

Universidad Nacional del Comahue

MAESTRÍA EN ESTADÍSTICA APLICADA

UNA METODOLOGÍA DE PRONÓSTICO DE
PRECIOS APLICADA AL MERCADO DEL
CORDERO PATAGÓNICO EN CONTEXTO DE
APRENDIZAJE ESTADÍSTICO

Trabajo de tesis presentado por:

Ana K. Haique

Licenciada en Economía

Egresada de la Universidad Nacional de Córdoba, 1999

Neuquén, 2024

Comisión Asesora de Tesis

Director

Gustavo Gimenez
Ingeniero Agrónomo
Magister Scientiae
Doctor en Ciencias Agrarias

Co-Director

Natalia Rubio
Profesora en Matemáticas
Magister en Estadística Aplicada

Co-Director

Maria L. Malaspina
Ingeniera Agrónoma
Magister en Agronegocios

Jurado de Tesis

Mg. Laura Lamfre
Dr. Pablo Martín Demetrio
Dr. Eduardo G. Aisen

Resumen

El trabajo de esta tesis aborda la creciente necesidad de contar con pronósticos confiables para los precios del sector primarios, como es el caso del precio del cordero patagónico en Argentina. El objetivo principal fue presentar una metodología dinámica de pronóstico que permita seleccionar entre un conjunto de algoritmos candidatos y realizar un pronóstico de corto plazo, preciso. Integrando marcos estadísticos y econométricos con técnicas avanzadas de aprendizaje estadístico. Se partió de un análisis exhaustivo de la serie histórica de precios proporcionada por el Instituto Nacional de Tecnología Agropecuaria (INTA). Se identificaron patrones estacionales y estructurales que fundamentan la aplicación de modelos como: Autoregresivo Integrado de Medias Móviles (ARIMA), Suavizamiento Exponencial en el Tiempo (ETS) y PROPHET, este último desarrollado por Facebook para manejar las peculiaridades de las series temporales del mundo real, como tendencias no lineales, estacionalidades anuales, semanales y diarias, y feriados o eventos especiales. La validación cruzada adaptada para series temporales fue esencial para evaluar la eficacia predictiva, revelando la importancia del reentrenamiento periódico frente a la variabilidad de los datos, a medida que transcurre el tiempo. Los resultados destacan la superioridad condicional del modelo ARIMA en el contexto

actual, y la metodología propuesta establece un enfoque adaptable que garantiza pronósticos actualizados y alineados con el mercado. En base a estos resultados se procedió a realizar un pronóstico de 12 meses para la variable de interés. Finalmente, se identificó futuras líneas de investigación que incluyen la exploración de nuevas técnicas de validación, algoritmos y la combinación de modelos para optimizar el pronóstico de precios como el del cordero patagónico.

Dedicatoria

A mi madre, cuya sabiduría y visión han sido una brújula en mi vida.

Reconocimientos

Mi profundo agradecimiento a mis Directores de Tesis, Mg Natalia Rubio y Mg. Maria Laura Malaspina y en especial al Dr. Gustavo Giménez, cuya disposición para guiar han sido fundamentales en este viaje. Su generosidad para compartir su tiempo y conocimiento sin restricciones, ha sido no solo una invaluable fuente de aprendizaje, sino también una inspiración constante. Gracias por su dirección y apoyo incondicional.

Certificación

Director: Gustavo Gimenez

Codirectora: Natalia Rubio

Codirectora: María Laura Malaspina

CERTIFICAN:

Que Ana Karina Haique ha realizado bajo nuestra dirección la tesis “UNA METODOLOGÍA DE PRONÓSTICO DE PRECIOS APLICADA AL MERCADO DEL CORDERO PATAGÓNICO EN CONTEXTO DE APRENDIZAJE ESTADISTICO”, la cual se encuentra finalizada y en condiciones de ser presentada. Y para que así conste a todos los efectos del interesado se expide el presente certificado en Neuquén a los 15 días del mes de abril del año 2.024.



Gustavo Gimenez



Natalia Rubio



Ma. Laura Malaspina

Índice general

1. Introducción	17
1.1. Descripción del problema	18
1.2. Antecedentes	21
1.2.1. El pronóstico de precios de bienes primarios	22
1.2.2. Otros estudios	26
1.3. Objetivos	30
1.4. Implicancia de la investigación realizada	31
1.4.1. Aportes teóricos	31
1.4.2. Aportes metodológicos	31
1.4.3. Aportes prácticos	32
2. Revisión de la Literatura	33
2.1. En relación al área de aplicación	33
2.2. En relación a los métodos estadísticos abordados	38
2.2.1. Series de tiempo	38
2.2.2. El aprendizaje estadístico y las series de tiempo	40
3. Metodología	59
3.1. Metodología	59
3.1.1. Recopilación y configuración inicial de datos	60

3.1.2. Análisis exploratorio de datos:	63
3.1.3. Preparación de los datos	65
3.1.4. Selección y ajuste de los modelos:	66
3.1.5. Validación cruzada en series temporales:	69
3.1.6. Generación de pronósticos y evaluación	70
3.1.7. Esquema del Flujo de Trabajo o <i>Workflow</i>	71
4. Resultados	75
4.1. Análisis Exploratorio de los Datos	75
4.2. La selección de los modelos	81
4.2.1. Ajuste del Modelo de Suavizamiento Exponencial	82
4.2.2. Ajuste del Modelo ARIMA	83
4.2.3. Ajuste del Modelo PROPHET	85
4.2.4. Validación cruzada	87
4.2.5. Pronóstico	90
4.3. Resultados en relación al Objetivo General	92
5. Discusión de Resultados	95
6. Conclusiones	101
6.1. Sugerencias para Investigación Futura	102
A. Apéndice	105
A.1. Código de R	105
A.1.1. Librerías	105
A.1.2. Datos	105
A.1.3. EDA	106
A.1.4. Preparación de Datos	107
A.1.5. Splits: Entrenamiento vs Validación	108
A.1.6. ETS	108
A.1.7. ARIMA	110
A.1.8. PROPHET	111
A.1.9. Resample	112
A.1.10. Validación Cruzada	113

ÍNDICE GENERAL

A.1.11.Refit	114
A.1.12.Pronóstico en la escala original	114
Referencias	116
Siglas	121

Índice de figuras

1.1. Línea de tiempo del ciclo productivo del Cordero Patagónico	20
2.1. Distribución de Cabezas de Cordero por Provincia. Datos obtenidos de SENASA (2022)	34
2.2. Distribución de Unidades Productivas de Cordero por Provincia. Datos obtenidos de SENASA (2022)	35
2.3. Ratio Cabezas/Unidades Productivas de Cordero por Provincia. Datos obtenidos de SENASA (2022)	36
3.1. Flujo de Trabajo para la Metodología de Pronóstico de Precio aplicada al Cordero Patagónico	73
4.1. Precio Promedio Nominal Mensual del Cordero Patagónico desde Abril 2014 hasta Noviembre 2023	76
4.2. Logaritmo del Precio Promedio Nominal Mensual del Cordero Patagónico desde Abril 2014 hasta Noviembre 2023	76
4.3. Estacionalidad Mensual del Logaritmo del Precio Promedio del Cordero Patagónico	77
4.4. Detección de Anomalías en el Logaritmo del Precio Promedio del Cordero Patagónico	78

4.5. Descomposición STL del Logaritmo del Precio Promedio del Cordero Patagónico	79
4.6. Correlograma de la Función de Autocorrelación y Autocorrelación Parcial del Logaritmo del Precio Promedio del Cordero Patagónico	81
4.7. Partición de la serie en grupo de entrenamiento y grupo de testeo	82
4.8. Comparación de Pronósticos de Modelos ETS para el Logaritmo del Precio Promedio del Cordero Patagónico	83
4.9. Comparación de Pronósticos del Modelo ARIMA con Deriva y Modelo SARIMA	84
4.10. Comparación de Pronósticos del Modelo PROPHET y PROPHET con Ajustes	86
4.11. Plan de Validación Cruzada en Serie Temporal para los Modelos de Pronóstico del logaritmo del Precio Promedio del Cordero Patagónico	87
4.12. Resultados de las Métricas de Rendimiento en la Validación Cruzada en Serie Temporal del Log_PPCCord	88
4.13. Pronóstico de 12 Meses mediante el Modelo ARIMA(1,1,0)(1,1,0)[12], Reajustado con Datos Completos	90
4.14. PPCord y su Pronóstico de 12 Meses mediante el Modelo ARIMA(1,1,0)(1,1,0)[12], Reajustado con Datos Completos, en pesos.	92

CAPÍTULO 1

Introducción

La práctica de pronosticar ha sido constante en la historia humana, motivada por el anhelo intrínseco de comprender y anticipar el futuro. Los pronósticos desempeñan un papel central en el proceso de elección de las estrategias y toma de decisiones, especialmente en el ámbito económico y financiero (Peralta y Darwin, 2019).

En el caso de los productores de bienes primarios, como se aborda en este trabajo de tesis, las expectativas sobre los precios de los productos tienen un fuerte impacto en sus decisiones (Ruiz Hernández, Barrios Puente, y Gómez Gómez, 2019). Estas expectativas afectan sus ingresos, márgenes de ganancia, así como la inversión y el empleo en el sector (Peralta y Darwin, 2019). Por lo tanto, una planificación racional implica disponer de pronósticos confiables de esta variable clave.

En los últimos tiempos, las herramientas de pronóstico han experimentado una evolución significativa gracias a los aportes provenientes de las ciencias estadísticas, econométricas e informáticas. Existen diversos enfoques que se esfuerzan por alcanzar este objetivo, como el estudio de los factores que determinan la formación del precio, o los modelos auto-

regresivos propuestos por la metodología de Box y Jenkins, que permiten que "los datos hablen por sí mismos" (Gujarati y Porter, 2011), por mencionar solo algunos.

Estas metodologías, giran entorno a distintos modelos que pueden resultar más o menos pertinentes dependiendo de cada escenario. La evolución de la tecnología informática ha hecho posible la aplicación de diversos modelos, lo que a su vez requiere de herramientas específicas para identificar aquellos que ofrecen pronósticos más precisos. En este trabajo se pretende plantear una metodología de pronóstico, que facilite valorar varios modelos, en contexto de **Aprendizaje Estadístico (AE)**. La premisa básica del **AE** consiste en aprender a partir de los datos, es decir, el uso de las observaciones para descubrir un proceso subyacente (Hastie, Tibshirani, y Friedman, 2009). En el caso de los datos temporales, como los de este estudio, el propósito del **AE** es capturar la forma en que se comporta la variable de interés, aprender a diferenciar entre fluctuaciones aleatorias de los datos en el pasado, que deben ignorarse, como podrían ser consecuencias de sequías o erupciones volcánicas que afectaron al sector, versus un patrón genuino que debe modelarse y extrapolarse (Hyndman y Athanasopoulos, 2021).

1.1. Descripción del problema

El pronóstico del precio de un producto juega un rol crucial en la mayoría de industrias y sectores, ya que permite a las empresas tomar decisiones informadas sobre estrategias de precios, gestión de inventario, planificación de la cadena de suministro y optimización general del negocio (Murpratiwi, 2021).

En este trabajo se ha seleccionado como variable de estudio el precio nominal promedio del cordero patagónico, *PPCord*. Específicamente el valor relevado se corresponde con el kilo/gancho, en el momento de la faena, que es percibido directamente por los criadores. Este precio ha sido relevado de manera mensual a lo largo de aproximadamente una década por

1.1. DESCRIPCIÓN DEL PROBLEMA

el Centro Regional Patagonia Norte y el Centro Regional Patagonia Sur del **Instituto Nacional de Tecnología Agropecuaria (INTA)**. La serie elegida, abarca el período comprendido desde abril del 2014 a noviembre del 2023.

La viabilidad de los emprendimientos pecuarios depende, en gran medida, de la relación entre el precio del producto y el precio de los insumos. Es por ello que un pronóstico del precio del cordero patagónico, permitiría conocer con mayor anticipación, la rentabilidad anual que se puede esperar en la temporada próxima siguiente.

La actividad económica de la cría del cordero cuyos precios son objeto de pronóstico, tiene como escenario a la región patagónica. Esta es conocida por ser un ambiente al que el cordero se encuentra bien adaptado. El clima de la Patagonia Argentina es de escasas lluvias, ventoso, y suelo menesteroso de pasturas, condiciones que no son aptas para cualquier ganado pero sí para el ovino. El cordero patagónico es un animal que camina mucho y se desplaza grandes distancias, por eso su carne es mas bien magra y de buen sabor (**MAGyP, 2021**). El cordero patagónico se cría a campo, con pasturas naturales que no tiene ningún tratamiento de residuos, plaguicidas ni ningún elemento que pueda alterar el gusto de la carne. Lo que lo posiciona como un producto muy reconocido por Unión Europea, Japón y China (**Ruiz, 2019**).

En 2014, se le otorgó la indicación geográfica de *cordero patagónico* a todos los corderos criados al sur del río Colorado, mediante la Resolución 326/2014 (**Res 326/2014 de la SAGyP, 2014**). Esta designación se debe a las condiciones únicas de la Patagonia, que permiten ofrecer un producto diferenciado.

El ciclo reproductivo del cordero es estacional. Desde marzo hasta mayo, se lleva a cabo el servicio de apareamiento. La gestación tiene una duración aproximada de 145 a 153 días, es decir, alrededor de cinco meses. Las ovejas comienzan a parir a partir de agosto, aunque la fecha varía según la ubicación de los campos. Durante poco más de un mes, los corderos se alimentan exclusivamente de leche materna. A partir de octubre, se inicia el destete, estos corderos se conocen como “primicia” y suelen tener un mayor precio debido a que son los primeros en salir al mercado.

