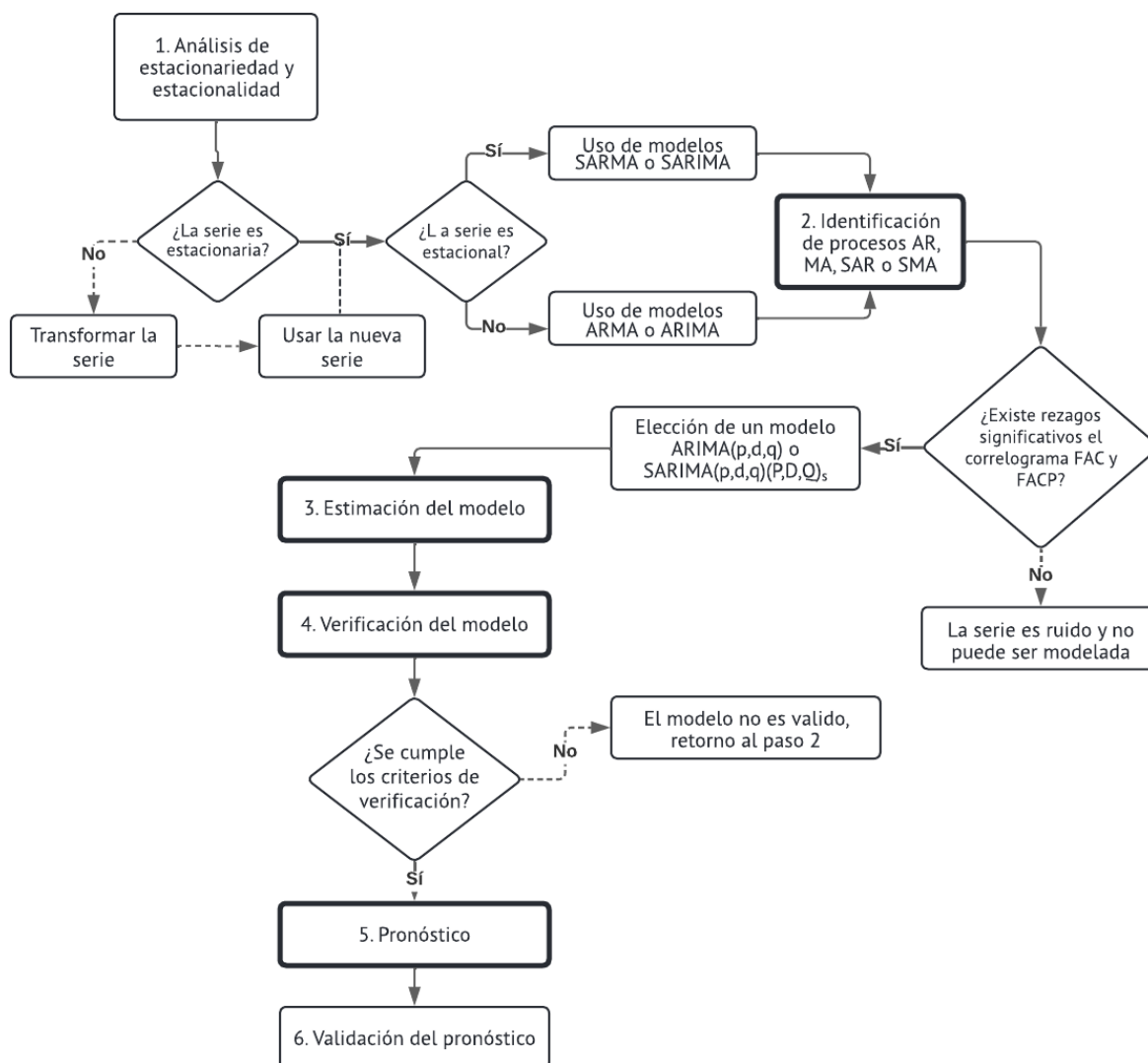
o de la forma

$$\nabla X_t = \nabla X_{t-12} - \phi_1 \nabla X_{t-1} - \phi_1 \nabla X_{t-13} + a_t + \Theta_1 a_{t-1}$$

2.2.9. Metodología de Box Jenkins

La metodología de Box Jenkins sigue las siguientes fases (Box et al., 2016):

- a) **Identificación del modelo:** Este es el primer paso de la metodología de Box Jenkins, donde se observa el comportamiento de la serie, este paso implica el uso de gráficas como los correlogramas, la función de autocorrelacion (ACF), la función de autocorrelacion parcial (PACF) y poder identificar los procesos autoregresivo, media móvil y la integración dentro de la serie.

Figura 16*Etapas de la metodología de Box Jenkins*

Nota: Adaptado de “Intermediate economics: Theory and applications”(pg.223) ¹

- b) Estimación de los parámetros del modelo:** Una vez identificado el modelo y los parámetros de orden p y q , se procede a estimar los parámetros desconocidos:

$$\theta = (\delta, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)'$$

según Hamilton (1994) para poder estimar los parámetros de la serie, los residuos a_t

deben ser independientes e idénticamente distribuidos (iid) y $a_t \sim \mathcal{N}(0, \sigma^2)$. Su probabilidad de densidad esta dada por (Box & Jenkins, 1976):

$$f(a_1, a_2, a_3, \dots, a_t) \propto (\sigma^2)^{-\frac{n}{2}} \exp \left\{ - \sum_{t=1}^n \frac{a_t^2}{2\sigma^2} \right\}$$

La función de verosimilitud se define como:

$$L(\theta) = \prod_{t=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{a_t^2}{2\sigma^2} \right),$$

Tomando logaritmos, se obtiene la log-verosimilitud:

$$\log L(\theta) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^n a_t^2,$$

$$L(\theta) = -\frac{n}{2} (\log(2\pi\sigma^2)) - \frac{1}{2\sigma^2} \sum_{t=1}^n a_t^2$$

la cual se maximiza con respecto a θ .

A partir del cual se define la función condicional de suma de cuadrados $S^* = \sum_{t=1}^n a_t^2$, con lo cual se estudia el comportamiento de la verosimilitud condicional de suma de cuadrados bajo el cual esta basado la librería del software R para estimar los parámetros del modelo.

Una vez estimado los parámetros, usamos los criterios de información para la elección del mejor estimador esperándose sea el mínimo posible.

- 1) **Criterio de información de AKAIKE (AIC):** Este criterio permite seleccionar modelos estimando la medida de la calidad relativa de un modelo estadístico para un conjunto de datos, utiliza como bondad de ajuste al criterio de máxima verosimilitud y viene dado por:

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

donde L representa la función de verosimilitud del modelo y k es el número de

parámetros estimados. En este sentido, el AIC permite comparar varios modelos, seleccionando el que tenga el valor más bajo como el mejor, ya que optimiza el equilibrio entre ajuste y simplicidad (Cotrina Araujo, 2020).

- 2) **Criterio de información de Bayes (BIC):** Este criterio fue desarrollado por Schwarz (1978) es una medida que evalúa el ajuste de un modelo estadístico y, por lo tanto, se utiliza como método de selección de modelos. Su fórmula es:

$$\text{BIC} = k \ln(n) - 2 \ln(L)$$

donde: k representa el número de parámetros del modelo, L es el valor de máxima verosimilitud, y n es el número de observaciones en el conjunto de datos (Cotrina Araujo, 2020).

- c) **Verificación del modelo:** En este paso hacemos uso de los residuos de la serie para comprobar que sean un proceso de ruido blanco y se distribuyan normalmente. Para ello hacemos uso de las siguientes pruebas:

1) **Prueba de Ljung-Box**

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_h = 0 \text{ (es ruido blanco)}$$

$$H_1 : \text{no es ruido blanco}$$

$$Q = n(n+2) \sum_{j=1}^h \frac{\hat{\rho}_j^2}{n-j} \sim \chi^2_{1-\alpha, h}$$

donde: n representa el tamaño de la muestra, h el numero de retardos y $\hat{\rho}_j$ es la autocorrelacion en el retardo h .

2) Jarque Bera:

H_0 : Los datos de la serie de tiempo tienen una distribución normal

H_1 : Los datos no siguen una distribución normal

$$JB = \left(\frac{n}{6}\right)A^2 + \left(\frac{n}{24}\right)(K - 3)^2 \sim \chi^2_{(2gl)}$$

donde: **n** representa el tamaño de la muestra, **A** representa la asimetría y **K** la curtosis.

Nota: La normalidad de los residuos no es un requisito estricto para el ajuste del modelo ARIMA, pero sí es fundamental para la validez de las inferencias estadísticas basadas en el modelo. Según Hyndman y Athanasopoulos (2018), si los residuos no presentan distribución normal, el modelo puede mejorarse y los intervalos de predicción podrían no ser confiables, ya que no se cumpliría con los supuestos necesarios para su correcta construcción.

d) **Pronóstico:** Una vez validado el modelo puede entonces utilizarse para predecir o pronosticar valores de la serie en períodos futuros. Esto se realiza aplicando el modelo en los valores de tiempo $t + 1, t + 2, \dots$

$$\hat{X}_{t+h} = c + \phi_1 \hat{X}_{t+h-1} + \phi_2 \hat{X}_{t+h-2} + \dots + \phi_p \hat{X}_{t+h-p} - \theta_1 a_{t+h-1} - \theta_2 a_{t+h-2} - \dots - \theta_q a_{t+h-q}$$

Se hace uso de las métricas de validación para elegir el mejor valor de pronóstico.

1) **Raíz del error cuadrático medio (RMSE):** Mide el promedio de las diferencias al cuadrado entre los valores predichos y los valores observados, proporcionando una idea de cuánto, en promedio, se desvía el modelo de los valores reales.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (X_t - \hat{X}_{t+h})^2}$$

- 2) **Error absoluto medio (MAE):** Mide el promedio de los errores absolutos entre los valores observados y los valores predichos por el modelo.

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |X_t - \hat{X}_{t+h}|$$

- 3) **Error porcentual absoluto medio (MAPE):** Es una medida útil para interpretar el error en relación con el tamaño de los valores observados, permitiendo comparar el rendimiento de un modelo en diferentes escalas.

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{X_t - \hat{X}_{t+h}}{X_t} \right| \times 100$$

- 4) **Suma de errores cuadrados (SSE):** Suma los errores al cuadrado entre las predicciones y los valores observados. Es útil para evaluar la calidad del modelo y es particularmente sensible a los errores grandes debido al cuadrado de las diferencias.

$$\text{SSE} = \sum_{t=1}^n (X_t - \hat{X}_{t+h})^2$$

2.3. Marco conceptual

1. Cobre

El cobre es un metal noble conocido por su alta conductividad eléctrica y térmica, resistencia a la corrosión, y capacidad de ser reciclado indefinidamente sin perder sus propiedades. Además, es fácil de moldear y cortar, y posee propiedades bactericidas. Su característica color rojo le ha valido el apodo de "metal rojo". Se usa tanto en su forma pura, como en cables y tuberías, como en aleaciones con otros metales, transmitiendo sus cualidades a estas combinaciones (Donoso, 2013).

2. Producción

En el aspecto de minería, producción se refiere a minerales metálicos que son comercializados principalmente como materia prima, en forma de concentrados ó productos refinados, y en contados casos con cierto valor agregado (Ministerio de Desarrollo Agrario y Riego, 2024).

3. Producción de cobre

Según International Copper Association (2024), la producción de cobre primario comienza con la extracción de minerales portadores de cobre. Hay dos formas básicas de extracción de cobre: la minería de superficie y la subterránea. Debido a que el cobre disminuye en concentraciones relativamente bajas en grandes áreas, la minería de superficie, o a cielo abierto, es el método de extracción predominante para el cobre en el mundo.

El proceso productivo del cobre abarca una serie de procedimientos destinados a obtener cobre de alta pureza, libre de impurezas. Comienza con la exploración de terrenos en busca de yacimientos minerales y culmina con la producción de cátodos o planchas de cobre con una pureza del 99,99 %, que se comercializan globalmente (CodelcoEduca, 2024).

4. Series estacionaria

Hanke y Reitsch (1996) define como aquella cuyas propiedades estadísticas básicas, como la media y la varianza, permanecen constantes en el tiempo. Se dice que una serie que no

presenta crecimiento o declinación es estacionaria.

5. **Serie estacional**

Son series que incorporan componentes que capturan y representan patrones recurrentes a lo largo del tiempo, típicamente en intervalos regulares, como días, meses, o trimestres. Estos patrones, conocidos como estacionalidad, reflejan variaciones que se repiten con una periodicidad fija y predecible dentro de la serie temporal (Box et al., 2016).

6. **Serie heterocedastica**

Hamilton (1994), define a una serie heterocedastica cuando las varianzas a lo largo del tiempo no son constantes.

7. **Pronóstico**

En estadística inferencial, el pronóstico es un proceso matemático mediante el cual se hace una estimación del valor futuro de una o más variables (Fierro Torres et al., 2022).

8. **Tiempo**

Magnitud física que permite ordenar la secuencia de los sucesos, estableciendo un pasado, un presente y un futuro, y cuya unidad en el sistema internacional es el segundo (Real Academia Española, 2024).

9. **Correlograma**

Es la representación gráfica de la **ACF** de una serie de tiempo, es un instrumento visual muy útil para decidir sobre la estacionariedad de dicha serie (Mauricio, 2007).

CAPÍTULO III

METODOLOGÍA

3.1. Tipo de investigación

El tipo de la investigación es aplicada ya que aprovecha los conocimientos logrados por la investigación básica o teórica para el conocimiento y solución de problemas inmediatos y cuantitativa porque evalúa los datos de manera científica o numérica con ayuda de la estadística (Sánchez et al., 2018).

3.2. Alcance de la investigación

Este estudio es descriptivo debido a que especifica propiedades y características de conceptos, correlacional porque determina en qué medida dos o más variables están relacionadas entre sí y predictivo porque permite cierto grado de predicción a pasado o futuro (Sánchez et al., 2018).

3.3. Diseño de investigación

El presente estudio es no experimental debido a que se observan fenómenos tal como se dan en su contexto natural, sin manipular ni intervenir en la variable, longitudinal y retrospectivo porque se recolectan datos en diferentes puntos del tiempo (pasado), para realizar el análisis respectivo (Hernández et al., 2018). .