El destete se completa en diciembre, justo a tiempo para las festividades navideñas y de fin de año. A partir de febrero, la oferta de cordero comienza a disminuir (Iglesias y Ghezan, 2010). La Figura 1.1 ilustra de manera gráfica este ciclo reproductivo junto con los puntos clave en los que los productores deben tomar decisiones estratégicas. Se destacan los principales *drivers* de costo, incluyendo la alimentación suplementaria, el manejo sanitario, y las tácticas de comercialización. Estos factores son cruciales para la optimización de ingresos y eficiencia operativa a lo largo del año. La proyección de precios en el mercado actúa como guía para estas decisiones estratégicas, convirtiéndose así en un insumo esencial para la planificación productiva.

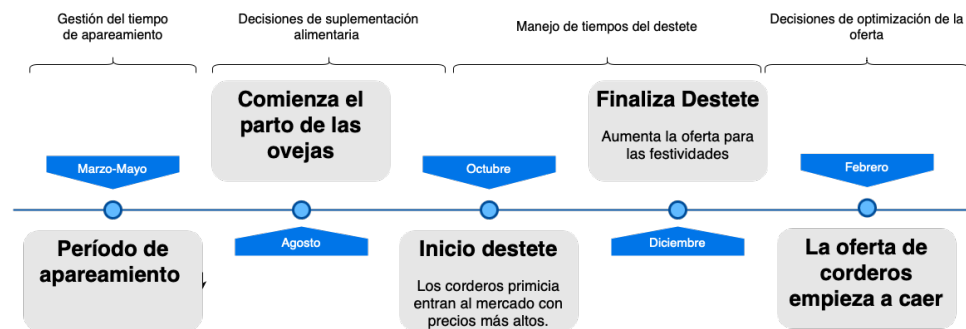


Figura 1.1: Línea de tiempo del ciclo productivo del Cordero Patagónico

Además de los criadores, el sector público también se beneficia de la información de los precios. Le sirve para diseñar incentivos, asistencias y planes dirigidos a cada sector productivo. La intervención del sector público puede garantizar la libre competencia en el mercado y la producción, así como evitar fluctuaciones de precios bruscas y pronunciadas o fomentar una actividad productiva en particular (Maino y Martínez, 1980) y para ello disponer de las proyecciones de los precios del sector en cuestión, es crucial.

Existen diversos métodos cuantitativos para realizar pronósticos, muchos de los cuales han sido desarrollados en disciplinas específicas con propósitos particulares (Hyndman y Athanasopoulos, 2021). Por ejemplo, modelos de simulación numérica para pronosticar el clima, métodos de re-

des de Petri en el ámbito de la logística, análisis de supervivencia que se utiliza para predecir la duración de la vida útil de un producto o fenómeno y métodos de series temporales desarrollados en el ámbito de la econometría para predecir indicadores económicos. Cada problema presenta sus propios desafíos, lo que implica que no existe un único modelo que permita realizar pronósticos eficientes en cualquier serie de tiempo. Es probable que, bajo un contexto de datos determinado, se puedan identificar varios modelos que sean candidatos viables para realizar el pronóstico. Por lo tanto, una vez que se haya ajustado un modelo, es necesario evaluar su rendimiento y precisión, y compararlo con otros posibles candidatos.

En el caso particular del mercado del cordero patagónico, no se han encontrado estudios de pronóstico específicos. Los análisis disponibles son de naturaleza descriptiva y proporcionan una representación de la serie histórica de precios en diferentes regiones, pero no incluyen proyecciones futuras.

A través de este trabajo, se busca desarrollar una metodología de pronóstico, teniendo como variable objetivo el *Precio Promedio del Cordero Patagónico (PPCord)* y ofrecer herramientas de evaluación que permitan determinar la eficacia de los diferentes modelos de pronóstico utilizados. La tesis se realizará bajo un enfoque auto-regresivo, es decir, permitiendo que los patrones pasados de la variable objetivo, predigan su trayectoria a futuro.

1.2. Antecedentes

El pronóstico de precios se refiere a la práctica de predecir los precios futuros basándose en datos históricos y tendencias pasadas (Gujarati y Porter, 2011), con el objetivo de tomar decisiones informadas sobre inversiones o planificación financiera. En el contexto del AE, existen diversas técnicas que se pueden emplear para mejorar la precisión en el pronóstico de precios (Hastie y cols., 2009). Al emplear modelos y algoritmos estadísticos avanzados, los agentes económicos tienen la oportunidad de

profundizar en su análisis y obtener una comprensión más completa de los patrones de precios. Esto les permite anticipar los movimientos del mercado con un mayor grado de precisión, lo que se traduce en una mayor rentabilidad y ventaja competitiva.

En este apartado se presentará un recorrido por varias investigaciones que abordan el reto de pronosticar variables temporales. Primeramente se presentaran antecedentes cuyo propósito es pronosticar precios de bienes primarios a fin de observar el tratamiento dado por sus autores; y luego se incluyen estudios que trabajan con algoritmos similares a los empleados aquí, o bien, que emplean herramientas de **AE**. Los trabajos seleccionados abordan el objetivo del pronóstico desde diferentes perspectivas por lo que se consideran antecedentes valiosos para esta investigación.

1.2.1. El pronóstico de precios de bienes primarios

En el campo del pronóstico de precios, una herramienta ampliamente utilizada es la familia de modelos **Autoregresivo Integrado de Medias Móviles (ARIMA)**. Tal es el tratamiento dado a los precios nominales pagados al productor de miel de abeja donde **Rojas, Sánchez, Mata, Coronado, y Valverde (2022)** trabajan con una serie histórica de precios aplicando la metodología de Box-Jenkins con y sin intervención. La opción sin intervención realiza el pronóstico teniendo en cuenta únicamente la serie del precio de la miel de abeja. La opción con intervención tiene en cuenta cambios estructurales y eventos significativos que pudieron haber afectado los precios de la miel, mediante la incorporación de variables dicotómicas. Por lo tanto este estudio propone que, sobre la base de estos modelos, los precios futuros de la miel pueden predecirse utilizando los precios del pasado. Este trabajo utilizó datos de 1966 a 2019 para entrenar o adaptar sus modelos ARIMA y luego se validó la capacidad de predicción de estos modelos utilizando datos más recientes, de 2018 y 2019. Este es un enfoque común para evaluar los modelos predictivos y para garantizar su capacidad de generalización a nuevos datos (**Hyndman y Athanasopoulos, 2021**). La bondad de ajuste de los modelos se evaluó empleando un análisis gráfico

de los residuos y medidas como el Error Estándar y el Error Porcentual Absoluto Medio o **Mean Absolute Percentage Error (MAPE)**. La predicción para el modelo sin intervención mostró un error porcentual absoluto medio (MAPE) de 8.16 %. Mientras que el modelo con intervención proporcionó un ajuste superior con un MAPE de 4.02 %, indicando una mayor precisión en las predicciones.

Otros investigadores de bienes primarios también han empleado modelos de la familia **ARIMA** como herramienta de pronóstico, por ejemplo, el “Análisis del precio de la manzana mediante un modelo SARIMA“ de **Ruiz Hernández y cols. (2019)**. SARIMA (**ARIMA** con estacionalidad), se consideró como el modelo más apropiado debido a su capacidad para manejar datos de series de tiempo que siguen un patrón estacional, como es el caso de la variable de interés de este trabajo. Las observaciones, 235 en total, fueron transformadas a logaritmos para evitar cambios pronunciados de la varianza a lo largo del tiempo, que pudieran ocasionar problemas en la aplicación del modelo. La bondad del ajuste se determinó mediante un análisis gráfico de los residuos. Este trabajo no emplea herramientas de **AE**, como reservar un porcentaje de datos para el testeo o la validación cruzada.

Haciendo foco en el sector pecuario, como puede ser el caso del mercado de la carne de cerdo, también se detecta el uso de los modelos **ARIMA** para el pronóstico de precios. **Zielińska-Sitkiewicz y Chrzanowska (2021)**, analizó las formas de predecir los precios de la carne de cerdo de clase E. Sus datos representaban los precios promedio mensuales pagados a los proveedores por animal al entregarlo al matadero, un rasgo que comparte con este trabajo, ya que se abordará los precios promedio mensuales del kilo gancho del cordero patagónico. Los datos cubren el período de mayo de 2004 a diciembre de 2019. El estudio utiliza dos métodos econométricos para predecir el precio de la carne de cerdo de clase E: Modelo de tendencia deslizante con segmentos de tendencias lineales constantes y un **ARIMA**. Los datos del período mencionado fueron utilizados como el conjunto de entrenamiento para desarrollar los modelos, y se reservaron las observaciones de los primeros tres meses de 2020 como conjunto de

prueba. La bondad de ajuste del modelo se estudió con las métricas tradicionales: error medio (**Mean Error (ME)**), error cuadrático medio (**Root Mean Squared Error (RMSE)**), error absoluto medio (**Mean Absolute Error (MAE)**), error porcentual medio (**Mean Percentage Error (MPE)**), error absoluto porcentual medio (**MAPE**) y el Coeficiente de Theil (Theil's U). El artículo concluye que tanto el modelo de tendencia deslizante como los modelos **ARIMA** son útiles para pronosticar los precios en el sector porcino. El primero es más ventajoso para predecir series caracterizadas por altas irregularidades o rupturas de tendencia. La versatilidad de los modelos **ARIMA**, por su parte, radica en su capacidad para adaptarse a estas dos situaciones: pueden manejar datos que requieren diferenciación para alcanzar estacionariedad y ser utilizados en datos que ya son estacionarios sin necesidad de diferenciación. Los autores concluyen que ambos muestran aceptabilidad y precisión de los pronósticos obtenidos.

En lo que respecta al mercado específico del cordero, los estudios analizados tienden a ser principalmente descriptivos. Un ejemplo de esto son los informes ocasionales emitidos por la *Agriculture and Horticulture Development Board* (AHDB), una importante asociación de agricultores en el Reino Unido. En estos documentos, la AHDB ofrece un desglose detallado de las tendencias históricas en el mercado de corderos, como se refleja en los informes de marzo de 2023 ([Hannah Clarke, 2023](#)) y abril del mismo año ([Tom Price, 2023](#)). Sin embargo, estos informes carecen de análisis predictivos y no proponen modelos de pronóstico que podrían anticipar las fluctuaciones futuras del mercado. La descripción proporcionada por la AHDB se centra en la documentación de datos pasados y actuales, sin adentrarse en la elaboración de predicciones basadas en dichos datos, lo cual podría ser de gran utilidad para los productores y comerciantes del sector ovino.

En relación al precio del cordero patagónico, el **INTA** publica un valioso relevamiento mensual¹ que caracteriza la situación de precios y oferta en diferentes regiones relacionadas con el mercado ovino. Sin embargo, no

¹Bolla, D. A., et al. (2023). Informe de precios de carne y ganado de la Patagonia. INTA.

cuenta con proyecciones de los precios a futuro.

Como se puede concluir, los modelos **ARIMA** son ampliamente reconocidos y empleados como una herramienta efectiva para el pronóstico de precios en el ámbito de los bienes primarios. Estos modelos son particularmente pertinentes cuando nos enfrentamos a series temporales que exhiben uniestacionalidad, es decir, un patrón estacional único que se repite a intervalos regulares, como es típico en los datos mensuales. En estas circunstancias, los modelos SARIMA, una extensión de **ARIMA** que incorpora componentes estacionales, resultan una alternativa adecuada para abordar y modelar dicha estacionalidad en la serie de tiempo.

Finalmente, en el ámbito de los productos pecuarios, la adopción de metodologías de pronóstico de precios está aún en una etapa incipiente. Una práctica recomendada en el desarrollo de modelos predictivos, como lo destaca **Hyndman y Athanasopoulos (2021)**, es la división del conjunto de datos en dos grupos distintos: uno para entrenar el algoritmo y otro para probar su rendimiento. Este enfoque es crucial ya que facilita la evaluación de la capacidad del modelo para generalizar y hacer predicciones precisas sobre datos que el modelo no ha visto previamente. No obstante, esta técnica no se ha adoptado ampliamente en el análisis de precios de los productos pecuarios, lo que representa una oportunidad significativa para mejorar la precisión y confiabilidad de los modelos de pronóstico en este sector.

En lo que respecta específicamente al mercado del cordero patagónico, se ha observado una notable ausencia de estudios que apliquen técnicas de pronóstico especializadas. Los informes y análisis que están disponibles tienden a ser de carácter descriptivo y adoptan una perspectiva principalmente económica. Estos estudios ofrecen una visión general de las tendencias de precios pasados, documentando y comentando sobre la serie histórica de precios. Sin embargo, carecen de una dimensión estadística avanzada que permita la creación de modelos capaces de proyectar hacia el futuro. Por consiguiente, se identifica un vacío en lo que respecta a la realización de pronósticos que serían de utilidad para los participantes del mercado. La capacidad de anticipar variaciones del mercado y fundamen-

tar decisiones en proyecciones futuras, además del conocimiento histórico, podría representar una ventaja significativa en la planificación y estrategia comercial.

1.2.2. Otros estudios

Como se vio en la sección anterior, los pronósticos de precios en general se realizan empleando modelos de la familia **ARIMA**. En este apartado analizamos algunos trabajos cuyo objetivo es brindar pronóstico. Las investigaciones se seleccionaron en función de los objetivos de esta tesis.

Se encontraron varios artículos cuya finalidad es comparar la efectividad de diferentes modelos a los fines del pronóstico. Tal es el caso del trabajo de tesis llevado a cabo por [Fernández Lavalle \(2019\)](#). Este trabajo evalúa la capacidad de pronóstico de diferentes modelos de series temporales propuestos para estimar la evolución de la serie trimestral de reservas internacionales argentinas en el período 1996-2018. Se comparan cuatro modelos: un modelo de *Random Walk*, un modelo univariado ARMA, un modelo multivariado **Vector Auto-Regressive (VAR)** y un modelo de corrección de errores **Vector Error Correction (VEC)**. El primero, se basa en la idea de que el valor en un momento dado depende únicamente del valor en el momento anterior más una perturbación aleatoria, no hay consideración de tendencias, patrones pasados ni influencias externas. El segundo pertenece a la familia **ARIMA**. El tercero, modelo **VAR**, es un enfoque estadístico utilizado para analizar y pronosticar el comportamiento conjunto de múltiples variables interrelacionadas en el tiempo y se utiliza para capturar las relaciones dinámicas entre variables que podrían influenciarse mutuamente. Y finalmente, el modelo **VEC** es una extensión de los modelos **VAR**. La parte central de un modelo **VEC** es el término de corrección de errores. Este término representa cómo las desviaciones de corto plazo de las variables del equilibrio a largo plazo, se ajustan en el siguiente período. Los resultados indican que el modelo univariado, ARMA, es el más efectivo para pronosticar los niveles de reservas internacionales cuando el horizonte temporal es un trimestre, y cuando se extiende a cuatro trimestres, el mo-

delo **VEC** muestra ganancias en términos de pronósticos. Esta conclusión es interesante ya que nos muestra que la capacidad de predicción puede cambiar en función del horizonte de tiempo. La performance de los modelos se ordenó en función de dos métricas: **RMSE** y **MAPE**, calculados fuera de la muestra.

El empleo de algoritmos alternativos al modelo **ARIMA**, como el modelo **PROPHET**, es más reciente. Tal es el caso del estudio para predecir los precios de las acciones de la bolsa de Myanmar (**Chan, 2020**). Este trabajo, analiza el rendimiento de estos dos modelos para diferentes unidades de tiempo (diario, semanal y mensual). El conjunto de datos cuenta con un total de 946 observaciones diarias. El artículo, describe un proceso que incluye la división de datos en conjuntos de entrenamiento y prueba, y la comparación de los precios pronosticados con los reales para evaluar la precisión de los modelos. Menciona que la selección del modelo debería hacerse utilizando validación cruzada, herramienta que se empleará en la metodología propuesta por esta investigación, para evitar el sobreajuste en los modelos predictivos (**Chan, 2020**). Sus conclusiones principales indican que, aunque ambos modelos son adecuados para predicciones a corto plazo, **PROPHET** superó a **ARIMA** para tres periodos de predicción, es decir, para un horizonte de tiempo mayor.

Otros autores que realizan un análisis comparativo de modelos son **Menculini y cols. (2021)**. Usando como modelo de referencia un **ARIMA**, trabajan con algoritmos como **PROPHET** y modelos de aprendizaje automático como **Long Short Term Memory (LSTM)**, tanto por sí solas como en combinación con **Convolutional Neural Network (CNN)**. Los modelos comparados son bastante diferentes, al igual que los conjuntos de datos que aceptan como entrada. **Prophet** mostró ventajas por su simplicidad y escalabilidad. Los modelos de redes neuronales empleados, se prestan directamente para llevar a cabo una regresión multivariante. Los investigadores se enfocaron en tres productos para realizar el análisis: *Carnaroli rice 1 kg × 10*, *Gorgonzola cheese 1/8 of wheel 1.5 kg* y *Cured aged ham 6.5 kg*. Su objetivo era predecir el precio promedio de venta para la semana siguiente de cada uno de estos productos. Para asegurar datos precisos,

excluyeron los registros posteriores al brote de la pandemia de COVID-19 y eliminaron los valores atípicos. Luego, volvieron a muestrear el conjunto de datos con una frecuencia semanal, acumulando el número de pedidos y calculando el precio promedio de venta para cada semana. Las observaciones disponibles para entrenamiento fueron 240, el set de validación fue de 52 y el set de testeo 62. El conjunto de validación, se utiliza durante la fase de desarrollo del modelo. Se usa para ajustar los hiperparámetros del modelo, seleccionar las mejores configuraciones y reducir el sobreajuste. El conjunto de testeo se reserva exclusivamente para evaluar el rendimiento generalizado del modelo después que este ha sido entrenado utilizando conjuntamente los sets de testeo y validación. En resumen, este trabajo resulta un valioso antecedente porque, por un lado, trabaja con dos de los algoritmos que emplearemos en esta investigación: **ARIMA** y **PROPHET** y por otra parte, emplea herramientas de aprendizaje estadístico e incluso trabaja con herramientas como ensamblaje de modelos, también conocido como combinación de modelos, que excede los objetivos de esta tesis.

El Cuadro 1.1, brinda un esquema de los antecedentes descritos en esta sección.

En conjunto, esta revisión ha explorado y comparado diversas metodologías de pronóstico de series de tiempo, ofreciendo distintos enfoques y herramientas que serán fundamentales para desarrollar una metodología de pronóstico adecuada para el precio del cordero patagónico.

1.2. ANTECEDENTES

Cuadro 1.1: Resumen de antecedentes revisados.

Antecedente	Objeto de Estudio	Metodología
Rojas (2022)	Precios de miel de abeja	ARIMA con y sin intervención
Ruiz (2019)	Precio de la manzana	SARIMA, transformación logarítmica
Zielinska (2021)	Precios de la carne de cerdo	ARIMA, Modelo de tendencia deslizando
AHDB (2023)	Mercado de corderos	Análisis descriptivo
INTA (2023)	Precio del cordero patagónico	Relevamiento mensual
Fernández	Reservas internacionales argentinas	Comparación de modelos (Random Walk, ARMA, VAR, VEC)
Chan (2020)	Precios de acciones	Comparación de ARIMA vs PROPHET
Menculini (2021)	Productos alimenticios	ARIMA, PROPHET, LSTM + CNN

1.3. Objetivos

Objetivo principal

Desarrollar una metodología de pronóstico adecuado aplicable al precio del cordero patagónico, que permita prever con anticipación su evolución en el mercado.

Objetivos específicos:

1. Realizar un análisis exhaustivo de la serie histórica de precios del cordero patagónico, recopilada por el INTA, identificando patrones estacionales, tendencias y posibles cambios estructurales.
2. Examinar y evaluar modelos de pronóstico, abarcando tanto los enfoques tradicionales arraigados en la econometría y la estadística, como los **ARIMA** y **Exponential Time Smoothing (ETS)**; así como modelos más contemporáneos vinculados al ámbito del aprendizaje automático, como el modelo PROPHET, y determinar su aplicabilidad al contexto del precio del cordero patagónico.
3. Aplicar una metodología de validación cruzada adecuada para series temporales para evaluar y comparar el rendimiento de los diferentes modelos de pronóstico en el precio del cordero patagónico.
4. Realizar un pronóstico para los 12 meses próximos siguientes.
5. Proporcionar recomendaciones y pautas para la selección y aplicación de modelos de pronóstico precios como el del cordero patagónico.

1.4. Implicancia de la investigación realizada

1.4.1. Aportes teóricos

Esta investigación aspira a enriquecer el campo del aprendizaje estadístico y el pronóstico de series de tiempo mediante un análisis comparativo de los modelos ETS, ARIMA y PROPHET, aplicados específicamente al caso del precio del cordero patagónico. Pretendemos evaluar la eficacia y aplicabilidad de estos modelos en el contexto mencionado, lo que constituye una indagación novedosa, especialmente en el mercado de carnes patagónicas. Se anticipa que este análisis ofrecerá *insights* valiosos para los productores de cordero patagónico, permitiéndoles anticipar con mayor precisión las fluctuaciones de precios, lo cual podría traducirse en una ventaja competitiva significativa y en una gestión del riesgo más eficiente.

Se espera que el estudio sugiera líneas de investigación futuras, como la evaluación del impacto de la longitud del horizonte de pronóstico en la precisión de los modelos y la exploración del ensamblaje de modelos (*ensembling*) para potenciar la precisión predictiva. Estas áreas representan oportunidades para ampliar el conocimiento y mejorar las prácticas de pronóstico en el ámbito de la estadística aplicada.

1.4.2. Aportes metodológicos

Con la intención de contribuir a un área de estudio emergente, este trabajo propone explorar el pronóstico del precio del cordero patagónico y desarrollar una metodología potencialmente extensible a otros contextos similares. Un aporte metodológico distintivo es la creación de un flujo de trabajo detallado en la Figura 3.1, utilizando R versión 4.2.1 (R Core Team, 2023), que sirva como guía para investigadores interesados en el pronóstico de precios. Este enfoque no solo fomentaría la replicabilidad y la transparencia científica sino que también establecería puntos de inicio para futuras investigaciones en este dominio.

En conjunto, este estudio no solo contribuiría al entendimiento específico del mercado del cordero patagónico sino que también aportaría sig-

nificativamente al diseño metodológico, al establecer un flujo de trabajo replicable y códigos accesibles, facilitando la adaptación de la metodología a investigaciones futuras y a otros mercados comparables.

1.4.3. Aportes prácticos

Este estudio tiene la intención de ofrecer nuevas herramientas de pronóstico para el precio del cordero patagónico, las cuales podrían ser de gran utilidad para instituciones como el INTA, entidades gubernamentales y organizaciones del sector pecuario.

Por otra parte, al proporcionar información precisa y oportuna sobre las tendencias y fluctuaciones esperadas en el precio del cordero patagónico, se empoderaría a los productores para tomar decisiones informadas que puedan tener un impacto significativo en la rentabilidad y el éxito de sus emprendimientos.

En última instancia, se busca que estas herramientas de pronóstico faciliten una mejor gestión y toma de decisiones en el sector, contribuyendo al crecimiento sostenible y la rentabilidad de la industria ovina en la región patagónica.

CAPÍTULO 2

Revisión de la Literatura

2.1. En relación al área de aplicación

En Argentina, la producción de carne ovina ha experimentado un notable crecimiento, en los últimos tiempos. Este aumento se debe principalmente a la inclusión de razas enfocadas en la producción de carne, el aumento en el número de criaderos dedicados a esta actividad, así como la reorientación de criaderos existentes hacia razas o genética especializadas en carne. También han influido cambios en los métodos de manejo y en la infraestructura, incluyendo la adaptación de salas de faena para esta actividad (MAGyP, 2021).

A nivel global, se observa un aumento en la demanda de carne, mientras que la cantidad de países con la capacidad para abastecerla de manera eficiente y sostenible se ha reducido significativamente. En este contexto, Argentina se posiciona como uno de los pocos países con la capacidad y potencial para abordar este desafío (Lauria y Uccelli, 2022).

El sector ovino en Argentina cuenta con dos importantes leyes: la Ley de Recuperación de la Ganadería Ovina N° 25.422 (Argentina, 2001) y la Ley

para la Promoción, Desarrollo y Consolidación de la Ganadería Ovina y de Llamas N° 27.646 (Argentina, 2021). Estas leyes, junto con el Programa para la Promoción del Consumo de Carne Ovina, reflejan el real interés de las políticas públicas en el fomento del sector (MAGyP, 2021). La producción ovina es una de las actividades agropecuaria más destacada en términos económicos y sociales en las provincias patagónicas, y su importancia radica en que no ha podido ser reemplazada por otras actividades debido a las características de los ecosistemas locales (Goizueta y Iglesias, 2013).

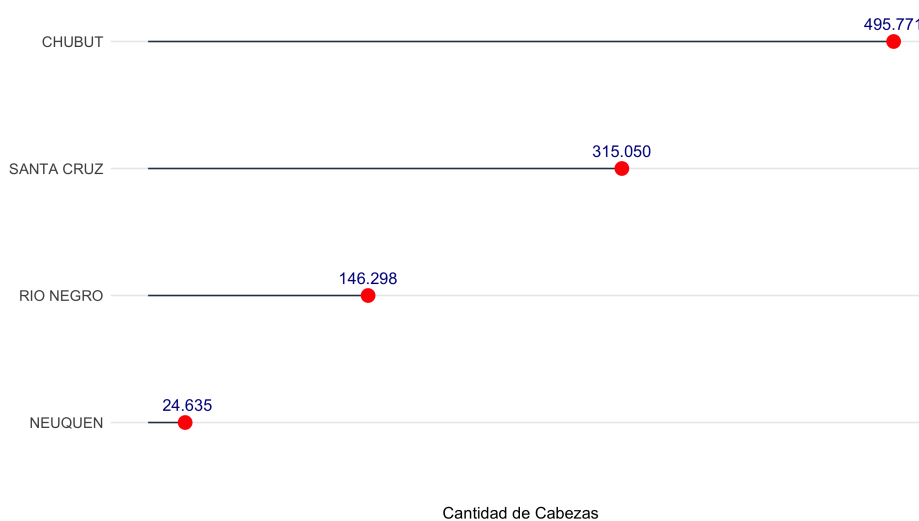


Figura 2.1: Distribución de Cabezas de Cordero por Provincia. Datos obtenidos de SENASA (2022)

Para caracterizar la oferta del cordero, podemos analizar la distribución de unidades de cabezas por provincia. Como se observa en la Figura 2.1, existe un importante concentración de la oferta en las provincias de Chubut y Santa Cruz. Chubut lidera la producción con 495,771 cabezas, seguido por Santa Cruz con 315,050 cabezas, Río Negro con 146,298 cabezas, y Neuquén con 24,635 cabezas (SENASA, 2022).

2.1. EN RELACIÓN AL ÁREA DE APLICACIÓN

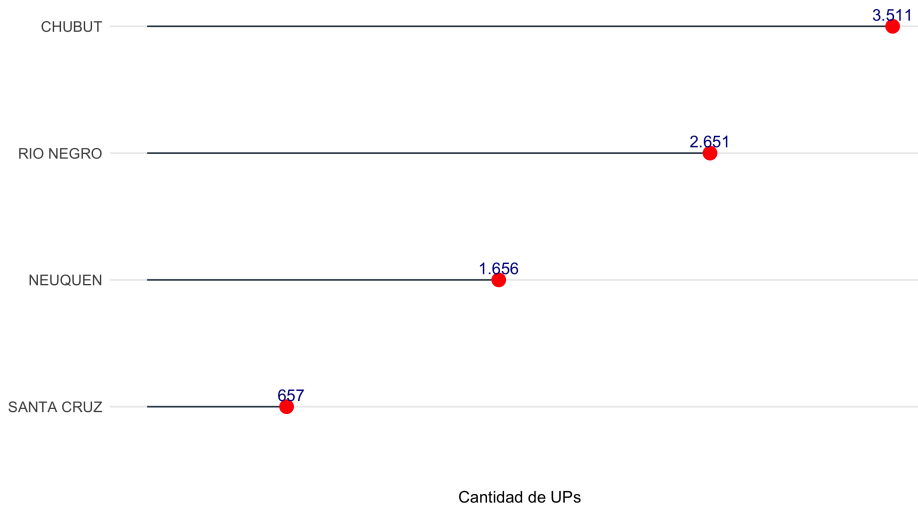


Figura 2.2: Distribución de Unidades Productivas de Cordero por Provincia. Datos obtenidos de SENASA (2022).