3.4. Población de estudio

La población está conformada por los datos mensuales de producción total de cobre en el Perú, reportados por el MINEM, desde Enero de 1999 hasta Agosto de 2024, sumando 308 observaciones. Este conjunto se utiliza en su totalidad en la investigación, dividiéndose metodológicamente en dos partes: de 1999 a 2023 para el modelado y entrenamiento, y de Enero a Agosto de 2024 para la validación del pronóstico.

3.5. Técnicas e instrumentos de recolección de datos

La técnica de recolección de datos para este estudio fue secundaria, debido a que se aprovechó la base de datos mensual de producción de cobre proporcionada por el Ministerio de Energía y Minas.

3.6. Procesamiento de datos

Con el propósito de cumplir los objetivos del estudio, se utilizó la base de datos del Ministerio de Energía y Minas (MINEM), correspondiente al periodo de Enero de 1999 a Agosto de 2024. La variable analizada fue la producción total mensual de cobre (TMF). Se aplicó la metodología Box-Jenkins para el análisis y pronóstico, utilizando el software RStudio. A continuación, se describen los pasos seguidos:

1. **Limpieza y preparación de datos:** Se verificó que los datos mensuales estuvieran completos y sin valores faltantes. Luego, se transformaron en una serie temporal con frecuencia mensual. La base fue dividida en dos subconjuntos: datos de entrenamiento (enero 1999 - diciembre 2023) y datos de prueba (enero - agosto 2024).
2. **Análisis exploratorio:** Se identificaron características como tendencia, estacionalidad, y posibles cambios estructurales mediante gráficos de línea y diagramas de caja. También se aplicaron pruebas estadísticas como Mann-Kendall, Dickey-Fuller, Levene y Kruskal-Wallis.
3. **Identificación del modelo:** Se evaluó la necesidad de estabilizar la varianza mediante la transformación de Box-Cox. Luego, se aplicó diferenciaciones regular y estacional (ambas de orden 1). Con los gráficos ACF y PACF, se propusieron 12 modelos SARIMA candidatos.
4. **Estimación y validación:** Los modelos fueron estimados mediante máxima verosimilitud. Se evaluó la significancia de los coeficientes y se validaron los residuos utilizando las pruebas de Ljung-Box (independencia) y Jarque-Bera (normalidad), además de gráficos Q-Q.

5. **Pronóstico y evaluación:** Se generaron pronósticos para el periodo Enero–Agosto 2024. Los valores estimados se compararon con los datos reales utilizando métricas como el RMSE y el MAPE para evaluar el desempeño de los modelos.

Y por ultimo se procedió a interpretar los resultados.

3.7. Operacionalización de variable

Tabla 1

Operacionalización de la variable

Variables	Definición Conceptual	Definición operacional	Indicadores	Escala
Producción mensual de cobre.	El proceso productivo del cobre es la serie de procedimientos que se realizan para alcanzar un cobre de alta concentración, es decir, libre de la mayor cantidad de impurezas. CodelcoEduca, 2024	Cantidad total de cobre extraído en Perú durante un período específico (mensual), medidos en toneladas métricas.	Producción mensual de cobre.	Razón

CAPÍTULO IV

RESULTADOS

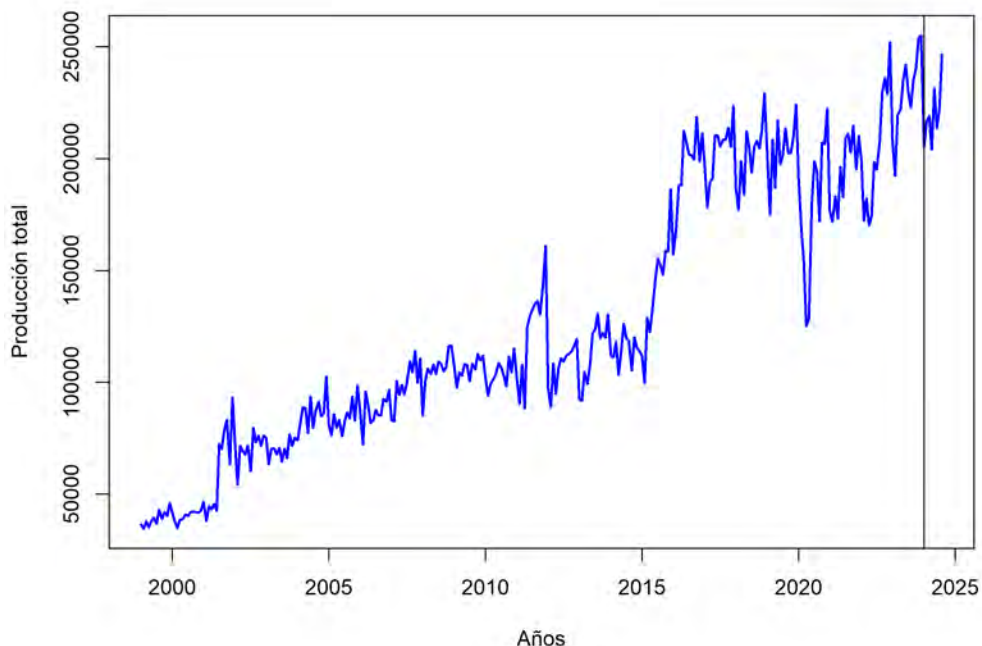
4.1. Exploración de la Producción mensual del cobre en el Perú.

En concordancia con la metodología se divide la serie en dos partes: data de entrenamiento (1999-2023) y la data de prueba (Enero-Agosto 2024), para evaluar la capacidad de pronóstico del mejor modelo.

Podemos apreciar en la Figura 17 la serie original de la producción total mensual de cobre en Perú (1999-2024), donde la línea vertical separa la serie en dos grupos: la data de entrenamiento y la data de prueba. Podemos apreciar que la serie tiene tendencia creciente; se ve un incremento alto de producción entre los meses del año 2015 debido a que este año iniciaron sus operaciones la minera Constancia y el mega proyecto de Toromocho. Para inicios del año 2020, se observa una disminución drástica de la producción, posiblemente por las medidas propuestas durante la pandemia del COVID-19. En este mismo año se observa que en los próximos meses va en aumento, indicando una recuperación en la producción de cobre en Perú.

Figura 17

Serie Original de la producción total mensual del cobre en Perú 1999-2024



A continuación nos centraremos a analizar el análisis explícito con la data de entrenamiento. En la Tabla 2 se muestran los estadísticos descriptivos de la producción total mensual del cobre en Perú durante el periodo de 1999 – 2023, se obtiene una tasa de crecimiento mensual promedio de 1,3 % TMF de producción total mensual de cobre con un desvío típico de 58 309.85 TMF. La producción total mensual mínima de cobre es de 34 553 TMF y la máxima alcanza 255 143 TMF.

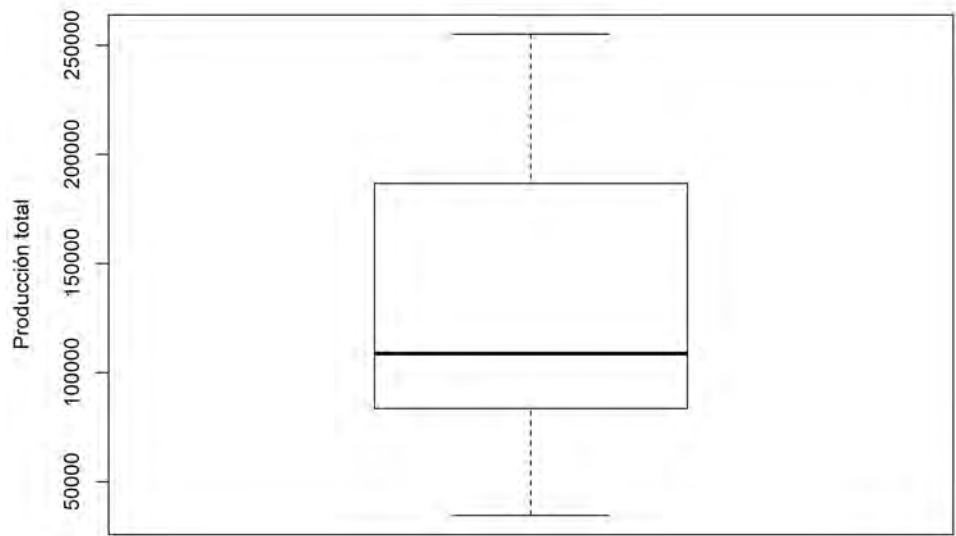
Tabla 2

Estadísticos descriptivos de la producción total de cobre en Perú (1999-2023)

Estadístico	Valor
Valor mínimo	34553 TMF
Primer cuartil (Q1)	83751 TMF
Mediana	108898 TMF
Tercer cuartil (Q3)	186436 TMF
Valor máximo	255143 TMF
Tasa mensual promedio	1,3 % TMF
Desviación estándar	58309,85 TMF

Figura 18

Diagrama de cajas de la Producción total mensual de cobre en Perú (1999-2023)

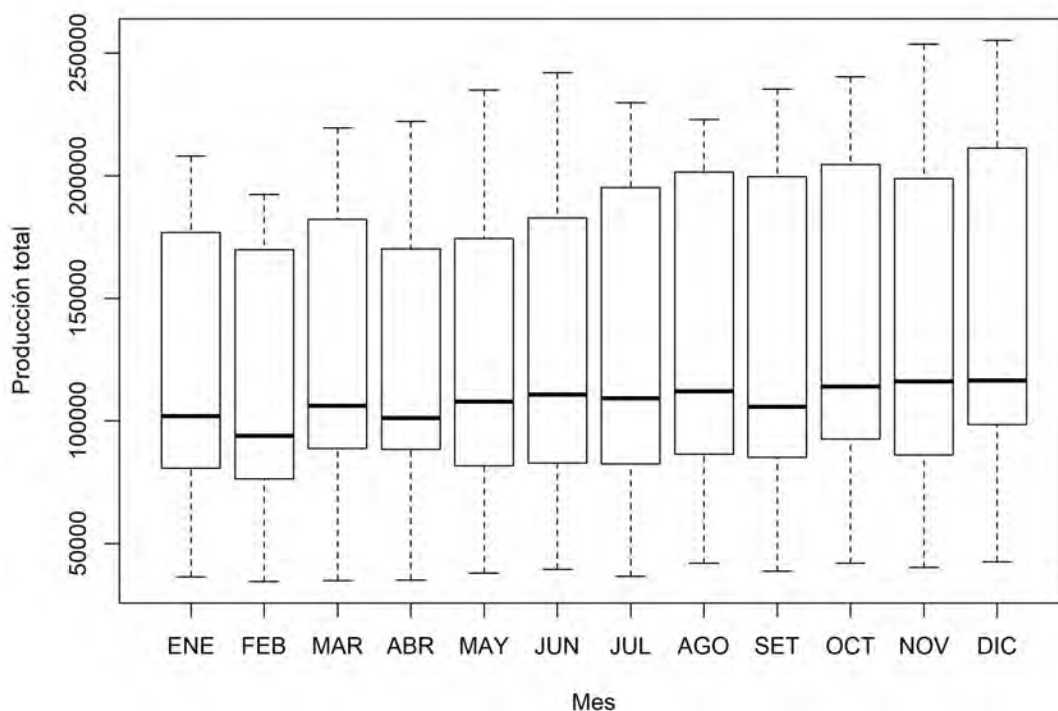


En la Figura 18, se visualiza el diagrama de cajas de la producción total mensual de cobre en Perú entre 1999 y 2023 presenta una asimetría positiva (cola más larga hacia valores altos). La

mediana se ubica cerca de los 110 000 TMF, mientras que el rango intercuartílico va de 83 751 a 186 436 TMF. No se observan valores atípicos extremos dentro del período analizado.

Figura 19

Diagrama de cajas de la producción total de cobre en Perú (1999-2023) mensual.



Por otro lado en la Figura 19 se visualiza las cajas con respecto a cada mes y se puede ver que no existen datos atípicos, el mes de Febrero tiene una mediana de producción de cobre más baja con respecto a los demás meses, las cajas tiene asimetría positiva, lo que significa que existe una cola más larga en la parte superior de la caja (lado derecho), la mediana de los 12 meses oscilan alrededor de 100 000 TMF. de producción de cobre. La producción de cobre en cada mes puede variar debido a las condiciones climáticas, a mantenimientos programados, etc.

4.2. Identificación de las características de la producción total mensual de cobre Perú (1999-2023).

Como se observa en la Tabla 3, los resultados indican que la serie presenta una tendencia significativa, confirmada por la prueba de Mann-Kendall ($p < 0.05$), lo que sugiere un cambio sistemático

en la media a lo largo del tiempo. La prueba de Kruskal-Wallis muestra diferencias significativas entre años ($p < 0.05$), pero no entre meses ($p = 0.3872$), evidenciando que las variaciones relevantes ocurren a nivel interanual y no responden a un patrón estacional dentro del año. Además, la prueba Dickey-Fuller no rechaza la hipótesis de no estacionariedad ($p = 0.1694$), indicando la necesidad de aplicar diferencias para estabilizar la media. En cuanto a la variabilidad, la prueba de Levene detecta heterocedasticidad significativa entre años ($p < 0.05$), mientras que a nivel mensual la varianza se mantiene homogénea ($p = 0.9796$).