En términos de Unidades Productivas (UPs) por provincia, si bien Chubut también lidera esta categoría, Santa Cruz pasa al final de la lista, indicando que esta provincia presenta unidades productivas más grande en comparación con el resto de las provincias patagónicas (ver Figura 2.2).

La Figura 2.3, permite observar la relación de cabezas por unidad productiva (UP), lo que indica que tan intensivas son las unidades productivas en cada provincia. Santa Cruz presenta la mayor concentración con 480 cabezas por UP, seguida por Chubut con 141 cabezas por UP, Río Negro con 55 cabezas por UP, y Neuquén con 15 cabezas por UP. La variabilidad en la relación Cabezas por UPs, sugiere enfoques en la cría de cordero patagónico diferentes entre provincias.

En lo que respecta a la demanda, la información disponible es más fragmentada y menos clara esto se debe en parte, a que “estimaciones no oficiales calculan que la faena oficial representaría solo el 50 % de la real” (Lauria y Uccelli, 2022).

Como informan [Lauria y Uccelli \(2022\)](#), durante el año 2021 se faenaron 936.477 cabezas ovinas en todo el territorio del país donde 15 establecimientos tienen el 80 % del total, 21 el 15 % y los más chicos que son 45, el 5 % del total de la faena. El total faenado por las provincias en estudio es de 724.267 cabezas de ganado ovino lo que representa el 77 % de la faena nacional. Finalmente, el cordero equivale al 60 % del total de la faena.

Para el año 2022, aún no se dispone de datos con el mismo nivel de detalle; sin embargo, el [INDEC \(2022\)](#), en su anuario informa un volumen de faena de 975.661 cabezas, es decir, un 4,2 % más que el año anterior. Además, el volumen de exportación ascendió a 4.318 cabezas, siendo su principal destino el Medio Oriente.

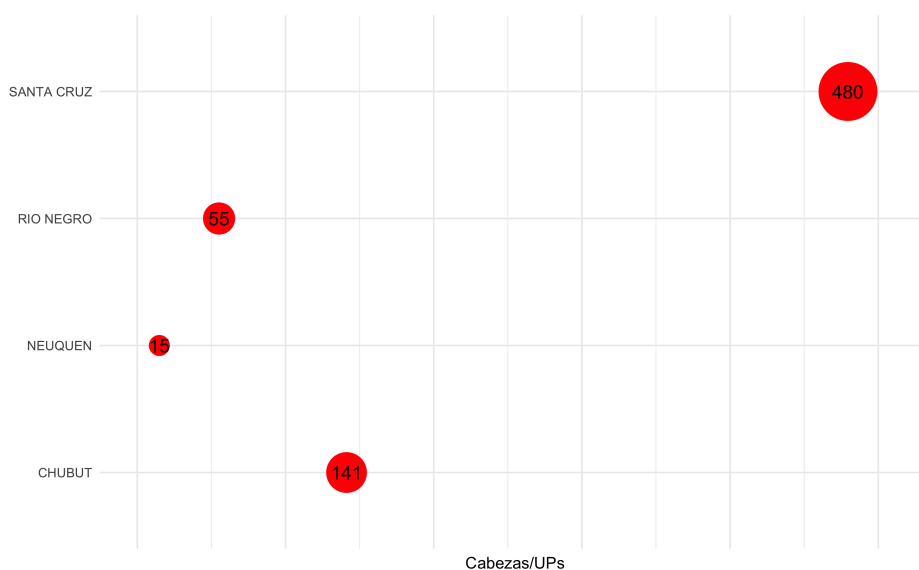


Figura 2.3: Ratio Cabezas/Unidades Productivas de Cordero por Provincia. Datos obtenidos de SENASA (2022).

El sector ovino, y en particular el cordero patagónico, presenta un gran potencial en Argentina, especialmente en un contexto con una creciente demanda de carne fresca en el mercado local y pocas alternativas disponibles para abordar esta situación ([Lauria y Uccelli, 2023](#)). Con el objetivo

2.1. EN RELACIÓN AL ÁREA DE APLICACIÓN

de promover la carne ovina argentina entre los consumidores nacionales y potenciar su cadena de valor, el Ministerio de Agricultura, Ganadería y Pesca llevó a cabo un detallado informe durante 2021. Este estudio exploró la presencia y la percepción de la carne ovina en diversos medios de comunicación y plataformas digitales, incluyendo medios gráficos digitales, audiovisuales, redes sociales de periodistas gastronómicos, cocineros e *influencers*, así como programas de televisión. A pesar de que la carne ovina, y en particular el cordero patagónico, no cuenta con una presencia amplia en estos medios, las menciones que se encuentran son mayormente positivas, asociándose con exclusividad, lujo, y momentos especiales, lo cual sugiere una imagen favorable que contribuye a su distinción como un producto gourmet y de alta calidad. Este aspecto emocional y de exclusividad, aunque actualmente limitado a un nicho de mercado, representa una oportunidad para los actores de la cadena de valor ovina. Los aspectos cuantitativos del estudio revelaron que “el lugar que ocupe la carne ovina (en el mercado) debe necesariamente estar acompañado por el factor precio. La tarea que queda por delante para los actores de la cadena es encontrar ese precio objetivo que lo distinga y lo haga competitivo, el cual no necesariamente debe ser más barato que otras carnes. El precio asociado a la calidad y la oferta serán las puertas de entrada para ganar terreno entre las carnes de elección”(MAGyP, 2021).

En síntesis, el incremento sostenido en la producción de carne ovina en Argentina puede atribuirse a diversos factores y se presenta como una buena oportunidad para el país en el marco de una demanda creciente de carne a nivel mundial. Estos hallazgos respaldan la relevancia y la necesidad de contar con herramientas precisas de pronóstico de precios, que no solo proporcionen estimaciones precisas sobre las tendencias futuras de precios, sino que también doten a los productores, con la capacidad de planificar con anticipación, ajustar sus estrategias de producción y comercialización de forma proactiva, y en definitiva, optimicen sus recursos para maximizar los beneficios.

Las políticas públicas y el interés gubernamental en el sector ovino, reforzados por leyes y programas de promoción, resaltan la importancia

de desarrollar métodos de pronóstico confiables y efectivos para potenciar la cadena de valor de la carne del cordero patagónico en el mercado. La promoción de este producto representa un desafío y una oportunidad para los actores de la cadena, y la aplicación de herramientas de aprendizaje estadístico, como las propuestas en este estudio, puede ser clave para sostener la competitividad en este mercado en constante evolución.

2.2. En relación a los métodos estadísticos abordados

En esta sección de la tesis, se presenta una revisión bibliográfica que sienta las bases teóricas y metodológicas de la investigación sobre el pronóstico del precio del cordero patagónico. El análisis se desarrolla en dos ejes principales que son fundamentales para encuadrar esta investigación: las series de tiempo y el aprendizaje estadístico.

2.2.1. Series de tiempo

El análisis de series de tiempo es una herramienta crucial en el campo de la economía, ya que permite estudiar la evolución de variables a lo largo del tiempo, identificar patrones y fluctuaciones históricas, y realizar predicciones sobre futuras condiciones económicas (Gujarati y Porter, 2011). En este contexto, una serie de tiempo se define como una secuencia de variables aleatorias indexadas en el tiempo, cuyas observaciones se miden en intervalos regulares (Wooldridge, 2010).

Hyndman y Athanasopoulos (2021), distinguen claramente tres métodos de abordaje para el estudio de series de tiempo. Por un lado, los “modelos explicativos” que buscan explicar qué causa la variación de la variable de respuesta. Por ejemplo, en el caso del precio de un bien agropecuario:

$$Precio_t = f(u\$s_t, Clima_t, OtrosPrecios_t...error) \quad (2.1)$$

2.2. EN RELACIÓN A LOS MÉTODOS ESTADÍSTICOS ABORDADOS

Esto implica que el *Precio* es explicado por un conjunto de regresoras, por ejemplo, el tipo de cambio (*u\$s*), el clima, otros precios, y un término de error.

También puede explicarse mediante un “modelo de series temporales”:

$$Precio_t = f(Precio_{t-1}, Precio_{t-2}, Precio_{t-3}...error) \quad (2.2)$$

En este caso, el *Precio* en el momento *t* depende de sus valores pasados en *t-1*, *t-2*, *t-3*, etc., y un término de error.

Y una tercer opción sería una combinación de ambas, conocida como “modelos de regresión dinámicos”:

$$Precio_t = f(Precio_{t-1}, u\$s_t, Clima_t, OtrosPrecios_t...error) \quad (2.3)$$

En este enfoque, el *Precio* en el momento *t* depende de sus valores pasados, así como de las variables predictoras: tipo de cambio (*u\$s*), clima, otros precios, y un término de error.

En sintonía con [Hyndman y Athanasopoulos \(2021\)](#), hay varias razones por las cuales en este trabajo se prefiere optar por un “modelo de series temporales”:

1. En la primera y tercera opción, planteadas más arriba, el sistema puede no estar completamente comprendido, una teoría que determine el comportamiento de la variable de respuesta, podría estar incompleta, y con frecuencia suele suceder. Y aunque lo estuviera, medir las relaciones planteadas, puede ser extremadamente difícil, por ejemplo, debido a la falta de disponibilidad de datos.
2. En la primera y tercera opción puede ser necesario prever los valores futuros de las propias predictoras para pronosticar la variable de interés.
3. Y si el objetivo es pronosticar, como en esta tesis, esto se podría lograr sin conocer necesariamente el “por qué” que involucra las relaciones entre variables.

Entonces, el análisis de modelos de series temporales se presenta como una herramienta eficaz y adaptable para el pronóstico. Al centrarse en las tendencias y patrones históricos inherentes a los datos, estos modelos ofrecen un enfoque pragmático que, si bien puede no desentrañar completamente las causas subyacentes de las fluctuaciones económicas, sí provee una base sólida para realizar predicciones confiables. Este enfoque práctico y centrado en el pronóstico es particularmente relevante en un entorno económico donde la rapidez y precisión de la información son cruciales para la formulación de estrategias y políticas efectivas.

2.2.2. El aprendizaje estadístico y las series de tiempo

El **AE** como herramienta para entender y modelar datos complejos, es un área relativamente nueva dentro de la estadística que se combina con desarrollos paralelos en el ámbito de la informática y el aprendizaje automático (James, Witten, Hastie, y Tibshirani, 2021). El aprendizaje estadístico aplicado a series de tiempo se enfoca en el estudio de métodos y modelos estadísticos que permiten identificar patrones y tendencias temporales en datos secuenciales. Su objetivo es desarrollar técnicas que permitan hacer inferencias, pronósticos y análisis de series de tiempo basados en los datos históricos disponibles. Los modelos de aprendizaje estadístico para series de tiempo se basan en principios estadísticos y probabilísticos, y sus objetivos pueden ser comprender y modelar las relaciones subyacentes entre las observaciones a lo largo del tiempo, identificar componentes como tendencias, estacionalidades y ruido, y realizar pronósticos precisos para el futuro (Shumway y Stoffer, 2017).

Los métodos del **AE** se pueden clasificar en cuatro áreas principales:

1. **Aprendizaje Supervisado:** En este enfoque, el modelo se entrena con un conjunto de datos etiquetados, lo que significa que cada entrada de datos (X_1, X_2, \dots, X_p) en el conjunto de entrenamiento está asociada con una respuesta o salida conocida (Y) (James y cols., 2021). El objetivo es aprender una función que, dada una entrada,

2.2. EN RELACIÓN A LOS MÉTODOS ESTADÍSTICOS ABORDADOS

puede predecir la salida deseada. Ejemplos típicos incluyen la regresión y la clasificación ([IFOA Data Science Research, 2023](#)).

2. **Aprendizaje No Supervisado:** Aquí, los modelos se entrenan utilizando un conjunto de datos que no está etiquetado. El objetivo es explorar la estructura subyacente de los datos para extraer patrones, detectar anomalías o agrupar datos en categorías homogéneas ([James y cols., 2021](#)). Un ejemplo común es el análisis de agrupamiento.
3. **Aprendizaje Semisupervisado:** El aprendizaje semisupervisado es una técnica de aprendizaje automático que utiliza tanto datos etiquetados (como en el aprendizaje supervisado) como no etiquetados (como en el aprendizaje no supervisado). En esta técnica, solo una parte de los datos en el conjunto de entrenamiento tiene etiquetas. Al utilizar datos no etiquetados junto con una cantidad menor de datos etiquetados, el aprendizaje semisupervisado puede mejorar la precisión de los modelos. Esto se debe a que el modelo puede aprender patrones tanto de los datos etiquetados como de los no etiquetados ([Hajaj, 2023](#)).
4. **Aprendizaje por Refuerzo o Aprendizaje Forzado:** Difiere significativamente de los otros tipos de aprendizaje automático. En lugar de ser entrenado con datos etiquetados o no etiquetados, en el aprendizaje por refuerzo, un agente aprende a tomar decisiones interactuando con un entorno. A través de este proceso, el agente recibe recompensas o penalizaciones basadas en las acciones que realiza, lo que le ayuda a aprender cuáles son las mejores acciones para maximizar la recompensa a lo largo del tiempo ([Hajaj, 2023](#)). Es ampliamente utilizado en áreas como juegos, navegación autónoma y sistemas de recomendación.