Tabla 3

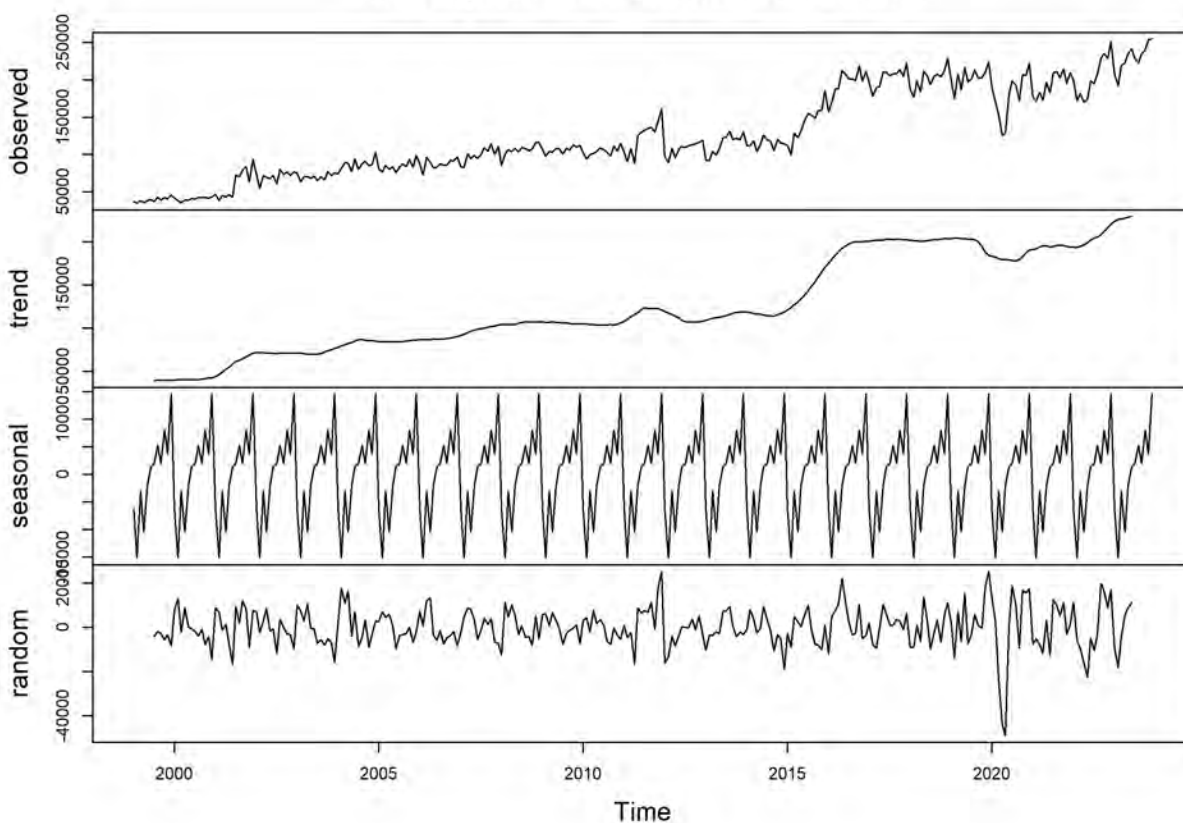
Pruebas de características de la serie de producción de cobre en el Perú (1999-2023)

Característica	Prueba	Hipótesis Nula	Estadístico	P-valor	Decisión
Tendencia	Mann-Kendall	No tiene tendencia	20,72	$2,2 \times 10^{-16}$	Rechazar H_0
Estacionalidad	Kruskal-Wallis	No estacional			
		- Anual	277,65	$2,2 \times 10^{-16}$	Rechazar H_0
		- Mensual	11,693	0,3872	No Rechazar H_0
Estacionariedad	Dickey-Fuller	No estacionario	-2,9653	0,1694	No Rechazar H_0
Heterocedastica	Levene	Homocedastica			
		- Anual	5,4682	$3,2 \times 10^{-13}$	Rechazar H_0
		- Mensual	0,3268	0.9796	No Rechazar H_0

La Figura 20, nos muestra la descomposición de la serie de producción de cobre (1999-2023) y sus componentes de tendencia, estacionalidad y aleatoriedad. Además, se puede observar que presenta tendencia creciente, tiene comportamiento estacional cada 12 meses durante el periodo de estudio.

Figura 20

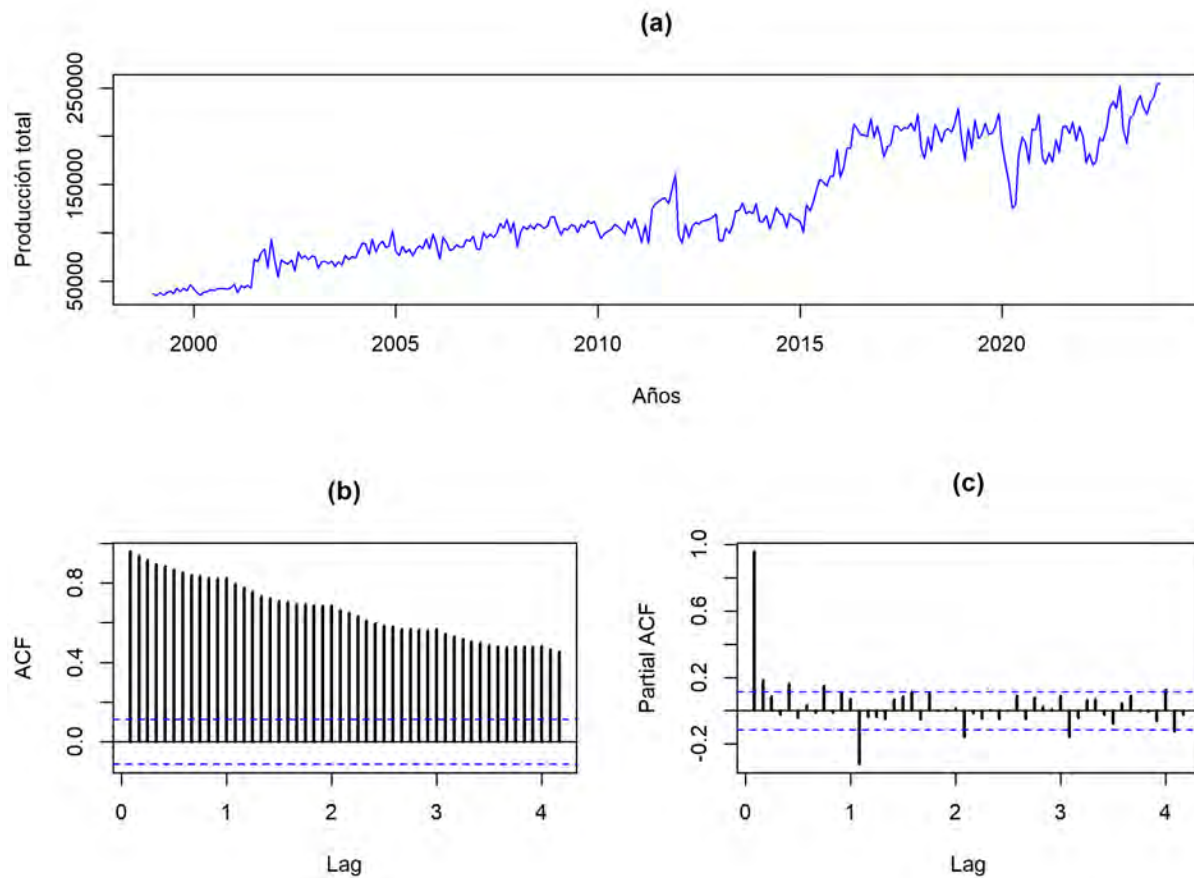
Descomposición de la serie de producción de cobre Perú (1999-2023)



Los correlogramas de la Figura 21, muestran que (b) tiene un decaimiento exponencial lento lo que asume que existe componentes de media móvil en la serie, de igual forma (c) muestra lags significativos lo que asume existencia de componentes autoregresivo en la serie, asimismo por las gráficas de los correlogramas y la prueba de Dickey Fuller mostrada en la Tabla 3 concluimos que la serie no es estacionario por lo cual debemos generar una transformación o una diferenciación regular.

Figura 21

Correlogramas de la producción total de cobre Perú (1999-2023)

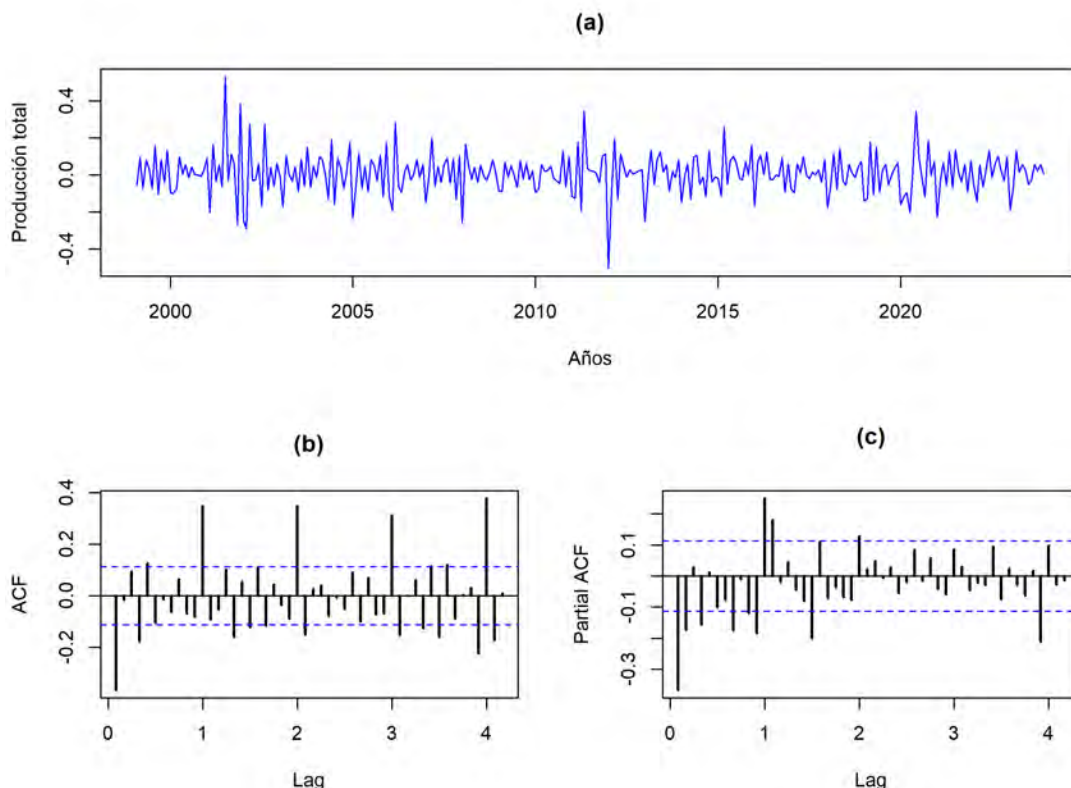


(a) Serie de la producción total mensual de cobre (1999-2023). (b) Función de autocorrelacion simple de la serie (a). (c) Función de autocorrelacion parcial de la serie(a).

Los resultados presentados en la Tabla 3, justifican la necesidad de aplicar transformaciones a la serie para estabilizar la varianza y alcanzar la estacionariedad, condiciones previas para un análisis adecuado de la estacionalidad y el modelado predictivo. En particular, se aplicó la transformación logarítmica para controlar la heterocedasticidad detectada, seguida de la primera diferencia para eliminar la tendencia y estabilizar la media. A continuación, se analiza la serie resultante.

Figura 22

Serie logarítmica diferenciada ($d=1$) y correlogramas de la producción total mensual de cobre Perú (1999-2023)



(a) Serie logarítmica diferenciada $d=1$. (b) Función de autocorrelacion simple. (c) Función de autocorrelacion parcial.

La Figura 22 muestra: (a) el comportamiento de la serie diferenciada, la cual ya no presenta una tendencia visible y oscila alrededor de una media constante cercana a cero, lo que sugiere que la serie podría ser estacionaria. En (b), el correlograma revela la presencia de rezagos significativos en los lags 12, 24, 36, ..., lo que indica un periodo estacional con frecuencia $s = 12$.

Tabla 4

Prueba estadística de tendencia y estacionariedad de la serie logarítmica diferenciada

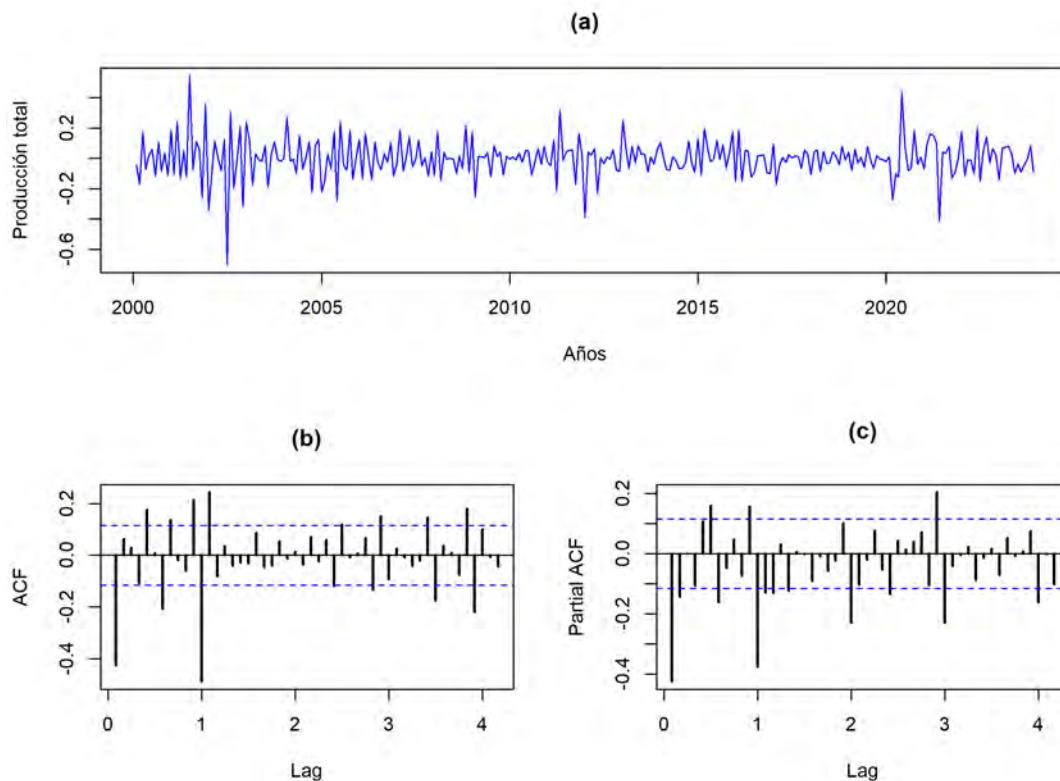
Prueba estadística	Hipótesis nula	Estadístico	P-valor	Decisión
Mann Kendal	No tiene tendencia	0,088	0,93	No Rechazar H_0
Dickey-Fuller	No estacionaria	-8,5271	0,01	Rechazar H_0

La Tabla 4 muestra los resultados de las pruebas de estacionariedad aplicadas a la serie logarít-

mica diferenciada. La prueba de Mann-Kendall no detecta tendencia significativa ($p\text{-valor} = 0.93$), mientras que la prueba de Dickey-Fuller rechaza la hipótesis de no estacionariedad ($p\text{-valor} = 0.01$), indicando que la serie es estacionaria tras la primera diferenciación ($d = 1$). Dado que se observó un periodo estacional ($s = 12$) en la Figura 22, se procede a aplicar una diferenciación estacional adicional ($D = 1$).

Figura 23

Serie logarítmica diferenciada ($D=1$) estacional de la producción de cobre Perú (1999-2023)



(a) Serie estacionaria y no estacional. (b) Función de autocorrelacion simple de la serie estacionaria no estacional. (c) Función de autocorrelacion parcial de la serie serie estacionaria no estacional.

La Figura 23 muestra la serie transformada, la cual es **estacionaria** tras aplicar una diferenciación regular ($d = 1$) y estacional ($D = 1$). El correlograma sugiere posibles órdenes $p = 0, 1, 2$; $q = 0, 1$; $P = 0, 1, 2$; y $Q = 1, 2$. Por ello, se plantearán doce modelos SARIMA para evaluar sus parámetros y seleccionar el más adecuado.

4.3. Identificación de modelos de series temporales para la serie de producción mensual de cobre en Perú (1999-2023).

Las estimaciones de los parámetros en la Tabla 5 corresponden al ajuste de modelos SARIMA sobre la serie transformada mediante logaritmo natural y diferenciada para estabilizar la varianza y eliminar la tendencia. Por tanto, el modelo fue ajustado sobre $\ln(X_t)$ diferenciado, no sobre la serie original.

Tabla 5

Modelos propuestos para la serie logarítmica diferenciada ($d=1$, $D=1$, $s=12$)

Modelo	Parámetros	Estimación	P-valor	Métricas
SARIMA1 (0,1,0)(1,1,1)₁₂	Φ_1	-0,0394	0,5563	AIC = -529,63
	Θ_1	-0,9256	0	BIC = -518,65
SARIMA2 (0,1,0)(2,1,0)₁₂	Φ_1	-0,6629	0	AIC = -467,32
	Φ_2	-0,3504	0	BIC = -456,34
SARIMA3 (1,1,1)(0,1,2)₁₂	ϕ_1	0,0253	0,8691	
	θ_1	-0,4416	0,0016	AIC = -570,69
	Θ_1	-0,9093	0	BIC = -552,39
	Θ_2	-0,0178	0,7850	
SARIMA4 (1,1,1)(1,1,1)₁₂	ϕ_1	0,0253	0,8691	
	θ_1	-0,4417	0,0016	AIC = -570,69
	Φ_1	0,0191	0,7860	BIC = -552,39
	Θ_1	-0,9286	0	
SARIMA5 (1,1,2)(0,1,1)₁₂	ϕ_1	-0,9269	0	
	θ_1	0,5298	0	AIC = -572,81
	θ_2	-0,4196	0	BIC = -554,51
	Θ_1	-0,9075	0	

Modelo	Parámetros	Estimación	P-valor	Métricas
SARIMA6 (1,1,2)(0,1,2)₁₂	ϕ_1	-0,9255	0	AIC = -570,93 BIC = -548,98
	θ_1	0,5272	0	
	θ_2	-0,4212	0	
	Θ_1	-0,8931	0	
	Θ_2	-0,0224	0,7266	
SARIMA7 (1,1,2)(1,1,1)₁₂	ϕ_1	-0,9258	0	AIC = -570,93 BIC = -548,98
	θ_1	0,5274	0	
	θ_2	-0,4212	0	
	Φ_1	0,0244	0,728	
	Θ_1	-0,9177	0	
SARIMA8 (2,1,0)(0,1,1)₁₂	ϕ_1	-0,4058	0	AIC = -571,94 BIC = -557,31
	ϕ_2	-0,1581	0,0071	
	Θ_1	-0,9151	0	
SARIMA9 (2,1,0)(1,1,1)₁₂	ϕ_1	-0,4083	0	AIC = -570,02 BIC = -551,73
	ϕ_2	-0,1578	0,0072	
	Φ_1	0,0197	0,7795	
	Θ_1	-0,9230	0	
SARIMA10 (2,1,0)(2,1,0)₁₂	ϕ_1	-0,4523	0	AIC = -517,56 BIC = -499,27
	ϕ_2	-0,1505	0,0108	
	Φ_1	-0,6344	0	
	Φ_2	-0,3254	0	
SARIMA11 (2,1,1)(0,1,1)₁₂	ϕ_1	0,1172	0,8094	AIC = -570,65 BIC = -552,35
	ϕ_2	0,0392	0,8579	
	θ_1	-0,5311	0,2677	
	Θ_1	-0,9209	0	

Modelo	Parámetros	Estimación	P-valor	Métricas
SARIMA12 (2,1,2)(0,1,1)₁₂	ϕ_1	-0,9722	0	AIC = -570,91 BIC = -548,95
	ϕ_2	-0,0469	0,7623	
	θ_1	0,5689	0,0002	
	θ_2	-0,3799	0,0090	
	Θ_1	-0,9044	0	

A continuación, se nombran los modelos que cumplieron con la significancia de los coeficientes de estimación con sus respectivos parámetros:

- **SARIMA2 (0,1,0)(2,1,0)₁₂**
- **SARIMA5 (1,1,2)(0,1,1)₁₂**
- **SARIMA8 (2,1,0)(0,1,1)₁₂**
- **SARIMA10 (2,1,0)(2,1,0)₁₂**

Ahora verificaremos si los residuos de los modelos elegidos cumplen con la prueba de raíz unitaria y si siguen una distribución normal.

Verificación del modelo $SARIMA2 (0,1,0)(2,1,0)_{12}$

En la Figura 24, podemos observar en el ACF y PACF que los residuos del modelo no son ruido blanco tienen lags significativos lo que podría indicar que no es un buen modelo.

Figura 24

Correlogramas de los residuos del modelo $SARIMA2 (0,1,0)(2,1,0)_{12}$

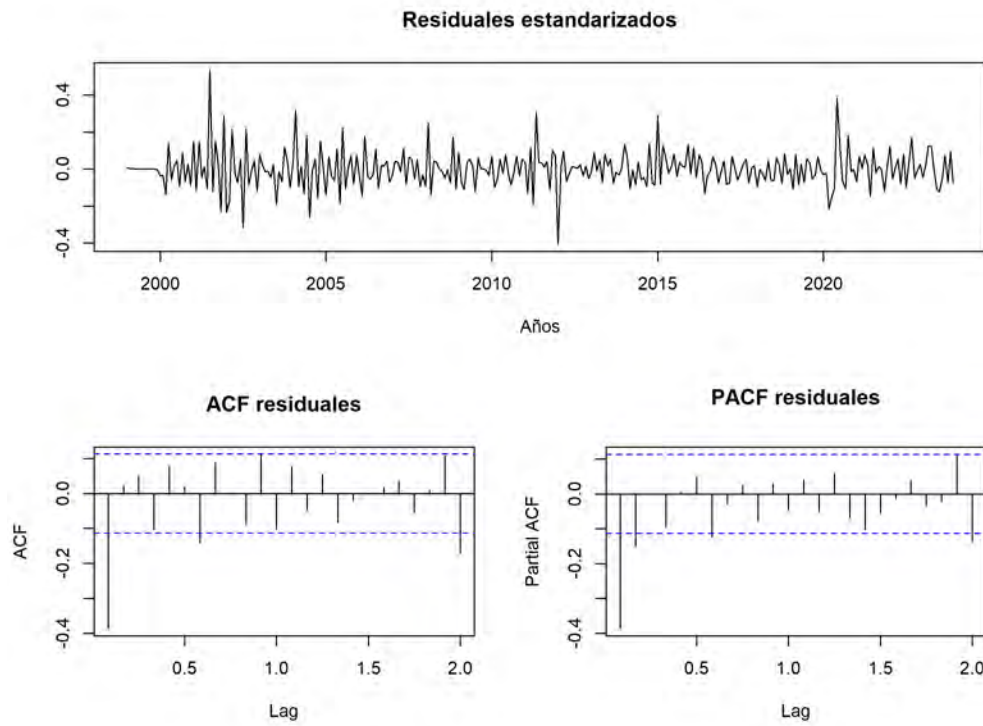


Tabla 6

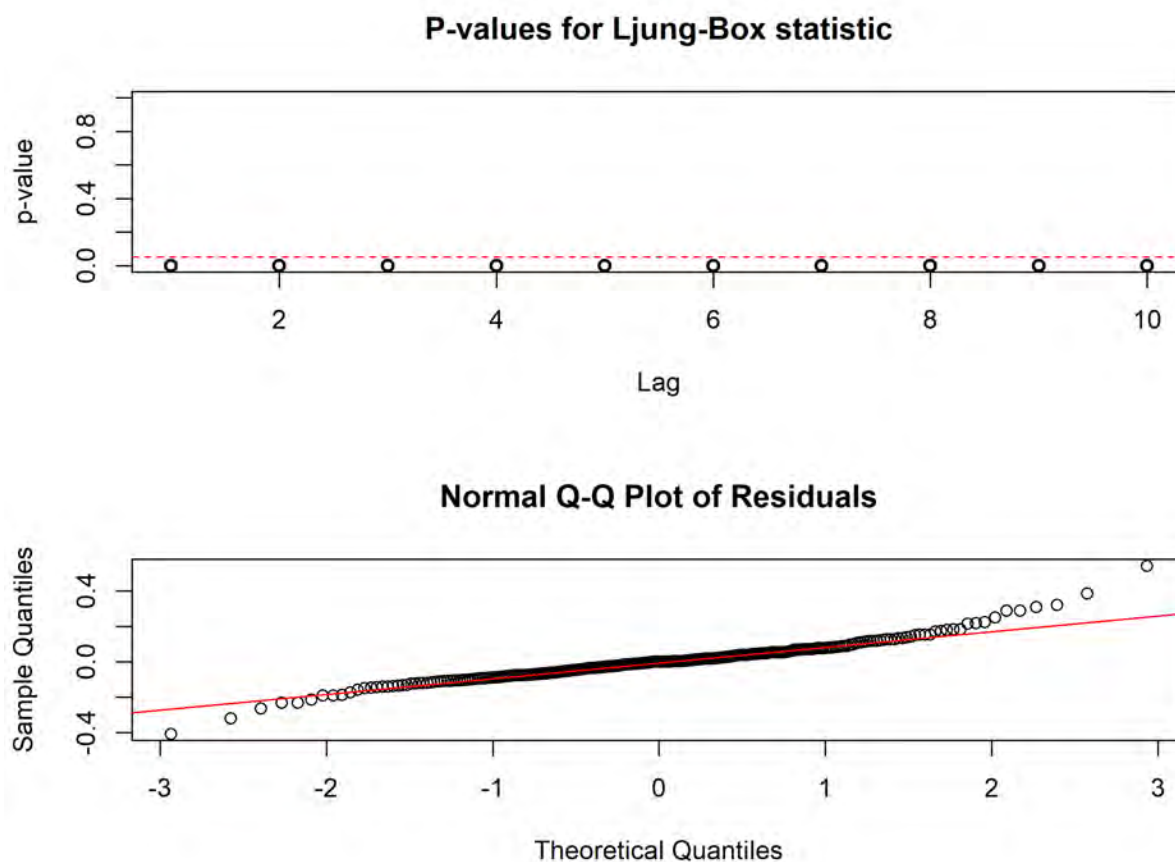
Pruebas de validación para los residuos del modelo $SARIMA2 (0,1,0)(2,1,0)_{12}$

Prueba estadística	Hipótesis nula	Estadístico	P-valor	Decisión
Ljung-Box	Es ruido blanco	69,517	$3,943 \times 10^{-10}$	Rechazar H_0
Jarque-Bera	Normal	203,83	$2,2 \times 10^{-16}$	Rechazar H_0

En la Tabla 6, se confirma que el modelo no es adecuado, ya que los p-valores de la prueba de Ljung-Box para varios lags son menores a 0,05, indicando autocorrelación significativa en los residuos. En la Figura 25, esto se observa claramente, ya que los p-valores se encuentran por debajo de la banda de referencia, y los residuos no siguen una distribución normal al no alinearse con la recta en el gráfico Q-Q.