En esta tesis, cada punto de datos en la serie de tiempo actúa como un dato etiquetado. Esto significa que para cada momento en el tiempo, se tiene tanto la entrada (el momento en el tiempo) como la salida (el valor de

la serie en ese momento). Además, el objetivo es pronosticar los valores futuros de la serie de tiempo, basándose en los datos históricos. Esto se alinea con la esencia del aprendizaje supervisado, donde el modelo aprende de los ejemplos pasados para hacer predicciones hacia el futuro. Es por ello que la naturaleza de los datos de series de tiempo y el enfoque de pronóstico del análisis se alinean con los principios del aprendizaje supervisado.

Por otra parte, es importante resaltar que el pronóstico para series de tiempo es una necesidad en muchas situaciones económicas y de negocios (Niaki, Seifi, y Azizi, 2020), donde la anticipación y comprensión de las tendencias temporales conlleva un valor estratégico incuestionable. Los buenos pronósticos capturan los patrones y relaciones genuinos que existen en los datos históricos, pero no replican eventos pasados que no volverán a ocurrir. En otras palabras, un modelo de pronóstico busca capturar la forma en que se mueven las cosas, aprendiendo a diferenciar entre una fluctuación aleatoria en los datos pasados, que debe ignorarse, y un patrón genuino que debe ser modelado y extrapolado (Hyndman y Athanasopoulos, 2021).

En el ámbito del aprendizaje estadístico aplicado al pronóstico de precios, se identifican diversos modelos con potencial de aplicación. En base a los antecedentes analizados, se seleccionaron tres modelos para desarrollar la metodología propuesta, dirigida al pronóstico de precios del cordero patagónico: el modelo de suavizamiento exponencial, el modelo autorregresivo de medias móviles y el modelo PROPHET

El modelo de suavizamiento exponencial (ETS)

El suavizamiento exponencial es un modelo estadístico que implica aplicar ponderaciones a observaciones pasadas de una serie temporal, con ponderaciones exponencialmente menores a medida que se encuentran más alejadas en el pasado (Lugo y Porto, 2020). En su versión más completa, conocida como método de Holt-Winters, el modelo logra captar razonablemente tanto la tendencia como la estacionalidad. Una de las

2.2. EN RELACIÓN A LOS MÉTODOS ESTADÍSTICOS ABORDADOS

principales ventajas del **ETS** es su simplicidad y facilidad de uso. El método **ETS** permite actualizar y mejorar los pronósticos a medida que se disponen de nuevos datos, lo que es esencial en entornos en tiempo real. Como limitante, si la serie es no estacionaria, el método puede proporcionar resultados poco confiables y no siempre captura cambios extremos o abruptos en la serie de tiempo por lo que puede ser menos preciso si el horizonte de pronóstico es muy largo (Hyndman y Athanasopoulos, 2021). Considerando sus ventajas, se incluye como uno de los modelos para realizar el pronóstico del precio del cordero patagónico.

El modelo **ETS** presenta tres versiones diferenciadas:

1. **Suavizamiento exponencial simple:** Si denotamos Y_1, Y_2, \dots, Y_T a la serie de datos históricos, en este caso el pronóstico es un promedio ponderado que es más sensible a las observaciones recientes:

$$\hat{Y}_{T+1|T} = \alpha Y_T + \alpha(1 - \alpha)Y_{T-1} + \alpha(1 - \alpha)^2 Y_{T-2} + \dots + \alpha(1 - \alpha)^{T-1} Y_1 \quad (2.4)$$

En la ecuación 2.4, $\hat{Y}_{T+1|T}$ hace referencia al valor estimado de Y para el momento $T + 1$, dado los datos Y_1, Y_2, \dots, Y_T . El parámetro α es conocido como el factor de suavización y toma valores entre 0 y 1. Representa el peso que se le otorga a la observación más reciente en el cálculo del pronóstico. Un valor de α cercano a 1 implica que se le da mayor importancia a los datos más recientes, lo que hace que el pronóstico sea más sensible a las fluctuaciones recientes. Por otro lado, un valor de α cercano a 0 hará que el pronóstico sea más estable y menos sensible a las últimas observaciones (Hyndman y Athanasopoulos, 2021).

Otra forma de expresar la ecuación (2.4) es como un promedio ponderado entre la observación más recientes, Y_T y el pronóstico previo $\hat{Y}_{T|T-1}$:

$$\hat{Y}_{T+1|T} = \alpha Y_T + (1 - \alpha)\hat{Y}_{T-1} \quad (2.5)$$

Una representación alternativa es en su forma de componentes, por ahora solo se tiene un componente llamado de nivel o valor suavizado

y se denota ℓ_t . Por lo tanto, se tiene:

$$\hat{Y}_{t+h}|t = \ell_t \quad (2.6)$$

Si $h = 1$, (2.6) devuelve los valores ajustados o suavizados. Y si $t = T$ devuelve el pronóstico más allá de los datos de entrenamiento. Respecto al valor de α , si bien esta está comprendidos entre 0 y 1, es común que en la práctica varíe entre 0,05 y 0,5 (Guerrero Melgarejo, 2020).

2. **Suavizamiento exponencial con tendencia - Método de Holt:** El suavizamiento exponencial simple devuelve una función de pronóstico plana, es decir, todos los pronósticos toman el mismo valor. Para incluir un componente de tendencia, el método de Holt¹, propone como ecuación de pronóstico la ecuación 2.7:

$$\hat{Y}_{t+h}|t = \ell_t + hb_t \quad (2.7)$$

La ecuación de nivel o valor suavizado, ℓ_t , ahora es:

$$\ell_t = \alpha Y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1}) \quad (2.8)$$

Donde b_t es el componente de tendencia:

$$b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1} \quad (2.9)$$

En la ecuación 2.9, b_t representa una estimación de la tendencia (pendiente) de la serie en el tiempo. β es el parámetro de la tendencia y $0 \leq \beta \leq 1$. Este parámetro controla la tasa de cambio o pendiente de la tendencia en el pronóstico. Un valor de β cercano a 1 dará mayor peso a la tendencia más reciente, lo que permitirá que

¹Holt, C. C. (1957). Forecasting seasonals and trends by exponentially weighted averages (ONR Memorandum No. 52). Carnegie Institute of Technology, Pittsburgh USA. Reprinted in the International Journal of Forecasting, 2004.

2.2. EN RELACIÓN A LOS MÉTODOS ESTADÍSTICOS ABORDADOS

el pronóstico se ajuste rápidamente a cambios en la tendencia. Un valor de β cercano a 0 hará que el pronóstico sea menos sensible a cambios en la tendencia y se ajuste más lentamente (Holt, 2004). La ecuación de la tendencia muestra que b_t es un promedio ponderado de la tendencia estimada en el momento t , basada en la diferencia $\ell_t - \ell_{t-1}$ y b_{t-1} , la estimación anterior de la tendencia.

La función de pronóstico ya no es plana, sino que tiene una tendencia. El pronóstico para h pasos adelante es igual al último nivel estimado más h veces el último valor estimado de la tendencia. Por lo tanto, los pronósticos son una función lineal de h , como lo describe la ecuación 2.7.

- 3. Suavizamiento exponencial con tendencia y estacionalidad - Método Holt-Winter:** La estacionalidad es un patrón recurrente y periódico en una serie de tiempo que se repite en intervalos regulares. Son comportamientos específicos que se observan en la serie de tiempo en momentos consistentes dentro de un año. Pueden deberse a factores como eventos climáticos, días festivos, temporadas de ventas, entre otros (Wooldridge, 2010). Como se dedujo en el Exploratory Data Analysis (EDA), la serie bajo estudio, presenta una estacionalidad mensual.

El método de Holt-Winter², es una extensión del método de Holt para captar la estacionalidad de la serie (Lugo y Porto, 2020). Tiene dos variaciones, la variante aditiva que es adecuada cuando la estacionalidad se mantiene razonablemente constante a lo largo de la serie y la variante multiplicativa cuando la estacionalidad cambia proporcionalmente al nivel de la serie (Hyndman y Athanasopoulos, 2021). En este trabajo corresponde aplicar el método aditivo, que queda descrito en el siguiente conjunto de ecuaciones:

²Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. Management Science, 6(3), 324–342.

$$\hat{Y}_{t+h|t} = \ell_t + hb_t + s_{t+h-m(k+1)} \quad (2.10)$$

Donde, k es la parte entera de $((h - 1)/m)$, lo que asegura que las estimaciones de los índices estacionales utilizados para el pronóstico provengan del último año de la muestra.

$$\ell_t = \alpha(Y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1}) \quad (2.11)$$

La ecuación del nivel es un promedio ponderado entre la observación ajustada estacionalmente $(Y_t - s_{t-m})$ y el pronóstico no estacional $(\ell_{t-1} + b_{t-1})$ para el momento t .

$$b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1} \quad (2.12)$$

La ecuación de tendencia (2.12) como se ve es la misma que en el método de Holt.

Y finalmente la componente estacional es igual a:

$$s_t = \gamma(Y_t - \ell_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m} \quad (2.13)$$

Como se puede apreciar la ecuación 2.13 resulta ser un promedio ponderado entre el índice de estacionalidad corriente $(Y_t - \ell_{t-1} - b_{t-1})$ y el índice de estacionalidad de la misma estación en el último año s_{t-m} . El parámetro γ controla la magnitud de la estacionalidad en el pronóstico. Si se utiliza la variante aditiva, un valor de γ cercano a 1 implicará que la estacionalidad se ajusta rápidamente a las fluctuaciones estacionales más recientes, mientras que un valor cercano a 0 hará que el pronóstico sea menos sensible a las variaciones estacionales (Winters, 1960).

Modelos ARIMA

ARIMA significa “*Auto Regressive Integrated Moving Average*” y también se conoce como metodología de Box-Jenkins³. Estos tienen una fuerte presencia en los trabajos de pronóstico de precios en general (Rojas y cols., 2022; Ruiz Hernández y cols., 2019; Zielińska-Sitkiewicz y Chrzanowska, 2021). Constan de tres componentes: autoregresión (AR), diferenciación (I) y media móvil (MA). El componente de autoregresión supone que el valor actual de la variable depende de sus valores pasados. La diferenciación implica transformar los datos tomando las diferencias entre observaciones consecutivas para lograr estacionariedad y capturar los patrones subyacentes. Por último, el componente de media móvil tiene en cuenta los términos de error y suaviza las fluctuaciones aleatorias en los datos (Gujarati y Porter, 2011). Dentro de sus principales ventajas se puede decir que los modelos ARIMA pueden proporcionar pronósticos estables y confiables cuando se ajustan correctamente y se eligen los órdenes adecuados para los componentes AR, I y MA. Sin embargo, encontrar los órdenes óptimos puede requerir un proceso de prueba y error, lo que puede ser más complicado en series de tiempo complejas aunque esta limitación ha disminuido significativamente con el desarrollo de funciones automatizadas en R Core Team. Al igual que los modelos ETS, pueden tener dificultades para hacer pronósticos precisos a largo plazo. Dada su amplia presencia y relevancia en la literatura académica, esta investigación incluye el modelo ARIMA.

1. **Modelo ARIMA sin estacionalidad:** El modelo $ARIMA(p, d, q)$ tiene tres componentes principales:

Componente Autoregresivo(AR): El término “Autoregresivo” hace referencia a que el valor actual de la serie de tiempo está relacionado con sus valores pasados. El orden de este componente está representado por p , que indica cuántos períodos anteriores se utilizan en el modelo. Si $p = 1$, entonces se utilizará el valor inmediatamente

³Box, G. E. P., & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control* (Revised Edition). San Francisco: Holden-Day

anterior para predecir el valor actual. Si $p = 2$, se considerarán los dos valores anteriores, y así sucesivamente.

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t \quad (2.14)$$

Donde $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$ son los valores anteriores de la serie de tiempo utilizados en el modelo, $\phi_1, \phi_2, \dots, \phi_p$ son los coeficientes autorregresivos y ε_t es el término de error en el momento t , que representa el ruido o la parte no explicada por el modelo (Hyndman y Athanassopoulos, 2021).

Componente de Integración (I): El término de “Integración” se refiere a la necesidad de realizar alguna diferenciación en la serie de tiempo para hacerla estacionaria.

Como definen Gujarati y Porter (2011): “En términos generales, se dice que un proceso estocástico es estacionario si su media y su varianza son constantes en el tiempo y si el valor de la covarianza entre dos periodos depende solo de la distancia o rezago entre estos dos periodos, y no del tiempo en el cual se calculó la covarianza.” Si una serie no es estacionaria, habitualmente, su integración suele resolver esa situación.

Si se aplica una diferenciación, $d = 1$, se verá así:

$$\Delta Y_t = Y_t - Y_{t-1} = \varepsilon_t \quad (2.15)$$

Donde ΔY_t representa la diferencia entre el valor actual y el valor anterior de la serie y ε_t es el término de error. Esta operación se puede repetir d veces hasta que se logre la estacionariedad. La estacionariedad es importante porque asegura que la serie de tiempo no tenga tendencia o estacionalidad, lo que facilita el modelado y pronóstico (Gujarati y Porter, 2011).

Componente de promedio Móvil (MA): El término de “Promedio Móvil” se refiere a que el valor actual de la serie de tiempo está relacionado con los errores pasados del modelo. El orden de este com-

2.2. EN RELACIÓN A LOS MÉTODOS ESTADÍSTICOS ABORDADOS

ponente está representado por q , que indica cuántos errores pasados se utilizan en el modelo para predecir el valor actual. Si $q = 1$, entonces se utilizará el último error para hacer la predicción. Si $q = 2$, se considerarán los dos errores pasados, y así sucesivamente.

$$Y_t = c + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \dots + \theta_q\varepsilon_{t-q} \quad (2.16)$$

Donde c es una constante (intercepto) en el modelo. ε_t es el término de error en el momento t . Los parámetros $\theta_1, \theta_2, \dots, \theta_q$ son los coeficientes de la media móvil. Y $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-p}$ son los errores pasados utilizados en el modelo (Gujarati y Porter, 2011).