Figura 25

Ljung-Box y Gráfico Q-Q de los residuales del modelo SARIMA2 (0,1,0)(2,1,0)₁₂

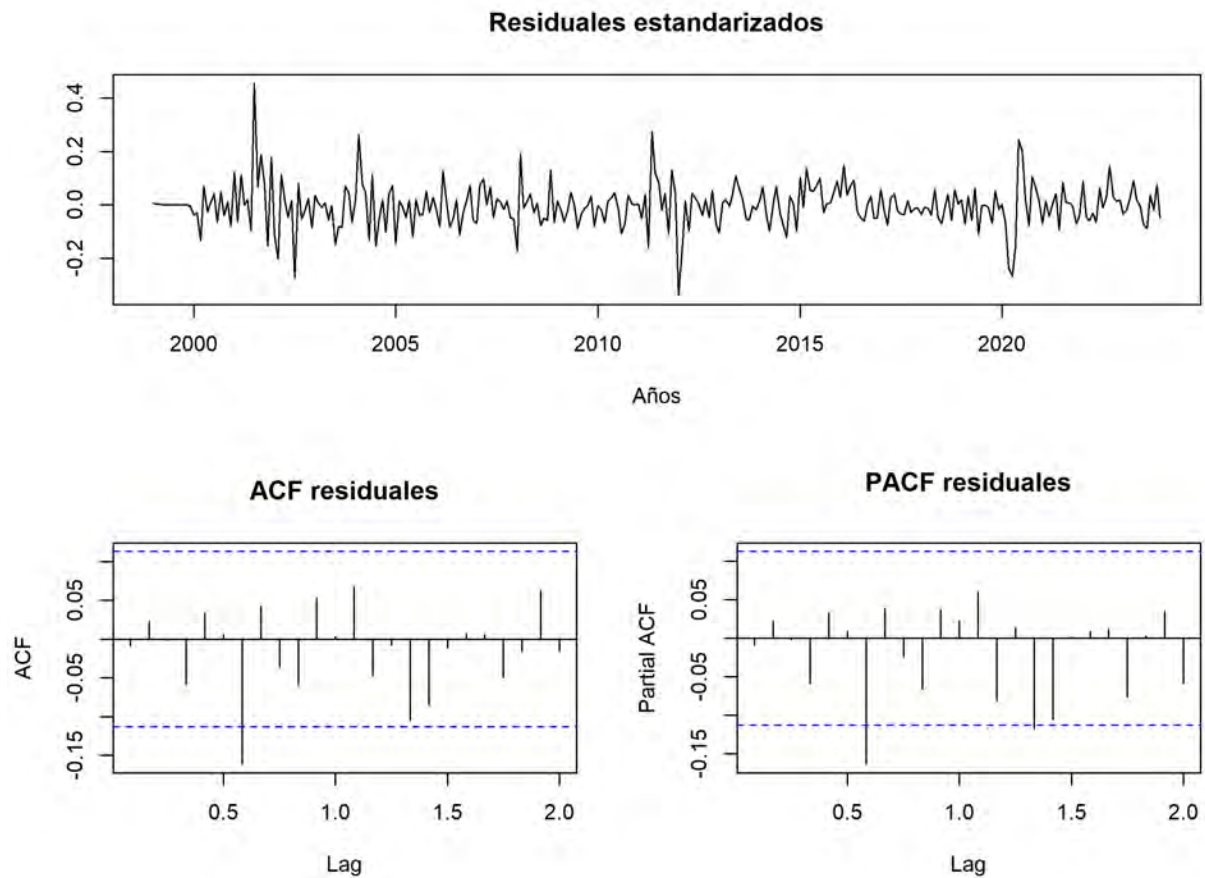


Verificación del modelo SARIMA5 (1,1,2)(0,1,1)₁₂

En la Figura 26, podemos observar algunos lags ligeramente sobresalientes, la mayoría están dentro de las bandas de confianza en el correlograma ACF, lo cual no indica claramente si el modelo es bueno.

Figura 26

Correlogramas de los residuales del modelo $SARIMA5(1,1,2)(0,1,1)_{12}$

**Tabla 7**

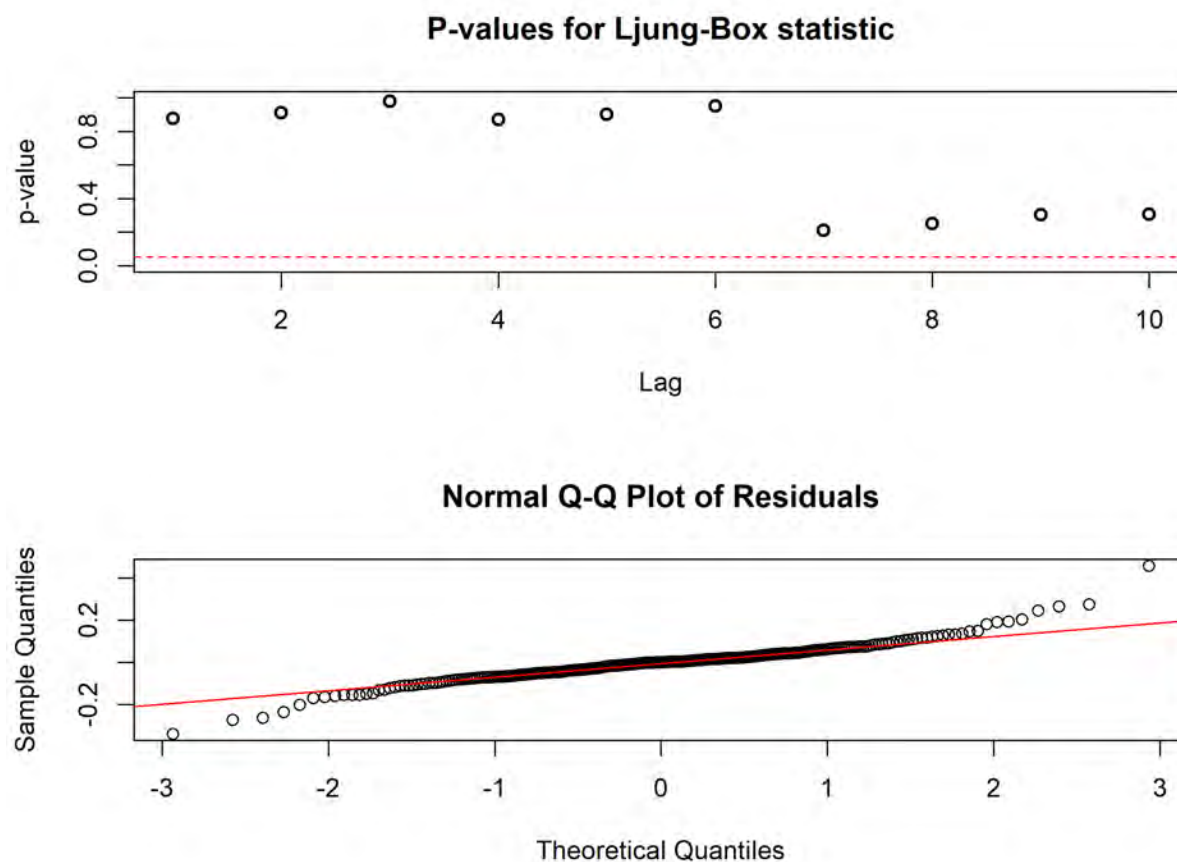
Pruebas de validación para los residuos del modelo $SARIMA5(1,1,2)(0,1,1)_{12}$

Prueba estadística	Hipótesis nula	Estadístico	P-valor	Decisión
Ljung-Box	Es ruido blanco	12,6	0,3987	No Rechazar H_0
Jarque-Bera	Normal	279,36	$2,2 \times 10^{-16}$	Rechazar H_0

En la Tabla 7, se confirma que el modelo es adecuado, ya que los p-valores de la prueba de Ljung-Box para varios lags son mayores a 0,05, indicando autocorrelación significativa en los residuos. En la Figura 27, esto se observa claramente, ya que los p-valores se encuentran por encima de la banda de referencia, y los residuos no siguen una distribución normal al no alinearse con la recta en el gráfico Q-Q.

Figura 27

Ljung-Box y Gráfico Q-Q de los residuales del modelo SARIMA5 (1,1,2)(0,1,1)₁₂

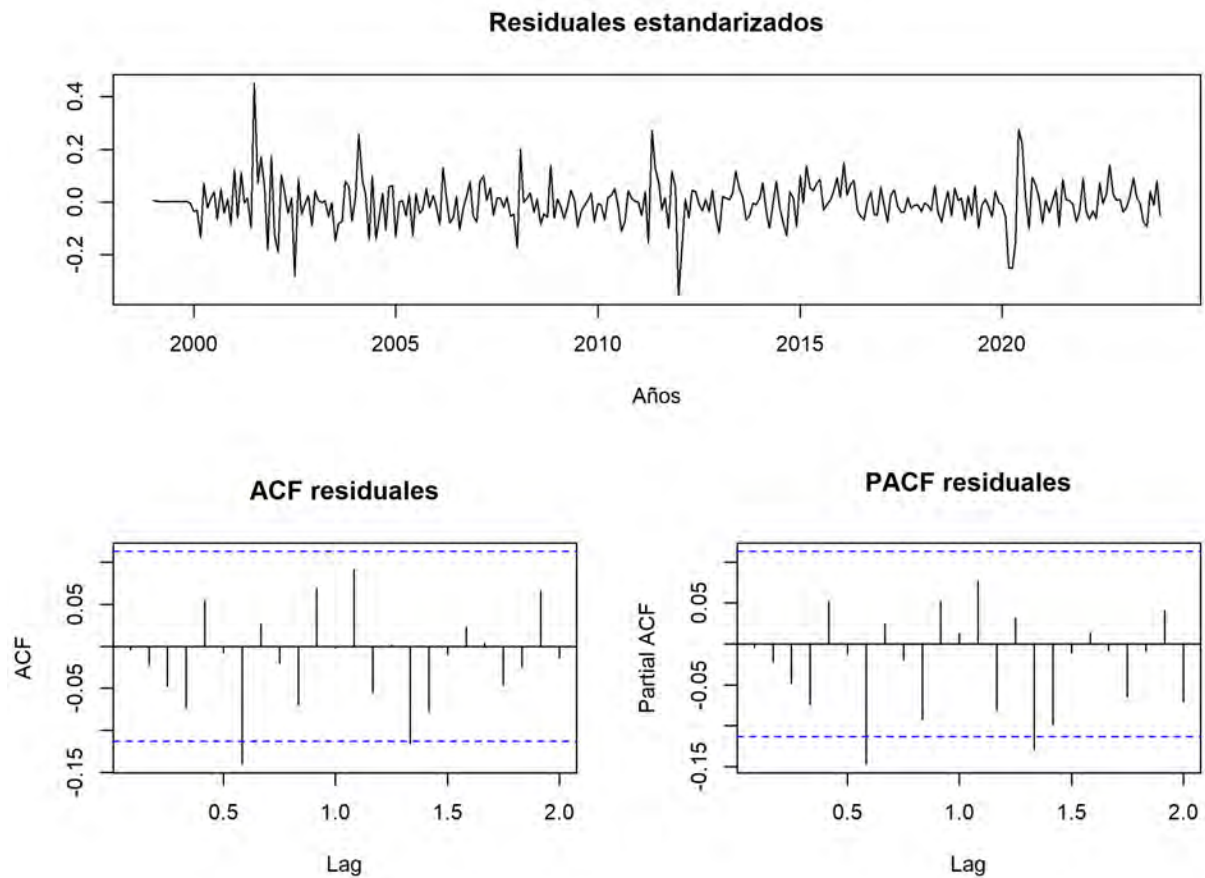


Verificación del modelo SARIMA8 (2,1,0)(0,1,1)₁₂

En la Figura 28, podemos observar que los ACF de los residuos del modelo tienen lags significativos lo que podría indicar que no se comportan como ruido blanco.

Figura 28

Correlogramas de los residuales del modelo $SARIMA8(2,1,0)(0,1,1)_{12}$



La Tabla 8, nos confirma que el modelo es adecuado, ya que los p-valores de la prueba de Ljung-Box para varios lags es mayor a 0,05, indicando autocorrelación significativa en los residuos. En la Figura 29, esto se observa claramente, ya que los p-valores se encuentran por encima de la banda de referencia, y los residuos no siguen una distribución normal al no alinearse con la recta en el gráfico Q-Q.

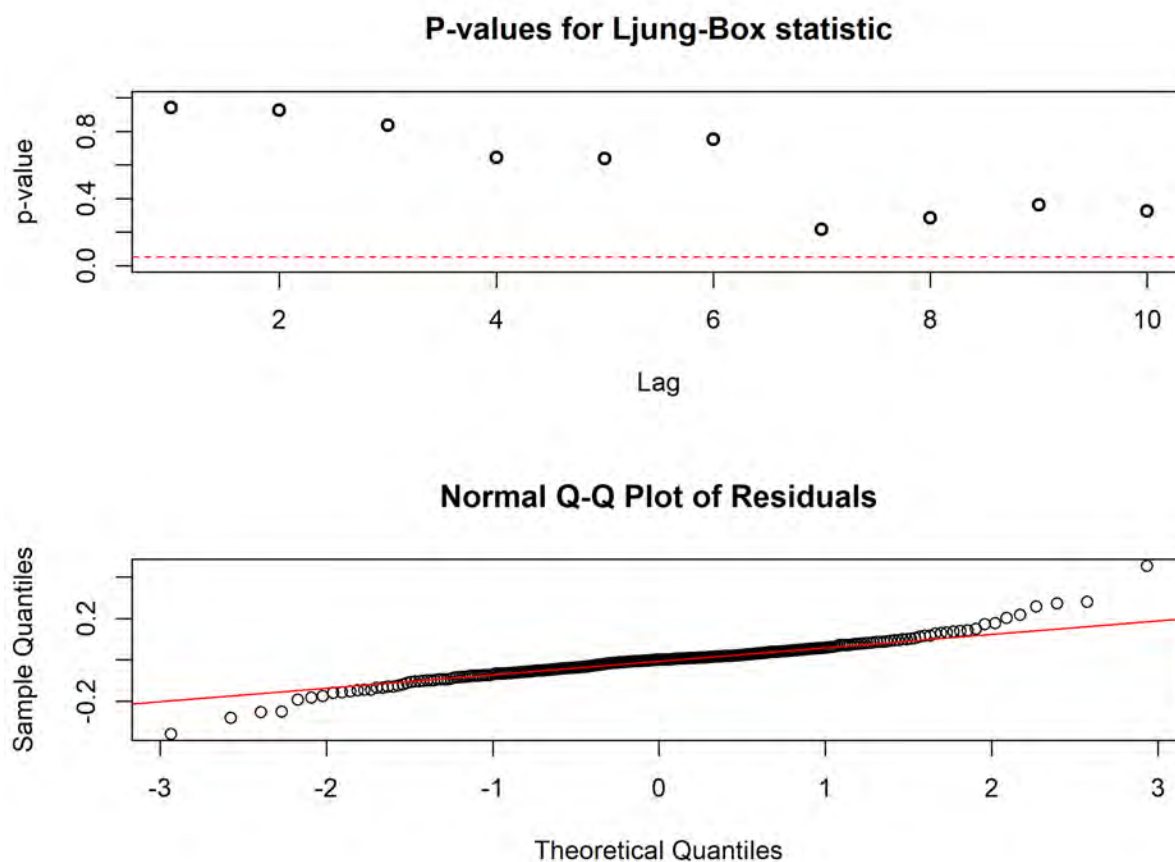
Tabla 8

Pruebas de validación para los residuos del modelo $SARIMA8(2,1,0)(0,1,1)_{12}$

Prueba estadística	Hipótesis nula	Estadístico	P-valor	Decisión
Ljung-Box	Es ruido blanco	12,926	0,3744	No Rechazar H_0
Jarque-Bera	Normal	287,19	$2,2 \times 10^{-16}$	Rechazar H_0