En resumen, de ser necesario, el modelo $ARIMA(p, d, q)$, combina estas tres componentes en una sola ecuación que se utiliza para modelar y pronosticar la serie de tiempo. La ecuación completa será la suma de las ecuaciones para cada componente con sus respectivos coeficientes. La notación completa del modelo $ARIMA(p, d, q)$ se puede expresar como:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (2.17)$$

2. Modelo ARIMA con estacionalidad

El modelo $SARIMA(p, d, q)(P, D, Q)_m$, también conocido como *Seasonal Autoregressive Integrated Moving Average*, es una extensión del modelo **ARIMA** que incluye componentes estacionales para capturar patrones cíclicos y estacionales presentes en una serie de tiempo (Ruiz Hernández y cols., 2019).

Componente Autoregresivo(AR): Al igual que en el modelo **ARIMA**, el componente AR del modelo SARIMA se refiere a las relaciones lineales entre el valor actual de la serie de tiempo y sus valores pasados, pero esta vez incluye términos estacionales. La ecuación de la componente AR se representa como:

$$Y_t = c + \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{i=1}^P \Phi_i Y_{t-im} + \varepsilon_t \quad (2.18)$$

Donde:

Y_t es el valor de la serie de tiempo en el momento t .

c es el intercepto del modelo.

ϕ_i son los coeficientes autoregresivos no estacionales, para $i = 1, 2, \dots, p$.

Y_{t-i} son los valores anteriores de la serie de tiempo utilizados en el componente AR no estacional.

Φ_i son los coeficientes autoregresivos estacionales, donde $i = 1, 2, \dots, P$.

Y_{t-im} son los valores anteriores de la serie de tiempo separados por el período estacional m .

Componente de Integración (I): Al igual que en el modelo ARIMA, el componente I del modelo SARIMA se refiere a la diferenciación de la serie de tiempo para lograr la estacionariedad. En este caso:

$$\Delta^d \Delta_m^D Y_t = (1 - B)^d (1 - B^m)^D Y_t = \varepsilon_t \quad (2.19)$$

Donde:

Δ^d representa el operador de diferencia no estacional de orden d , que implica realizar d diferenciaciones no estacionales para lograr la estacionariedad de la serie de tiempo.

Δ_m^D representa el operador de diferencia estacional de orden D , que implica realizar D diferenciación estacionales para lograr la estacionariedad de la serie de tiempo.

B es el operador de rezago no estacional, que desplaza los valores de la serie de tiempo en un paso hacia atrás.

B^m es el operador de rezago estacional, que desplaza los valores de la serie de tiempo en un múltiplo del período estacional m .

2.2. EN RELACIÓN A LOS MÉTODOS ESTADÍSTICOS ABORDADOS

Componente de promedio Movil (MA)

$$Y_t = c + \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \sum_{i=1}^Q \Theta_i \varepsilon_{t-im} \quad (2.20)$$

Donde:

Y_t es el valor de la serie de tiempo en el momento t .

c es el intercepto del modelo.

ε_t es el término de error en el momento t .

θ_i son los coeficientes de la media móvil no estacional, donde $i = 1, 2, \dots, q$.

ε_{t-i} son los errores pasados utilizados en el componente MA no estacional.

Θ_i son los coeficientes de la media móvil estacional, donde $i = 1, 2, \dots, Q$.

ε_{t-im} son los errores pasados utilizados en el componente MA estacional, donde im representa el múltiplo del período estacional m .

El modelo SARIMA(p, d, q)(P, D, Q)m combina todas estas componentes, ecuaciones 2.18, 2.19 y 2.20, en una sola ecuación que se utiliza para modelar y pronosticar la serie de tiempo.

Modelo PROPHET

Prophet es un modelo desarrollado por Facebook (Taylor y Letham, 2018) que se utiliza para el pronóstico y análisis de datos de series temporales, especialmente en el contexto de pronósticos de demanda y tendencias a corto y medio plazo. Fue diseñado para ser fácil de usar, escalable y capaz de manejar datos con cambios en la tendencia y efectos estacionales (Chan, 2020). También permite incluir regresores adicionales si se dispone de variables explicativas que puedan influir en la serie de tiempo.

Este modelo distingue en la serie en tres componentes:

$$Y(t) = g(t) + s(t) + h(t) + \varepsilon_t \quad (2.21)$$

Donde:

Y_t es el valor de la serie de tiempo en el momento t .

$g(t)$ es la componente de tendencia que modela la dirección general del cambio en la serie de tiempo.

$s(t)$ es la componente de efectos estacionales que captura las variaciones periódicas en la serie de tiempo.

$h(t)$ es la componente de días festivos que modela los efectos específicos de fechas festivas o eventos especiales.

ε_t es el término de error en el momento t , que representa la variación no explicada por el modelo.

A continuación se explica los tres componentes que identifica PROPHET en cada serie:

1. **Tendencia:** Una de las principales ventajas de este modelo es que permite identificar tendencias no lineales. Para ello utiliza el concepto de *change points*. Emplea una aproximación basada en la segmentación de la serie de tiempo en múltiples segmentos lineales entre los *change points* detectados (Hyndman y Athanasopoulos, 2021). Cada segmento representa una tendencia constante en la serie antes de que ocurra un cambio en el comportamiento. La función no lineal de tendencia que se ajusta a los datos, se compone de una serie de segmentos lineales que se unen en los puntos de cambio. El uso de *change points* en la identificación de la tendencia en Prophet permite que el modelo capture cambios en la dirección de la serie de tiempo de manera más precisa, lo que es especialmente útil cuando la serie muestra patrones de tendencia no lineales o cuando hay eventos significativos que afectan la dirección de la serie

$$g(t) = (k + \sum_{i:t>s_i})\delta_i t + (m + \sum_{j:t>s_j})\gamma_j \quad (2.22)$$

Donde:

k es una constante que representa el nivel base de la tendencia en el inicio de la serie de tiempo ($t = 0$).

2.2. EN RELACIÓN A LOS MÉTODOS ESTADÍSTICOS ABORDADOS

δ_i es una constante que representa el cambio lineal en la tendencia en el momento s_i .

m es una constante que representa el cambio base en la tasa de crecimiento de la tendencia.

γ_i es una constante que representa el cambio no lineal en la tasa de crecimiento en el momento s_j .

Específicamente, la ecuación tiene dos sumas:

- a) $\sum_{i:t>s_i} \delta_i$: Esta suma recorre todos los cambios lineales en la tendencia que han ocurrido hasta el momento actual t , después de cada punto de cambio s_i . Es decir, para cada punto en el tiempo t , esta suma acumula todos los cambios (δ_i) que han ocurrido después de cada punto de cambio en la tendencia, antes de t .
- b) $\sum_{j:t>s_j} \gamma_j$: esta suma acumula todos los cambios no lineales en la tasa de crecimiento (γ_j) después de cada punto de cambio no lineal s_j hasta el momento actual t .

En definitiva, la función de tendencia $g(t)$ en PROPHET tiene una parte lineal que depende del nivel base k , los cambios lineales δ_i en los puntos de cambio s_i , y una parte no lineal que depende del cambio base m y los cambios no lineales γ_j en los puntos de cambio s_j . Esta combinación de términos lineales y no lineales permite que el modelo Prophet se ajuste a patrones de tendencia más complejos y no lineales en la serie de tiempo.

2. **Efectos estacionales:** El modelo Prophet identifica los efectos estacionales y los modela utilizando una aproximación basada en series de Fourier. La aproximación de Fourier es una técnica matemática que descompone una función periódica en una serie de funciones seno y coseno de diferentes frecuencias (Menculini y cols., 2021). Al utilizar la aproximación de Fourier, el modelo Prophet puede capturar patrones estacionales complejos y no lineales que pueden no ser

fácilmente representados por componentes simples de estacionalidad aditiva o multiplicativa. Esto proporciona una mayor flexibilidad al modelo para adaptarse a estacionalidades cambiantes, lo que es particularmente útil cuando la serie de tiempo muestra estacionalidades que varían en intensidad o duración a lo largo del tiempo, es decir, multiestacionalidad.

Y finalmente, su API (*Application Programming Interface*) y sintaxis intuitiva facilitan el proceso de ajustar modelos y obtener pronósticos (Chan, 2020). Por estas razones, PROPHET ha sido seleccionado como una opción adecuada y prometedora para abordar el pronóstico del precio del cordero patagónico.

Con la inclusión de estos tres modelos, se busca aplicar una metodología que permita realizar un análisis comparativo exhaustivo y determinar cuál de ellos ofrece un pronóstico más preciso y confiable para el precio del cordero patagónico en función de las características específicas de los datos y la naturaleza de la serie de tiempo estudiada. Es importante señalar que la gama de modelos disponibles para abordar esta tarea es extensa y variada, y los tres mencionados son solo una selección de las opciones disponibles. Otras técnicas, como los modelos de redes neuronales, modelos estacionales o de regresión, también tienen relevancia en el campo del pronóstico y podrían ser objeto de análisis en futuras investigaciones. En este contexto, el presente trabajo busca ofrecer un enfoque académico y riguroso al evaluar y comparar la capacidad predictiva de estos tres modelos, siendo conscientes de que el espectro de posibles enfoques y métodos es amplio y diverso.

Validación Cruzada y Comparación de los Modelos

La validación cruzada es una técnica crucial en el aprendizaje estadístico, fundamental para evaluar la eficacia predictiva y el ajuste de los modelos estadísticos y de aprendizaje automático. Se basa en el principio de dividir el conjunto de datos en partes, para luego utilizar estas divisiones de manera iterativa como conjuntos de entrenamiento y prueba. El objetivo

2.2. EN RELACIÓN A LOS MÉTODOS ESTADÍSTICOS ABORDADOS

es evaluar cómo el modelo se desempeña con datos que no ha visto durante el entrenamiento, proporcionando una medida robusta de su capacidad predictiva general.

Existen varios enfoques de validación cruzada, siendo la validación cruzada de K -folds uno de los más utilizados. En este método, el conjunto de datos se divide en k particiones (o “folds”) de tamaño similar. En cada iteración, uno de los k folds se utiliza como conjunto de prueba y el resto como conjunto de entrenamiento. Este proceso se repite k veces, con cada uno de los k folds utilizado exactamente una vez como conjunto de prueba. Finalmente, el rendimiento del modelo se evalúa promediando los resultados obtenidos en cada iteración vinculados al error de prueba, ofreciendo una visión global del rendimiento del modelo (James y cols., 2021).

Dada la naturaleza de las series de tiempo, que presentan una estructura dependiente del tiempo, la división aleatoria del conjunto de datos que se realiza en la validación cruzada tradicional podría generar la pérdida de la estructura temporal, afectando la calidad del pronóstico. Para evitar este problema en nuestro análisis, se implementa una variante de validación cruzada estudiada por Bergmeir y Benítez (2012) que respeta el orden temporal de los datos. En este enfoque, llamado por Bergmeir y Benítez (2012) “punto de origen móvil para el conjunto de evaluación” los conjuntos de entrenamiento contienen todas las observaciones hasta cierto punto en el tiempo, mientras que los conjuntos de prueba contienen las observaciones posteriores, asegurando que no haya datos futuros en el proceso de entrenamiento.

La selección del número adecuado de k es un desafío clave en la validación cruzada. En la práctica, es común utilizar valores como $k = 5$ o $k = 10$, ya que han demostrado empíricamente proporcionar estimaciones del error de prueba con un sesgo y una varianza equilibrados (James y cols., 2021). Un enfoque extremo consiste en utilizar un k igual al número de observaciones, conocido como *leave-one-out cross-validation* (LOOCV), pero puede resultar computacionalmente costoso y generar un mayor sesgo. Por otro lado, un valor más pequeño para k reducirá el sesgo pero aumentará la varianza en las estimaciones de error (James y cols., 2021). En este

estudio, se trabaja con un valor de $k = 5$ para encontrar un equilibrio adecuado entre sesgo y varianza en las estimaciones de rendimiento de los modelos de pronóstico de precios del cordero patagónico. De esta manera, se obtienen resultados confiables y sólidos para la evaluación comparativa de los modelos seleccionados (Bergmeir y Benítez, 2012).

La comparación de modelos constituye uno de los pilares fundamentales de esta tesis y es la esencia misma del aprendizaje estadístico ya que permite seleccionar el modelo que mejor se ajusta a los datos observados, bajo ciertos criterios de evaluación. Para lograr esa evaluación, es necesario calcular y comparar medidas de evaluación robustas que reflejen adecuadamente el ajuste de los distintos modelos. Los estudios que han abordado la comparación entre modelos de pronóstico han utilizado diversas métricas para este propósito. Por ejemplo, Fernández Lavallo (2019) ordena la performance de los modelos en función de métricas como el RMSE y el MAPE. Por su parte, Chan (2020) emplea el MAPE para evaluar el rendimiento comparativo entre los modelos ARIMA y PROPHET. Asimismo, la investigación comparativa de Menculini y cols. (2021) utiliza el RMSE, el MAE, el MAPE y el ME.

Es esencial tener en cuenta que ninguna medida de evaluación es perfecta por sí sola y la elección de las métricas de precisión adecuadas dependerá de la naturaleza particular de los datos y los objetivos del análisis. Por tanto, es recomendable considerar múltiples medidas y técnicas de validación cruzada para obtener una evaluación completa y robusta del rendimiento de los modelos de pronóstico (Hyndman y Athanasopoulos, 2021). Al combinar diversas métricas, como el MAE, el RMSE, y el MAPE, esta tesis garantiza una evaluación integral que permitirá una comprensión más profunda y precisa del desempeño de los modelos seleccionados para el pronóstico del precio del cordero patagónico.

Mediante esta comparación exhaustiva de modelos y el análisis de sus resultados, se obtendrán conclusiones sólidas y fundamentadas que proporcionarán una visión clara sobre qué modelo es más adecuado para este contexto específico. Asimismo, esta evaluación permitirá identificar las for-

2.2. EN RELACIÓN A LOS MÉTODOS ESTADÍSTICOS ABORDADOS

talezas y limitaciones de cada modelo, proporcionando una base sólida para futuras investigaciones y mejoras en el pronóstico del precio del cordero patagónico.

Las métricas empleadas para evaluar los modelos, se basan en el error de pronóstico que es la diferencia entre el valor observado y su valor pronosticado.

$$e_{T+h} = Y_{T+h} - Y_{T+h|T} \quad (2.23)$$

Donde el grupo de entrenamiento esta dado por $\{Y_1, Y_2, \dots, Y_T\}$ y el grupo de prueba por $\{Y_{T+1}, Y_{T+2}, \dots\}$.

El error de pronóstico se puede resumir de diferentes formas:

1. La media del valor absoluto del error, **MAE**, por sus siglas en inglés:

$$MAE = mean(|e_t|) \quad (2.24)$$

2. La raíz de la media de los errores al cuadrado, **RMSE**, por sus siglas en inglés:

$$RMSE = \sqrt{mean(e_t^2)} \quad (2.25)$$

Tanto **MAE** como **RMSE** dependen de la escala de los datos originales. En estudios con una única serie de tiempo esto no es inconveniente, pero cuando se disponen de múltiples series otras medidas pueden resultar mas adecuadas (Hyndman y Athanasopoulos, 2021).

3. El error porcentual esta dado por $p_t = e_t/Y_t 100$, su principal ventaja es que es adimensional y su desventaja es que en aquellas series que presentan valor cero, toma valor infinito o indefinido (Hyndman y Athanasopoulos, 2021). La media del valor absoluto del error porcentual, **MAPE**, por sus siglas en inglés se calcula:

$$MAPE = mean(|p_t|) \quad (2.26)$$

La integración de múltiples métricas ofrece una evaluación más holística y precisa, esencial para la selección del modelo predictivo. Este enfoque no solo facilita conclusiones fundamentadas sobre el modelo más eficaz en

este contexto, sino que también sienta las bases de una metodología adecuada para futuras investigaciones en el campo del análisis predictivo.

En resumen, esta tesis tiene como objetivo principal diseñar una metodología dinámica de pronóstico para los precios del cordero patagónico, aprovechando las capacidades del aprendizaje estadístico aplicado a series de tiempo. Este enfoque se sustenta en la aplicación de tres modelos dentro del marco del aprendizaje supervisado, tales como el **ETS**, **ARIMA** y **PROPHET**, seleccionados por su relevancia y adaptabilidad a las peculiaridades del mercado del cordero patagónico. Este trabajo propone una aproximación sistemática y detallada para entender los patrones que presenta el precio del mercado, incorporando un enfoque de validación cruzada adaptado a series de tiempo, lo cual permite una evaluación precisa y objetiva del desempeño predictivo de cada modelo. Al integrar meticulosamente estos elementos, la metodología propuesta se propone como una respuesta a la búsqueda de soluciones prácticas y fundamentadas para enfrentar los desafíos, del mercado de carne de cordero.

CAPÍTULO 3

Metodología

3.1. Metodología

Desde el punto de vista de la metodología, este trabajo se puede clasificar como un estudio observacional. En el campo de la economía y la estadística, este tipo de estudio se refiere a un método de investigación que implica recopilar datos a través de la observación directa de fenómenos o eventos del mundo real relacionado a los procesos económicos (Paldam, 2021).

A partir de datos recabados del “*Informe de Precio de Carne y Ganado de la Patagonia*”¹, proporcionado por el INTA, se ha centrado la atención en el precio del Cordero Liviano de hasta 12 kilos para Faena, abarcando un periodo desde abril de 2014 hasta noviembre de 2023.

En la fase inicial, se llevó a cabo un análisis exploratorio de datos, empleando herramientas especializadas para la gestión y transformación temporal de los datos, lo que permitió una primera aproximación a las características fundamentales de la serie de tiempo estudiada. Este paso preli-

¹Bolla, D. A., et al. (2023). *Informe de precios de carne y ganado de la Patagonia*. INTA.

minar fue esencial para preparar el terreno para las etapas de modelado estadístico y aprendizaje automático, en las cuales se evaluaron modelos **ARIMA**, **ETS** y **PROPHET** por su capacidad para adaptarse y predecir con precisión las dinámicas de precios observadas. La selección de estos modelos se basó en una revisión exhaustiva de la literatura científica y en la adaptabilidad de cada modelo a las especificidades de los datos en estudio.

La evaluación de los modelos, mediante técnicas de validación cruzada diseñadas específicamente para series de tiempo, constituyó un paso crítico para asegurar la fiabilidad y precisión de los pronósticos generados. Este proceso iterativo de ajuste y selección del modelo más adecuado culminó en la generación de pronósticos que reflejan las tendencias futuras de los precios en el mercado objetivo. La metodología aplicada, por lo tanto, no solo satisface los objetivos de esta investigación sino que también contribuye al entendimiento de las fluctuaciones de precios en el sector agropecuario, subrayando la importancia de una práctica de reevaluación continua frente a las dinámicas cambiantes del mercado.

3.1.1. Recopilación y configuración inicial de datos

Esta tesis se basó en el análisis de series de tiempo usando datos extraídos del informe ([Tejada y et al., 2023](#)), publicado conjuntamente por los centros regionales Patagonia Norte y Patagonia Sur, dependientes del **INTA**. Los precios informados incluye los concernientes al mercado ovino, bovino, porcino y caprinos de la región Patagónica. Los mismos se recopilan con una regularidad mensual. El relevamiento comprende diferentes regiones: Patagonia Norte A (Norte de Río Negro y Partido de Patagones), Patagonia Norte B (Provincia de Neuquén y Cordillera de Río Negro), Chubut y Santa Cruz. Los precios de ovinos se discriminan en cuatro categorías: Adultos para Faena, Cordero Liviano de hasta 12 kilos para Faena, Cordero Pesado Mayor a 12 kilos para Faena y Oveja de Refugio para Invernada. Todos los precios corresponden al kilo limpio al gancho, excepto Oveja de Refugio que es por animal. Estos precios son sin IVA, recibidos

por los productores.

Con fin de estudiar el precio de un mercado con características similares, en esta tesis se trabajó con las series referida al precio del Cordero Liviano de hasta 12 kilos para Faena. El periodo analizado comprende desde abril de 2014 a noviembre de 2023, lo que totalizó 117 meses.

Se tomó la decisión de emplear valores nominales de precios. Los precios nominales reflejan el costo de bienes y servicios en términos corrientes, sin ajustar por los efectos de la inflación, capturando así el precio de mercado tal como se presenta en el momento de la transacción. En contraste, los precios reales se ajustan tomando en cuenta la inflación, ofreciendo una medida que refleja el verdadero poder adquisitivo al eliminar el impacto de la variación de precios a lo largo del tiempo. Esta decisión se basó en varios factores claves que se alinean con los objetivos de pronóstico a corto plazo de esta tesis:

1. Trabajar con precios reales es principalmente útil cuando se trabaja en proyectos de mediano y largo plazo. Como el horizonte de pronóstico es de un año hacia adelante, los valores nominales pueden proporcionar información suficiente para un análisis de corto plazo.
2. Esta elección es consistente con la forma en que se publican oficialmente los datos del mercado de la carne en la Patagonia.
3. Los precios nominales son más fáciles de comunicar al público en general, lo que es importante cuando se divulgan resultados de un análisis a una audiencia no especializada.
4. Los precios nominales pueden mostrar más variabilidad y tendencias que los precios reales, que se suelen mantener constante a lo largo del tiempo, lo que permite sacar provecho de los métodos de aprendizaje estadístico, que se basan en la detección de patrones en los datos.
5. De esta forma, se evita la complejidad inherente a la selección de un índice de precios adecuado para ajustar los valores a términos reales.

Las series de precios de cada región presentaron algunos datos faltantes: uno para la región Patagonia Norte A, ocho para la región Patagonia Norte B, tres para Chubut y 27 para Santa Cruz. Para evitar las limitaciones de los datos faltantes y poder plantear un modelo univariado de serie de tiempo, se tomó como variable de respuesta, el Precio Promedio Mensual entre regiones y se denotó como *PPCord*. Aunque hay variabilidad en los precios individuales, el promedio ofrece una visión general de la tendencia de precios en la macro-región, Patagonia. Dicha estrategia es similar a la empleada por [Zielińska-Sitkiewicz y Chrzanowska \(2021\)](#), quien también trabajó con precios promedios. Esta manipulación previa se realizó mediante hoja de cálculo.

Se combinaron técnicas de análisis de series de tiempo, aprendizaje estadístico, validación cruzada y revisión de la documentación científica. La metodología se desarrolló utilizando el lenguaje de programación R version 4.2.1 ([R Core Team, 2023](#)).

En el marco de esta tesis, se aprovecharon diversos paquetes de análisis de series de tiempo que forman parte del ecosistema *Tidymodels* ([Kuhn, Vaughan, y Wickham, 2021](#)). Este está diseñado para facilitar y estandarizar el proceso de modelado estadístico y aprendizaje automático. La filosofía detrás de este ecosistema, es hacer que el modelado sea más transparente, reproducible y fácil de realizar mediante la adopción de principios de *Tidyverse*, otro ecosistema, conocido por su enfoque estructurado y coherente para el análisis de datos ([Wickham, 2021](#)). *Tidymodels* promueve la idea de que el proceso de modelado debe ser claro y transparente, desde la preparación de los datos hasta la evaluación del modelo. Al adherirse a esta filosofía, los usuarios pueden crear flujos de trabajo más consistentes y reutilizables, lo que facilita la comparación de diferentes modelos y enfoques.

Los paquetes esenciales utilizados en esta tesis son *Timetk*, *Modeltime* y *Forecast*, que están diseñados para trabajar de manera fluida y complementaria.

Timetk, se centra en la manipulación y preprocesamiento de datos específicos de series de tiempo. Ofrece funcionalidades para realizar tareas como

limpieza de datos, agregación temporal, transformaciones y visualización, todo ello en línea con los principios de Tidyverse.

`Forecast` y `Modeltime` son paquetes que pueden utilizarse en conjunto para el modelado y pronóstico de series de tiempo. `Modeltime` presenta un enfoque más estructurado y automatizado, mientras que `Forecast` ofrece una amplia gama de modelos y técnicas tradicionales de pronóstico. La relación entre estos dos paquetes es que `Modeltime` puede aprovechar algunos de los modelos y técnicas disponibles en `Forecast` como parte de su conjunto de herramientas.

Se realizó una preparación inicial sobre los datos para garantizar la calidad de los mismos, como la identificación y manejo de valores atípicos o datos faltantes. Para esto se trabajó con paquetes de R como `lubridate` (Grolemund y Wickham, 2011) y `Timetk` (Dancho y Vaughan, 2022), como se describe en la siguiente sección.

3.1.2. Análisis exploratorio de datos:

Como se mencionó, la serie de tiempo cuenta con un total de 117 observaciones mensuales correspondiente al periodo abril-2014 a noviembre-2023.

Se utilizó el paquete `lubridate` (Grolemund y Wickham, 2011) para dar formato de serie de tiempo a los datos. Esto se hace porque, si bien los precios son captados correctamente como números (formato *numeric*), R desconoce las fechas al importarlas, las mismas son tomadas como *string*, es decir, texto. Luego de utilizar la función `dmy()` la columna que describe al tiempo queda con formato *date*, (ver apéndice A.1.2), condición requerida para continuar trabajando en el software.

Para la visualización de la serie, en su escala original y su versión logarítmica, se empleó la función `plot_time_series()` del paquete `Timetk` (Dancho y Vaughan, 2022). Tras realizar la transformación, se ejecutaron varias representaciones para un análisis más profundo. Con la función `plot_seasonal_diagnostics()`, se realizaron los diagramas de caja con diferente periodicidad. La detección gráfica de anomalías se efectuó con la

función `plot_anomaly_diagnostics()`, que para identificar los valores atípicos, considera aquellos datos que caen por debajo del cuartil uno (Q1) o por encima del cuartil tres (Q3) en una magnitud equivalente a tres veces el rango intercuartílico (Dancho y Vaughan, 2022). Las representaciones de la autocorrelación y la autocorrelación parcial se llevó a cabo con 25 rezagos, con la función `plot_acf_diagnostics()`. El último diagnóstico gráfico implicó descomponer la serie con la función `plot_stl_diagnostics()`, que utiliza el método *Seasonal - Trend - Loess* (STL)² (ver apéndice A.1.3). El procedimiento parte de la idea que la serie temporal se puede descomponer en todos o algunos de los siguientes componentes:

1. Tendencia (T), evolución de la serie en el largo plazo.
2. Fluctuación cíclica (C), de carácter periódico, en el mediano plazo en torno a la tendencia.
3. Variación estacional (S), recoge aquellos comportamientos de tipo regular y repetitivo que se dan un período de tiempo, generalmente igual o inferior a un año.
4. Y los movimientos irregulares (e), que pueden ser aleatorios y recogen los efectos accidentales, o erráticos.

La asociación de estos cuatro componentes en una serie de tiempo, puede responder a esquemas aditivos, multiplicativos o una combinación de ambos. Este trabajo, se ajusta al esquema aditivo:

$$Y_t = T_t + C_t + S_t + e_t \quad (3.1)$$

Si bien hay otros métodos de descomposición, por lo general, los resultados son muy similares (Parra, 2019).

²Propuesto originalmente por Cleveland (1979) y desarrollado por Cleveland y Devlin (1988)

3.1.3. Preparación de los datos

Se procedió a estandarizar la serie. Al estandarizar los datos, se puede simplificar el proceso de selección de modelos y la comparación de diferentes modelos, ya que todos estarán en la misma escala. Se guardaron los parámetros de estandarización para luego realizar el proceso inverso.