Figura 29

Ljung-Box y Gráfico Q-Q de los residuales del modelo SARIMA8 (2,1,0)(0,1,1)₁₂

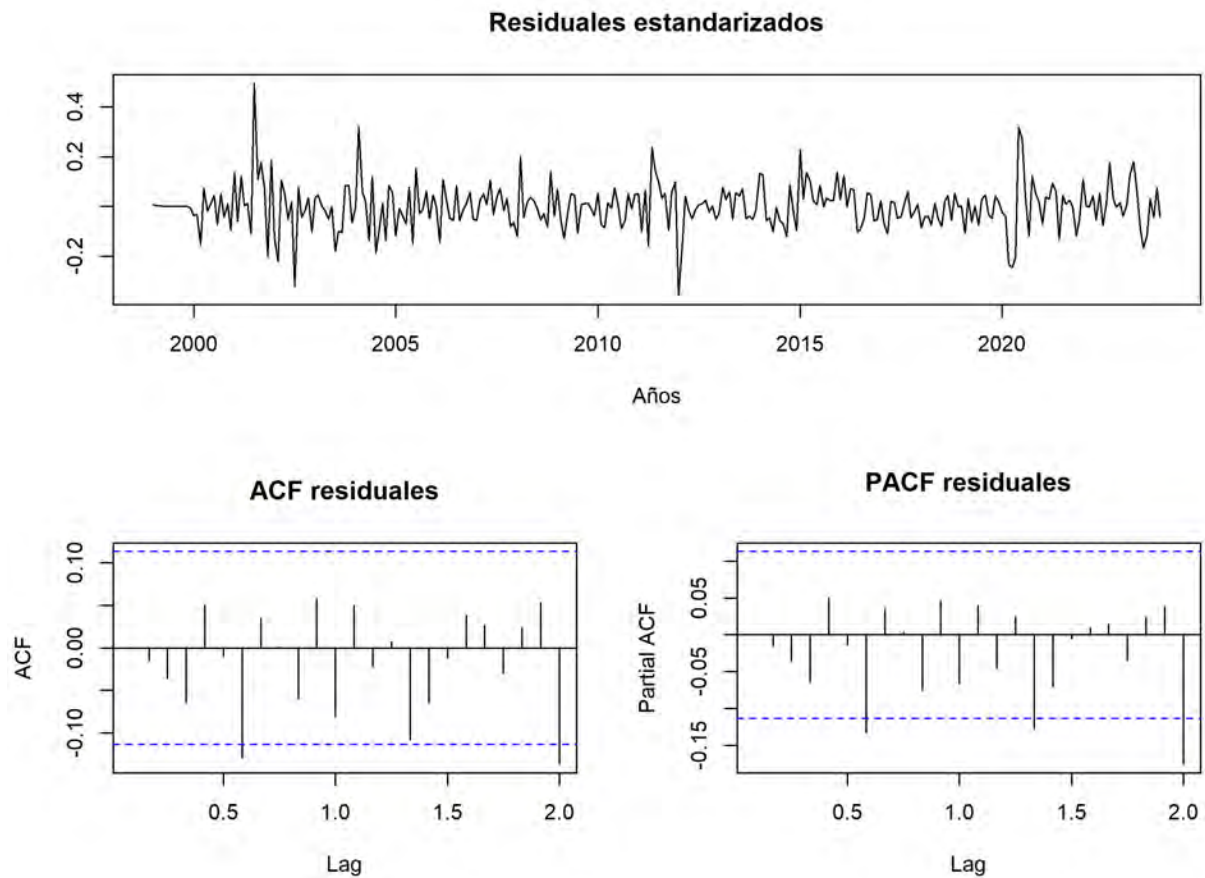


Verificación del modelo SARIMA10 (2,1,0)(2,1,0)₁₂

En la Figura 30, podemos observar que los ACF de los residuos del modelo tienen lags significativos lo que podría indicar que no se comportan como ruido blanco.

Figura 30

Correlogramas de los residuales del modelo $SARIMA_{10}(2,1,0)(2,1,0)_{12}$



La Tabla 9, confirma que el modelo es adecuado, ya que los p-valores de la prueba de Ljung-Box para varios lags es mayor a 0,05, indicando autocorrelación significativa en los residuos. En la Figura 31, esto se observa claramente, ya que los p-valores se encuentran por encima de la banda de referencia, y los residuos no siguen una distribución normal al no alinearse con la recta en el gráfico Q-Q.

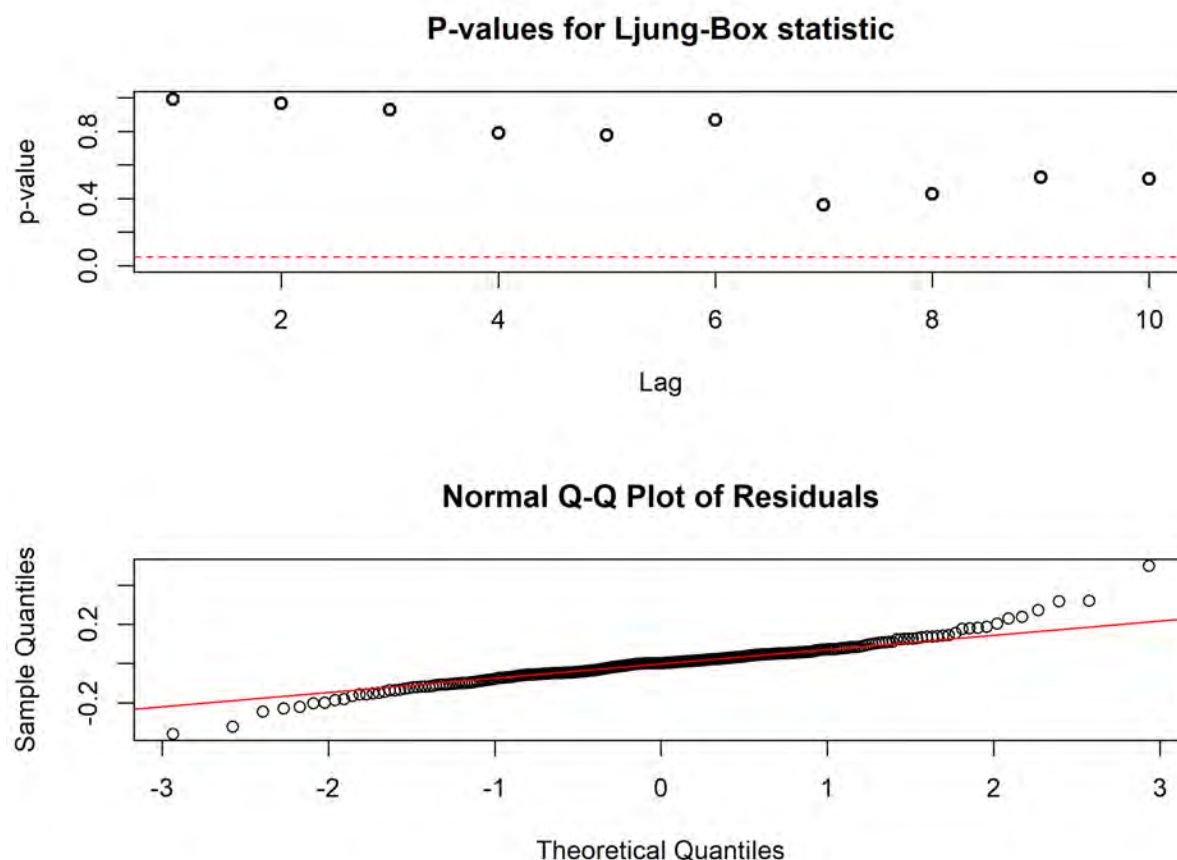
Tabla 9

Pruebas de validación para los residuos del modelo $SARIMA_{10}(2,1,0)(2,1,0)_{12}$

Prueba estadística	Hipótesis nula	Estadístico	P-valor	Decisión
Ljung-Box	Es ruido blanco	12,248	0,4259	No Rechazar H_0
Jarque-Bera	Normal	193,31	$2,2 \times 10^{-16}$	Rechazar H_0

Figura 31

Ljung-Box y Gráfico Q-Q de los residuales del modelo SARIMA10 (2,1,0)(2,1,0)₁₂



4.4. Validación del modelo de la producción mensual de cobre para el año 2024

De los modelos SARIMA propuestos, se puede observar que el modelo 5, SARIMA(1,1,2)(0,1,1)₁₂, modelo 8, SARIMA(2,1,0)(0,1,1)₁₂ y el modelo 10, SARIMA(2,1,0)(2,1,0)₁₂, cumplen con los criterios establecidos. En particular, sus parámetros resultaron significativos y sus residuos se comportan como un proceso independiente, lo que indica que estos modelos son adecuados para capturar la estructura temporal de la serie analizada.

Tabla 10

Valores pronosticados para la producción total mensual de cobre para el año 2024.

Modelos	Real	SARIMA5	SARIMA8	SARIMA10
Enero	205377	222175	223486	221571
Febrero	216752	207831	208273	204424
Marzo	219011	228539	230049	221886
Abril	203905	216596	216837	214671
Mayo	231509	231736	233334	229865
Junio	213578	243450	243302	236289
Julio	222389	244678	245981	241088
Agosto	246568	252480	252810	243657

En la Tabla 10, se puede ver que el modelo SARIMA10 parece ser el modelo mas adecuado para predecir la producción de cobre, ya que proporciona valores mas cercanos a los datos reales.

Figura 32

Valores pronosticados del modelo 5, 12 de la producción mensual de cobre en Perú, periodo Enero 2024 a Agosto 2024.

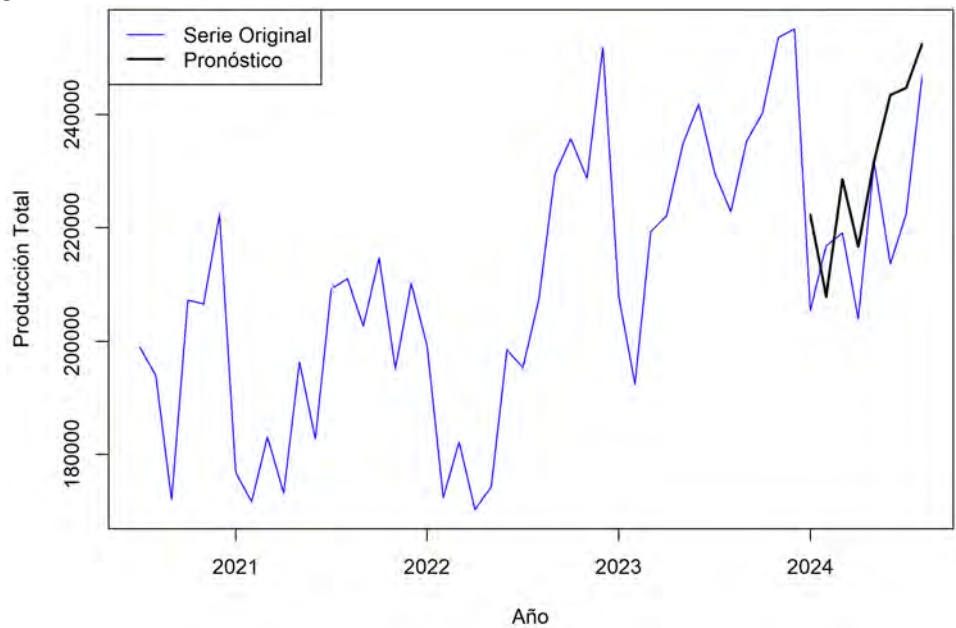
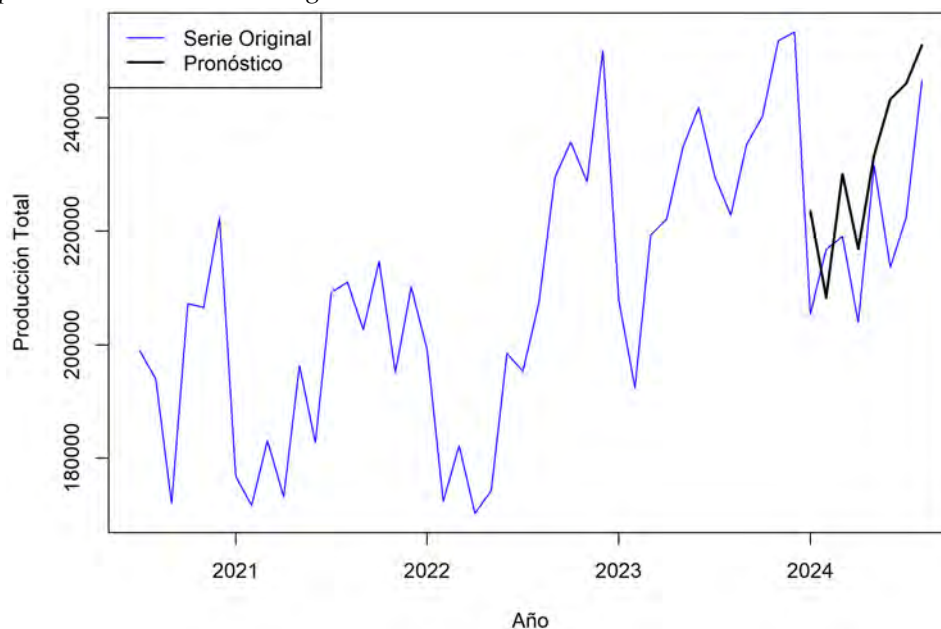
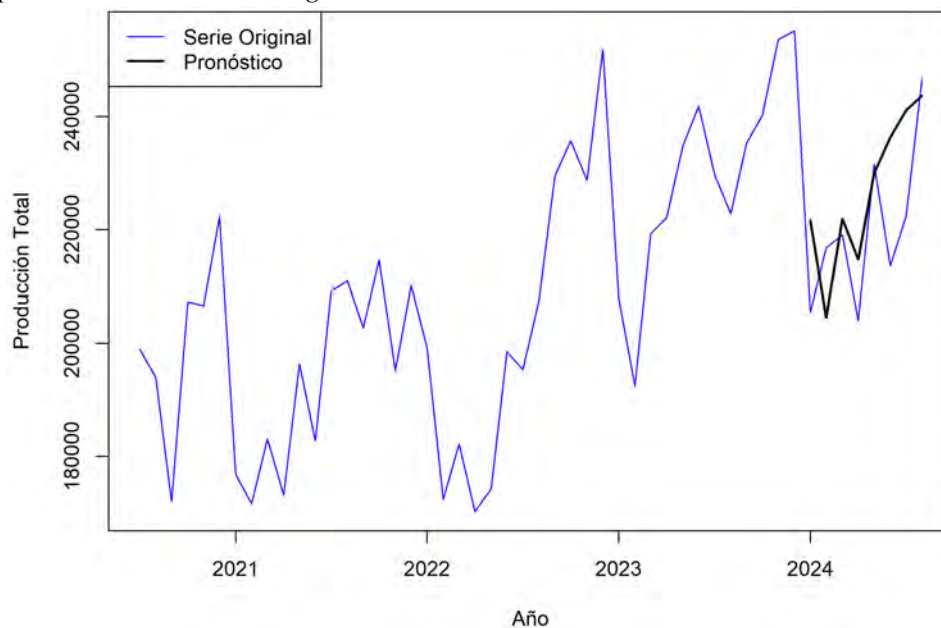


Figura 33

Valores pronosticados del modelo 8, $SARIMA(2,1,0)(0,1,1)_{12}$ de la producción mensual de cobre en Perú, periodo Enero 2024 a Agosto 2024.

**Figura 34**

Valores pronosticados del modelo 10, $SARIMA(2,1,0)(2,1,1)_{12}$ de la producción mensual de cobre en Perú, periodo Enero 2024 a Agosto 2024.



Validación del Pronóstico

A continuación se evalúa el mejor modelo de pronóstico con el menor error.

Tabla 11

Comparación de la evaluación de los errores de pronósticos de los modelos SARIMA propuestos.

Modelos	RMSE	MAE	MAPE	SSE
SARIMA5	15959.87	13279.75	6.17 %	203774
SARIMA8	16474.58	13992.62	6.49 %	217129
SARIMA10	13299.51	11016	5.14 %	141501

La Tabla 11, presenta las métricas de validación de los modelos SARIMA5, SARIMA8 y SARIMA10 para el pronóstico de la producción de cobre en 2024. Se observa que SARIMA10 tiene los menores valores en RMSE, MAE, MAPE y SSE, lo que indica una mayor precisión en sus estimaciones. A comparación, SARIMA5 y SARIMA8 muestran errores más altos, lo que indica un menor desempeño en la predicción.

Desarrollo de modelo 10, SARIMA(2,1,0)(2,1,0)₁₂

$$(1 - \Phi_1 B^{12} - \Phi_2 B^{24})(1 - \phi_1 B - \phi_2 B^2)(1 - B)(1 - B^{12})X_t = a_t$$

$$(1 - \Phi_1 B^{12} - \Phi_2 B^{24})(1 - \phi_1 B - \phi_2 B^2)(X_t - X_{t-1} - X_{t-12} + X_{t-13}) = a_t$$

Aplicando componente autoregresivo regular

$$X_t - (1 + \phi_1)X_{t-1} + (\phi_1 - \phi_2)X_{t-2} + \phi_2 X_{t-3} - (1 + \phi_1)X_{t-12} + (1 + \phi_1)X_{t-13} + (\phi_2 - \phi_1)X_{t-14} - \phi_2 X_{t-15}$$

Aplicando componente autoregresivo estacional

$$\begin{aligned}
& X_t - (1 + \phi_1)X_{t-1} + (\phi_1 - \phi_2)X_{t-2} + \phi_2 X_{t-3} - (1 + \Phi_1)X_{t-12} + (1 + \phi_1)(1 + \Phi_1)X_{t-13} \\
& - (\phi_1 - \phi_2)(1 + \Phi_1)X_{t-14} - \phi_2(1 + \Phi_1)X_{t-15} + (\Phi_1 - \Phi_2)X_{t-24} - (1 + \phi_1)(\Phi_1 - \Phi_2)X_{t-25} \\
& + (\phi_1 - \phi_2)(\Phi_1 - \Phi_2)X_{t-26} + \phi_2(\Phi_1 - \Phi_2)X_{t-27} + \Phi_2 X_{t-36} - \Phi_2(1 + \phi_1)X_{t-37} \\
& + \Phi_2(\phi_1 - \phi_2)X_{t-38} + \phi_2 X_{t-39}
\end{aligned}$$

Ecuación final.

$$\begin{aligned}
X_t = & (1 + \phi_1)X_{t-1} - (\phi_1 - \phi_2)X_{t-2} - \phi_2 X_{t-3} + (1 + \Phi_1)X_{t-12} - (1 + \phi_1)(1 + \Phi_1)X_{t-13} \\
& + (\phi_1 - \phi_2)(1 + \Phi_1)X_{t-14} + \phi_2(1 + \Phi_1)X_{t-15} - (\Phi_1 - \Phi_2)X_{t-24} + (1 + \phi_1)(\Phi_1 - \Phi_2)X_{t-25} \\
& - (\phi_1 - \phi_2)(\Phi_1 - \Phi_2)X_{t-26} - \phi_2(\Phi_1 - \Phi_2)X_{t-27} - \Phi_2 X_{t-36} + \Phi_2(1 + \phi_1)X_{t-37} \\
& - \Phi_2(\phi_1 - \phi_2)X_{t-38} - \phi_2 X_{t-39} + a_t
\end{aligned}$$

Reemplazando los valores de: $\phi_1 = -0,4523$, $\phi_2 = -0,1505$, $\Phi_1 = -0,6344$, $\Phi_2 = -0,3254$

$$\begin{aligned}
X_t = & 0,5477X_{t-1} + 0,3018X_{t-2} + 0,1505X_{t-3} + 0,3656X_{t-12} - 0,2003X_{t-13} + 0,1103X_{t-14} \\
& + 0,0550X_{t-15} + 0,3090X_{t-24} - 0,1691X_{t-25} + 0,0933X_{t-26} + 0,0465X_{t-27} - 0,3254X_{t-36} \\
& + 0,1783X_{t-37} - 0,0982X_{t-38} + 0,1505X_{t-39} + a_t
\end{aligned} \tag{9}$$

4.5. Pronóstico de la producción mensual de cobre en Perú (Septiembre 2024 - Agosto 2025)

Una vez validado nuestro modelo, siendo el SARIMA(2,1,0)(2,1,0)₁₂ el que mejor capta el comportamiento de la serie de la producción mensual de cobre en Perú y haber sido el que mejor validación muestra en función a la data de prueba. Procedemos a realizar el pronóstico para la producción de cobre en Perú desde Septiembre 2024 a Agosto de 2025 (Tabla 12).

Tabla 12

Pronóstico de la producción de cobre según el modelo SARIMA(2,1,0)(2,1,0)₁₂

Año	Meses	Pronóstico
2024	SEPTIEMBRE	253076
2024	OCTUBRE	261967
2024	NOVIEMBRE	256625
2024	DICIEMBRE	271469
2025	ENERO	238838
2025	FEBRERO	215944
2025	MARZO	235927
2025	ABRIL	228864
2025	MAYO	240581
2025	JUNIO	255861
2025	JULIO	252336
2025	AGOSTO	255989

DISCUSIÓN

De acuerdo a la investigación de Mutele y Carranza, titulada “*Análisis estadístico de la producción de oro en Sudáfrica utilizando técnicas de modelado ARIMA, VAR y ARNN: extrapolación de la producción futura de oro, agotamiento de recursos y reservas e implicancia en la exploración de oro de Sudáfrica*”, analizaron datos anuales y llegaron a la conclusión que el modelo ARIMA (1,1,0) resulto ser el mas eficaz para la producción futura de oro, en el estudio no se menciona explícitamente un análisis de la significancia de los valores estimados de los parámetros en los modelos ARIMA utilizados, los cuales en comparación con el presente estudio donde si se realizo este análisis para poder elegir el modelo adecuado.

La investigación realizada por Contreras Cerpa utilizó modelos de redes neuronales recurrentes (RNN), GRU y LSTM para prever la producción de cobre, incorporando variables externas como el consumo de agua y energía. En este estudio el modelo LSTM demostró ser el más preciso, lo que sugiere que técnicas de aprendizaje profundo pueden ofrecer una mejor predicción de variables complejas en la minería. En el presente estudio se utilizaron modelos SARIMA, pero aplicar modelos de redes neuronales podría ser una extensión valiosa para mejorar la precisión de las proyecciones en el futuro.

En el trabajo realizado por Aycaya-Paco et al. desarrollaron una herramienta en R-studio haciendo uso de los modelos ARIMA y el modelo de espacios de estados para pronosticar la extracción del cobre en el Perú, dicho estudio al igual que el nuestro validan el modelo usando el análisis residual con la prueba de Ljung Box pero analizan la normalidad con la prueba de Shapiro Wilk distinta a la que usamos en nuestro estudio que fue la de Jarque Bera, donde se pudo observar que siguen una distribución distinta a la Normal.

El trabajo titulado “*Pronóstico del precio mensual de Cobre con Modelos de Series de Tiempo*” realizado por Cotrina Teatino et al. pronostico el precio mensual del cobre mediante cinco modelos de series temporales: promedio móvil, promedio móvil ponderado, suavizado exponencial simple, suavizado exponencial de Holt y suavizado exponencial de Winters, resultando con menor error porcentual absoluto medio (MAPE) el modelo de suavización exponencial simple. Dicho estudio

empleó una metodología distinta a la utilizada en el presente trabajo, sin embargo ambos coinciden en la importancia de seleccionar el modelo óptimo en función de métricas de precisión, como el MAPE, lo cual minimiza el error del pronóstico, asegurando una mejor capacidad predictiva.

Por otro lado, en el trabajo realizado por Chafloque Céspedes et al., titulado “*Modelación de las dinámicas, estimación y análisis de las exportaciones mineras: caso peruano*”, se compararon los métodos predictivos de ARIMA y SARIMA para la estimación de las exportaciones mineras, siendo entre ellas el modelo SARIMA el que ofrecía mejor desempeño en la predicción. En nuestro estudio, dado que la serie de producción de cobre presentaba estacionalidad, se exploraron exclusivamente diferentes modelos SARIMA, sin comparar con el modelo ARIMA, de manera que capture de manera óptima la dinámica temporal de la serie. En ambos estudios resaltan la importancia de los pronósticos en el sector minero.

Como se mencionó en el marco teórico, la no normalidad de los residuos puede afectar la precisión de los intervalos de pronóstico. En este estudio, tanto la prueba de Jarque-Bera como los gráficos Q-Q mostraron que los residuos de los modelos ajustados no siguen una distribución normal. Si bien esto no invalida por completo los resultados obtenidos, sí limita la capacidad del modelo para generar predicciones con un grado de confianza adecuado. En ese sentido, los modelos utilizados podrían mejorarse incorporando un análisis más detallado de la variabilidad en los errores, a través de enfoques como los modelos ARCH o GARCH.

CONCLUSIONES

1. Se identificaron las características de la serie original de la producción total de cobre en Perú en el periodo de 1999-2023, la cual presento tendencia creciente, no fue estacionaria, presento estacionalidad y heterocedasticidad cada 12 meses, por ello aplicamos transformaciones logarítmicas a la serie original con una diferencia regular ($d = 1$) y una diferencia estacional ($D = 1$) en un período de $s = 12$ meses, donde se logró alcanzar estacionariedad, como confirmó la prueba de Dickey-Fuller.
2. Se analizaron doce modelos SARIMA distintos, de los cuales solo tres superaron la prueba de Ljung-Box: **SARIMA(0,1,0)(2,1,0)₁₂**, **SARIMA(1,1,2)(0,1,1)₁₂** y **SARIMA(2,1,0)(2,1,0)₁₂**.
3. Se evaluaron las métricas de pronóstico para estos modelos, determinándose que el modelo SARIMA(2,1,0)(2,1,0)₁₂ presentó el mejor desempeño, con un MAPE de 5,14 % consolidándose como la mejor opción según la metodología de Box-Jenkins. La ecuación (9) describe este modelo.

RECOMENDACIONES

- Se recomienda, para futuros estudios, explorar la combinación de modelos SARIMA con componentes ARCH/GARCH, como lo propuesto en Oktaviani et al. (2021), con el objetivo de mejorar la captura de la heterocedasticidad en la serie de producción de cobre. Este tipo de modelos es especialmente útil cuando se detecta variabilidad condicional en los residuos, situación común en series económicas
- Además, se podría llevar un estudio a futuro sobre el seguimiento continuo de la producción de cobre en el país. Haciendo uso de modelos de series temporales junto con técnicas de estadística espacial que podrían ayudar a identificar patrones y cambios significativos a lo largo del tiempo.
- Por último, sería muy útil aplicar modelos de series temporales multivariadas, ya que permitirían analizar la producción de cobre junto con otros factores clave, como los precios internacionales, la demanda del mercado, el clima y las políticas económicas. Esto daría una visión más completa de cómo interactúan estas variables y ayudaría a tomar decisiones más informadas.

REFERENCIAS

- Astudillo, G., Carrasco, R., Fernández-Campusano, C., & Chacón, M. (2020). Copper Price Prediction Using Support Vector Regression Technique. *Applied Sciences*, 10(19). <https://doi.org/10.3390/app10196648>
- Aycaya-Paco, Y. B., Vilca-Mamani, L. D., & Torres-Cruz, F. (2023). Peru Mining: Analysis and Forecast of Mining Production in Peru Using Time Series and Data Science Techniques. *arXiv*. <https://arxiv.org/abs/2307.06293>
- Banco Central de Reserva del Perú. (2023). Informe Económico Anual [Recuperado el 08 de Agosto del 2025]. <https://www.bcrp.gob.pe>
- Banco Mundial. (2019). Impacto de la Volatilidad de Precios en el Sector Minero Peruano [Recuperado el 08 de Agosto del 2025]. <https://www.worldbank.org>
- Billingsley, P. (1995). *Probability and Measure* (3rd). Wiley.
- Box, G. E. P., & Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2), 211-243. <https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>
- Box, G. E. P., & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day.
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time Series Analysis Forecasting and Control* (5th). Wiley.
- Brockwell, P. J., & Davis, R. A. (2016). *Introduction to Time Series and Forecasting* (3rd). Springer. <https://doi.org/10.1007/978-3-319-29854-2>
- Centro de investigación de economía y negocios globales. (2024). Reporte de exportaciones [Recuperado el 27 de Agosto del 2024]. https://www.cien.adexperu.org.pe/wp-content/uploads/2024/03/CIEN_REX_ene_2024.pdf
- Chafloque Céspedes, M. R., Chung Alva, V. M., & Cotrina Camacho, A. B. (2018). Modelación de las dinámicas, estimación y análisis de las exportaciones mineras: caso peruano. *Alternativas*, 19(3), 37-45. <https://doi.org/10.23878/alternativas.v19i3.264>

- CodelcoEduca. (2024). ¡Conoce y aprende del proceso productivo del cobre! [Recuperado el 20 de junio del 2024]. https://www.codelcoeduca.cl/codelcoeduca/site/edic/base/port/proceso_productivo.html
- Contreras, A. J., Aziry, C. Z., Martínez, J. L. F., & Sánchez, D. P. (2016). Análisis de series de tiempo en el pronóstico de la demanda de almacenamiento de productos perecederos. *Estudios Gerenciales*, 32(141), 387-396. <https://doi.org/https://doi.org/10.1016/j.estger.2016.11.002>
- Contreras Cerpa, P. S. (2021). *Proyecciones de producción de cobre considerando consumo de agua, energía y emisiones de gases de efecto invernadero, con redes neuronales recurrentes*. [Tesis de Maestría]. Universidad de Concepción de Chile. <https://repositorio.udec.cl/items/ecf062c1-43c0-4868-b0b8-7c2bbfc29447>
- Correa, J. C., Iral, R., & Rojas, L. (2006). Estudio de potencia de pruebas de homogeneidad de varianza. *Revista colombiana de estadística*, 29(1), 57-76.
- Cotrina Araujo, N. A. (2020). *Análisis predictivo de series temporales de temperatura corporal* [Tesis de licenciatura]. Pontificia Universidad Católica del Perú. <https://repositorio.pucp.edu.pe/index/handle/123456789/173215>
- Cotrina Teatino, M. A., Marquina Araujo, J. J., Noriega Vidal, E. M., Vega Gonzalez, J. A., Arango Retamozo, S. M., Portilla Rodriguez, H. R., & Castillo Chung, A. R. (2022). Copper Monthly Price Forecast with Time Series Models. *Exponential Technologies and Global Challenges: Moving toward a new culture of entrepreneurship and innovation for sustainable development*. <https://laccei.org/LEIRD2022-VirtualEdition/full-papers/FP6.pdf>
- Cryer, J. D., & Chan, K.-S. (2008). *Time Series Analysis with Applications in R* (2nd). Springer.
- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association*, 74(366), 427-431. <https://doi.org/10.2307/2286348>
- Donoso, M. J. (2013). El mercado del cobre a nivel mundial: evolución, riesgos, características y potencialidades futuras. *Ingeniare: Revista chilena de ingeniería*, 21(2), 248-261.

- Fierro Torres, C. Á., Castillo Pérez, V. H., & Torres Saucedo, C. I. (2022). Análisis comparativo de modelos tradicionales y modernos para pronóstico de la demanda: enfoques y características. *RIDE. Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 12(24), e048. <https://doi.org/10.23913/ride.v12i24.1203>
- Gonzales, L. F. (2019). Desafíos Ambientales y Sociales en la Minería Peruana. *Revista Ambiental Peruana*, 12(1), 30-48.
- González Casimiro, M. P. (2009). Análisis de series temporales: Modelos ARIMA. *Economía Aplicada III/Ekonomia Aplikatua III*, UPV/EHU.
- Grimmett, G. R., & Stirzaker, D. R. (2001). *Probability and Random Processes* (3rd). Oxford University Press.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.
- Hanke, J. E., & Reitsch, A. G. (1996). *Pronósticos en los negocios* (5ta). Prentice Hall Hispanoamericano S.A.
- Henríquez, M. I. I., Coello, M. E. S., & Sánchez, C. D. T. (2022). Modelo de pronóstico de la exportación minera en el Perú, 2020. *Revista de investigación estadística (ISSN: 2708-1125)*, 4(1).
- Hernández, R. S., Fernández, C. C., & Baptista, M. d. P. L. (2018). *Metodología de la investigación: la ruta cuantitativa, cualitativa y mixta*. Mc Graw Hill Education.
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice* [Disponible en <https://otexts.com/fpp3/>]. OTexts.
- Instituto de Ingenieros de Minas del Perú. (2024). Producción Peruana de cobre rompe récord histórico: Moquegua, Arequipa y Áncash son las regiones líderes [Recuperado el 10 de Julio del 2024]. <https://iimp.org.pe/actualidad-minera/produccion-peruana-de-cobre-rompe-record-historico-moquegua-arequipa-y-ancash-son-las-regiones-lideres>
- Instituto Nacional de Estadística e Informática. (2021). Estadísticas de la Producción Minera en el Perú [Recuperado el 08 de Agosto del 2025]. <https://www.inei.gob.pe>

- Instituto Peruano de Economía. (2024). Hablemos sobre el aporte de la minería en la economía [Recuperado el 10 de Junio del 2024]. <https://www.ipe.org.pe/portal/hablemos-aporte-de-la-mineria-en-la-economia-en-el-peru/#:~:text=La%20miner%C3%ADa%20tiene%20un%20rol,la%20reducci%C3%B3n%20de%20la%20pobreza>.
- International Copper Association. (2024). Ciclo de Vida del Cobre [Recuperado el 03 de Junio del 2024]. <https://internationalcopper.org/es/sustainable-copper/about-copper/copper-life-cycle/>
- Mamani Quispe, L. A. (2019). Exportación de cobre y su impacto en el crecimiento económico del Perú en el periodo 2010 – 2018.
- Mauricio, J. A. (2007). Análisis de series temporales. *Universidad Complutense de Madrid*.
- Ministerio de Desarrollo Agrario y Riego. (2024). Producción minera [Recuperado el 03 de Junio del 2024]. <https://www.midagri.gob.pe/portal/44-sector-agrario/recurso-mineral/336-produccion-minera>
- Ministerio de energía y minas. (2024). MINEM: Perú conservará posicionamiento atractivo para la actividad minera en los próximos años [Recuperado el 27 de Agosto del 2024]. <https://www.gob.pe/institucion/minem/noticias/674559-minem-peru-conservara-posicionamiento-atractivo-para-la-actividad-minera-en-los-proximos-anos>
- Ministerio de Energía y Minas del Perú. (2022). Informe Anual de la Producción Minera en el Perú. <https://www.gob.pe/minem>
- Ministerio de Energía y Minas del Perú. (2023). Reporte Anual de Producción Minera [Recuperado el 08 de Agosto del 2025]. <https://www.gob.pe/minem>
- Montalvan, M. M. E., & Silva, V. G. A. (2020). Factores económicos del cobre que influyen en el crecimiento económico del Perú. Periodo 2010-2019. *Universidad San Ignacio de Loyola*.
- Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2008). *Introduction to Time Series Analysis and Forecasting*. Wiley.
- Montgomery, D. C., & Runger, G. C. (2012). *Estatística Aplicada e Probabilidade para Engenheiros* (5.^a ed.). LTC.

- Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the Theory of Statistics* (4.^a ed.). McGraw-Hill.
- Morettin, P. A., & Toloi, C. M. C. (2006). *Análise de Séries Temporais* (2nd). Egard Blucher.
- Mutele, L., & Carranza, E. J. M. (2024). Statistical analysis of gold production in South Africa using ARIMA, VAR and ARNN modelling techniques: Extrapolating future gold production, Resources–Reserves depletion, and Implication on South Africa’s gold exploration. *Resources Policy*, 93, 105076. <https://doi.org/https://doi.org/10.1016/j.resourpol.2024.105076>
- Oktaviani, F., Miftahuddin & Setiawan, I. (2021). Forecasting sea surface temperature anomalies using the SARIMA ARCH/GARCH model. *Journal of Physics: Conference Series*, 1882(1), 012020. <https://doi.org/10.1088/1742-6596/1882/1/012020>
- Organización para la Cooperación y el Desarrollo Económicos. (2018). Minería y Desarrollo Sostenible en América Latina [Recuperado el 08 de Agosto del 2025]. <https://www.oecd.org>
- Orihuela, E., Soto, A. J., Jiménez, H. B., & Huayhuas, L. A. (2023). Un modelo vectorial autorregresivo (VAR) aplicado a las exportaciones de materias primas (oro y cobre) y la balanza de pagos: el caso peruano (2012–2022). *Compendium: Cuadernos de Economía y Administración*, 10(1), 63-75. <https://doi.org/10.46677/compendium.v10i1.1167>
- Real Academia Española. (2024). *Tiempo* [Recuperado el 02 de Septiembre del 2024]. <https://dle.rae.es/tiempo>
- Ríos, C. H. (2008). Series de tiempo. *Santiago: Universidad de Chile*.
- Rosales, R., Perdomo, J., Morales, C., & Urrego, J. (2012). Intermediate economics: Theory and applications. *Universidad de los Andes*.
- Sánchez, H. C., Reyes, C. R., & Sáenz, M. K. (2018). *Manual de término en invetiación científica, tecnología y humanítica*. Univeridad Ricardo Palma.
- Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 461-464.
- Shumway, R. H., & Stoffer, D. S. (2016). *Time Series Analysis and Its Applications With R Examples*. Springer.

- Suárez, J. M. (2020). Análisis del Crecimiento Minero en el Perú: 2000-2019. *Revista Peruana de Economía*, 15(2), 45-62.
- Teixeira-Gandra, C. F. A., Damé, R. d. C. F., Simonete, M. A., Bacelar, L. C. S., Disconzi, P. B., & dos Santos, J. P. (2014). Modelagem estocástica: previsão das temperaturas para a localidade de Pelotas/RS/Brasil. *Revista Agro@ mbiente On-line*, 8(1), 18-28.
- U.S. Geological Survey. (2023). Mineral Commodity Summaries 2023 [Recuperado el 08 de Agosto del 2025]. <https://pubs.usgs.gov/periodicals/mcs2023/mcs2023.pdf>

ANEXOS

Anexo A: Matriz de consistencia

Formulación del problema	Objetivos	Variables	Metodología de investigación
<p>1.- Problema General: ¿Cuál es el modelo de series de tiempo que mejor se ajusta para pronosticar la producción mensual de cobre en Perú con datos históricos (1999-2023) y validar con datos de Enero a Agosto del 2024?</p> <p>2.- Problemas Específicos: 2.1.- ¿Cuáles son las características presentes en la serie de producción mensual de cobre en Perú (1999-2023)? 2.2.- ¿Cuales son los modelos de series temporales que tienen mejor ajuste para la serie de producción mensual de cobre en Perú desde 1999 a Diciembre del 2023? 2.3.- ¿Cuál es el modelo que tiene mejor capacidad de validación de la producción mensual de cobre para los meses del ultimo año (Enero-Agosto 2024)?</p>	<p>1.- Objetivo General Determinar el modelo de series de tiempo que mejor se ajusta para pronosticar la producción mensual de cobre en Perú con datos históricos (1999-2023) y validación con datos de Enero y Agosto de 2024.</p> <p>2.- Objetivos Específicos: 2.1. Identificar las características presentes en la serie de producción mensual de cobre en Perú (1999-2023). 2.2. Analizar los modelos de series temporales que ofrecen el mejor ajuste para la serie de producción mensual de cobre en Perú desde 1999 a Diciembre del 2023. 2.3. Determinar el modelo con la mejor capacidad de validación de la producción mensual de cobre para los meses del último año (Enero-Agosto 2024).</p>	<p>Producción mensual de cobre (toneladas métricas finas).</p>	<p>Tipo de Investigación: El tipo de investigación es aplicada y cuantitativa.</p> <p>Alcance de la investigación: El estudio tiene alcance descriptivo, correlacional y predictivo.</p> <p>Diseño de investigación: El presente estudio es no experimental, longitudinal retrospectivo.</p> <p>Población: La producción total mensual de cobre de las mineras del Perú desde Enero de 1999 a Agosto de 2024, constituido por 308 observaciones según registro del MINEM, que contiene la data completa con relación a la serie de interés.</p>

Anexo B: Solicitud de la data solicitada al Ministerio de energía y minas



ALEXANDRA CONDORI CHOCCE <144961@unsaac.edu.pe>

Solicitud de Acceso a la Información Pública

frai@minem.gob.pe <frai@minem.gob.pe> 13 de mayo de 2024, 21:37
Para: 144961@unsaac.edu.pe

Estimada / Estimado: ALEXANDRA CONDORI CHOCCE

Se ha enviado una solicitud de Acceso a la Información Pública con los siguientes datos:

DATOS DEL SOLICITANTE	
D.N.I.	[REDACTED]
Nombres	ALEXANDRA
Apellidos	CONDORI CHOCCE
Teléfono	[REDACTED]
Correo	144961@UNSAAC.EDU.PE
Dirección	[REDACTED]
DATOS DE LA SOLICITUD	
Nº Expediente	37468 [REDACTED]
Fecha de Solicitud	13/05/2024 21:37:02
Situación Solicitud	PENDIENTE
Forma Entrega	CORREO ELECTRÓNICO
Información Solicitada	SOLICITO BASE DE DATOS DE LA PRODUCCION MINERA METALICA Y NO METALICA POR MESES DESDE ENERO DE 1999 A LA FECHA. TAMBIEN DE LA BASE DE DATOS DE LOS PRINCIPALES INDICADORES MACROECONOMICOS POR MESES DESDE ENERO DE 1999 A LA FECHA. REQUIERO DE ESTA BASE DE DATOS PARA PODER REALIZAR UN TRABAJO DE INVESTIGACION DE MI UNIVERSIDAD.

Atentamente.
Ministerio de Energía y Minas

Anexo C: Base de datos de la producción total mensual de cobre en Perú (1999-2024)

Para una visualización de la base de datos utilizada en el presente estudio y recopilada del Ministerio de Energía y Minas (MINEN) puede dirigirse al siguiente enlace: (<https://github.com/alexacondori/PRODUCCIONCOBRE>)