Estadístico	Valor
Media	5.17
Desvío Estándar	1.109

Fue necesario extender la base de datos en unidades de tiempo equivalentes al horizonte de pronóstico deseado. La elección del horizonte de pronóstico debe ser evaluada con atención, considerando las necesidades y objetivos específicos de la investigación, así como la disponibilidad de datos y la precisión requerida (Hyndman y Athanasopoulos, 2021). Los valores de pronóstico de diferentes horizontes tienen diferentes propiedades estadísticas. A medida que aumenta el horizonte de pronóstico, aumenta la incertidumbre y la varianza de las predicciones. Esto significa que las predicciones a corto plazo pueden ser más precisas y confiables que las predicciones a largo plazo y por otra parte, el desempeño de los métodos de pronóstico puede depender del horizonte utilizado como se mencionó en Bergmeir y Benítez (2012).

El objetivo de esta tesis es realizar el pronóstico del precio del cordero patagónico un ciclo productivo hacia adelante, es decir 12 meses. Con este horizonte el número de filas de la tabla para pronóstico pasa a ser 129, es decir, las 117 observaciones originales más los 12 meses para pronóstico (ver apéndice A.1.4)

Realizadas estas adecuaciones, los 117 datos, se particionan en dos: el conjunto de entrenamiento y el conjunto de testeo. Esto se realiza con la función `time_series_split()`, del paquete `Timetk`. Los algoritmos se ajustan utilizando el conjunto de entrenamiento, y los modelos ajustados se utilizan para predecir las respuestas para las observaciones en el conjunto

de testeo (James y cols., 2021).

El porcentaje de datos utilizados para el entrenamiento fue del 80%. El esquema 80-20, es propuesto por Hyndman y Athanasopoulos (2021). De esta forma se reservan 24 observaciones para que integren el grupo de testeo, que en función de la cantidad total de datos disponibles representa aproximadamente un 20,5%. El mencionado autor sostiene que el conjunto de prueba debería ser al menos tan grande como el horizonte de pronóstico requerido. Como en este caso es de 12 meses, queda satisfecho ese requisito. También advierte que para una evaluación efectiva del modelo de pronóstico, resulta esencial contar con un conjunto de prueba representativo que contenga datos "no vistos", es decir, datos que no fueron utilizados en el proceso de entrenamiento del modelo. Esto permite evaluar el rendimiento del modelo en situaciones futuras y garantizar la fiabilidad de las predicciones. La gráfica que muestra la partición hecha se confecciona con la función `plot_time_series_cv_plan()`.

3.1.4. Selección y ajuste de los modelos:

Basándose en la revisión de la documentación científica, se seleccionaron los modelos ARIMA y ETS y PROPHET.

Se ajustaron los modelos seleccionados utilizando técnicas de estimación y optimización de parámetros, teniendo en cuenta las características específicas de la serie de tiempo. Esta tarea se llevo a cabo con los paquetes `modeltime` (Dancho, 2023) y `tidymodels` (Kuhn y cols., 2021).

Para ajustar el modelo ETS se trabajó con la función `exp_smoothing()`, que permite configurar las características del modelo. El ajuste se logra estableciendo en la función `set_engine()` qué modelo a utilizar, en este caso "ets" y se provee la formula del modelo en la función `fit()` (ver apéndice A.1.6). Este procedimiento se repitió para las tres versiones de ETS analizadas: suavizamiento exponencial simple, suavizamiento exponencial con tendencia y suavizamiento exponencial con tendencia y estacionalidad.

Para aplicar el algoritmo de los modelos ARIMA, se usó la función

`arima_reg()`. Existen dos alternativas de especificación de modelo: “`auto_arima`” y “`arima`”.

La primera opción es la automatizada (`auto_arima`), que utiliza métodos de búsqueda y evaluación automática para explorar diversas combinaciones de órdenes (p, d, q) y selecciona el modelo ARIMA que tenga el valor más bajo de AIC (Criterio de Información de Akaike) o BIC (Criterio de Información Bayesiana). Esta opción proporciona un punto de partida sólido para el ajuste del modelo más adecuado, pero en algunos casos, pueden requerirse ajustes manuales para obtener un rendimiento aún mejor. La segunda alternativa, simplemente denominada `arima`, permite al investigador introducir manualmente los valores de p, d, q, P, D y Q . La información recabada durante el análisis exploratorio de datos (EDA) es útil para establecer estos valores. En base a esto, se corrió inicialmente la opción automatizada que arrojó como resultado un $ARIMA(0, 1, 0)$. Sin embargo, no logró capturar la estacionalidad mensual, presente en los datos. Para abordar esta limitación, se configuró un modelo SARIMA, utilizando la opción manual `arima`. Tomando como base el resultado de la versión automatizada y tras una revisión de la información del EDA, se configuró un modelo SARIMA de orden $(1, 1, 0)(1, 1, 0)_{12}$, que agrega la presencia de un componente estacional con un período de 12 meses (ver apéndice A.1.7).

`Prophet_reg()` es una función que permite generar una especificación del modelo PROPHET antes de ajustarlo. El software utiliza aproximación por máxima verosimilitud (*approximate maximum likelihood estimation*) para ajustar los parámetros del modelo.

Primeramente, se ejecutó el modelo PROPHET con los ajustes por defecto. La versión por defecto implica un *número de puntos de cambio potenciales* (`changepoint_num`) de 25, un *rango para los puntos de cambio* (`changepoint_range`) de 0.8, lo significa que no puede existir un punto de cambio después del primer 80% de los datos y un valor para la *escala previa de los puntos de cambio* (`prior_scale_changepoints`) de 0.05, que controla la fuerza con la que se permite que la tendencia cambie en los puntos de cambio. La función permite configurar la estacionalidad diaria y semanal, pero en base a la naturaleza de los datos estas no son requeri-

das.

A continuación se procedió a configurar manualmente algunos parámetros para mejorar la calidad del modelo:

Se cambió el número de puntos de cambio potenciales (parámetro `changeoint_num`) de 25 a 4 que se selecciono por prueba y error buscando el menor valor, tal que no incremente las métricas del error, **MAE**, **MAPE**, **RMSE**. Un valor grande innecesario de este hiperparámetro, puede acarrear algunos problemas. Por ejemplo, cuando el modelo se ajusta a tal punto que modela las fluctuaciones aleatorias o ruido de los datos históricos, en lugar de capturar las tendencias y patrones subyacentes que son generalizables a nuevos datos. Este sobreajuste, puede resultar en pronósticos poco precisos y poco fiables en datos futuros, por otro lado, se agrega una complejidad innecesaria al modelo, lo que afecta negativamente la calidad del ajuste y las predicciones.

Lo segundo que se configuró es el rango de los puntos de cambio, aumentándolo de 0.8 a 0.9. Las ventajas de este cambio es que permite que el modelo se ajuste a cambios más sutiles en la tendencia de la serie de tiempo. Es importante no excederse ya que podría implicar tanto sobreajuste como inestabilidad (ver apéndice **A.1.8**).

Por último, se ajustó el hiperparámetro `prior_scale_changeoints`, que pasó de su valor por default igual a 0.05 a 0.2. “Valores más grandes permiten que el modelo se ajuste a fluctuaciones estacionales más grandes, mientras que valores más pequeños atenúan la estacionalidad” (**Dancho, 2023**). Así, al aumentar el valor de `prior_scale_seasonality`, se permitió que el modelo sea más flexible, lo que significa que puede captar fluctuaciones estacionales más pronunciadas y complejas. Podría tomar valores incluso mayores a 1 pero desde luego que, un ajuste en exceso, puede provocar sobreajuste y ruido en las predicciones.

La personalización de estos hiperparámetros se realizó mediante un proceso de prueba y error, supervisando cuidadosamente el impacto en el **MAE** para asegurar que los ajustes mejoraran la calidad del modelo sin incrementar el error. De este modo, se buscó un equilibrio entre la capacidad del modelo para captar la complejidad de los datos y su generalización a

nuevos datos.

3.1.5. Validación cruzada en series temporales:

Una vez identificadas las mejores versiones de cada algoritmo, se procedió a realizar la comparación de la precisión de los modelos mediante el uso de la metodología de validación cruzada. Dicha precisión solo se puede determinar considerando qué tan bien funciona un modelo en datos nuevos, que no se utilizaron para ajustar el modelo (Hyndman y Athanassopoulos, 2021). La validación cruzada implica dividir los datos disponibles en conjuntos de entrenamiento y prueba. Los conjuntos de entrenamiento se usan para construir el modelo de pronóstico, mientras que el conjunto de prueba se usa para evaluar el desempeño del modelo comparando los valores predichos con los valores reales. (Li y cols., 2020)

En el caso de las series de tiempo, el conjunto de entrenamiento correspondiente consiste solo en observaciones que ocurrieron antes de las observaciones que forman el conjunto de pruebas (Hyndman y Athanassopoulos, 2021). A diferencia de otros tipos de datos, como los conjuntos de datos independientes y aleatorios, las series de tiempo tienen una estructura temporal y dependencia entre las observaciones, lo que hace que la validación tradicional (como la validación cruzada clásica) no sea apropiada, ya que puede llevar a evaluaciones sesgadas o poco realistas del rendimiento del modelo (Bergmeir, Hyndman, y Koo, 2018).

El proceso comienza determinando el número de particiones, *slices* o *folds*. Es importante asegurarse de que el número de ventanas sea lo suficientemente grande para obtener una cantidad razonable de evaluaciones y obtener una estimación robusta del rendimiento del modelo. Un número típico de *folds* puede estar en el rango de 5 a 10 (Bergmeir y cols., 2018). En este estudio, se fijó el valor de k en 5, para que en la partición más pequeña el conjunto de entrenamiento contenga al menos 43 observaciones. Estas 5 particiones se aplicaron al conjunto de entrenamiento original, que consiste en las primeras 92 observaciones. El objetivo es llevar a cabo la validación cruzada en los 5 *folds* para discernir el algoritmo más adecuado

y luego compararlo con el el rendimiento final del modelo seleccionado en el conjunto de prueba original (Lyashenko, 2023).

En cada partición, el grupo de prueba contiene 24 observaciones, coincidiendo con el grupo de validación original. Cada grupo de prueba comienza 6 meses después del inicio del grupo de prueba anterior. Este paso se encuentra representado en la Figura 4.11. Estas configuraciones se definen en la función `time_series_cv()`. Para las visualizaciones gráficas se usó: `tk_time_series_cv_plan()` y `plot_time_series_cv_plan()` (ver apéndice A.1.9).

Para medir el desempeño de cada modelo se evaluó, tanto el valor medio como la dispersión de cada una de las métricas empleadas (MAE, MAPE y RMSE), para las subseries resultantes del proceso de partición.

Se seleccionó el modelo de pronóstico más adecuado en función de su rendimiento en términos de precisión y confiabilidad en las predicciones. Las funciones utilizadas, correspondientes al paquete `resample` (Dancho, 2021), fueron: `modeltime_fit_resamples()` para el trabajo de partición y `modeltime_resample_accuracy()` para la comparación de las métricas. Y a fin de representar el desempeño de los modelos se empleó la función `plot_modeltime_resamples()` (ver apéndice A.1.10)

3.1.6. Generación de pronósticos y evaluación

A partir el modelo de pronóstico más adecuado identificado en el paso anterior, se procedió a realizar su reajuste con el 100% de los datos. Una vez reajustado el modelo de mejor desempeño, se generaron los pronósticos futuros del precio del cordero patagónico para una temporada, es decir, 12 meses y su representación en una gráfica se realizó con la función `plot_modeltime_forecast()` del paquete `modeltime` (Dancho, 2023). Reentrenar el modelo óptimo con la totalidad de los datos es un paso crítico para garantizar una mayor precisión en las proyecciones futuras. Durante la fase de calibración y selección del modelo, se utilizó una porción de los datos para entrenar y ajustar los parámetros del modelo. Sin embargo, al reservar una parte de los datos para la validación cruzada, se restrin-

gió parcialmente la cantidad de información disponible para el proceso de aprendizaje del modelo.

Reajustar el modelo con el conjunto completo de datos, se aprovecha toda la información histórica, patrones y tendencias que existen en la serie de tiempo. Esto permite al modelo comprender mejor las relaciones complejas y las interacciones que pueden estar presentes en los datos. Al tener una visión más completa y detallada del comportamiento pasado de la serie, el modelo tiene una base sólida para realizar proyecciones más precisas y confiables hacia el futuro.

La función `modeltime_forecast()`, por defecto construye un intervalo de pronóstico utilizando un nivel de confianza del 95%. Esto implica que se espera que el valor real esté dentro de ese intervalo con una confianza del 95% (Dancho, 2023). El mencionado intervalo se calcula tomando en cuenta la variabilidad observada en los datos de calibración durante el proceso de validación cruzada utilizado durante la calibración del modelo. En consecuencia, el intervalo de pronóstico será constante para todos los valores comprendidos en el horizonte de pronóstico.

Por último, se volvió a los datos a su escala original para facilitar su uso y comunicación.

3.1.7. Esquema del Flujo de Trabajo o *Workflow*

La metodología de pronóstico propuesta en esta tesis se estructura a través de un flujo de trabajo que se encuentra detallada en la Figura 3.1. Inicialmente, se recopilan y preparan los datos históricos de precios, utilizando herramientas como `lubridate` y `timetk` para un manejo y transformación temporal adecuado. Luego, se configuran y ajustan los diversos modelos predictivos con `modeltime` y `tidymodels`, seleccionando las mejores versiones de cada uno a través de un proceso iterativo. La validación cruzada, realizada con `modeltime.resample`, es clave en este proceso, proporcionando una evaluación rigurosa y comparativa del rendimiento de los modelos. El algoritmo más adecuado se elige basado en su precisión y estabilidad, resultando en un pronóstico fiable y actualizado. Este

enfoque integral y dinámico no solo subraya la adaptabilidad y precisión de los modelos elegidos sino que también establece una práctica de reevaluación continua, esencial para adaptarse a las dinámicas cambiantes del mercado.

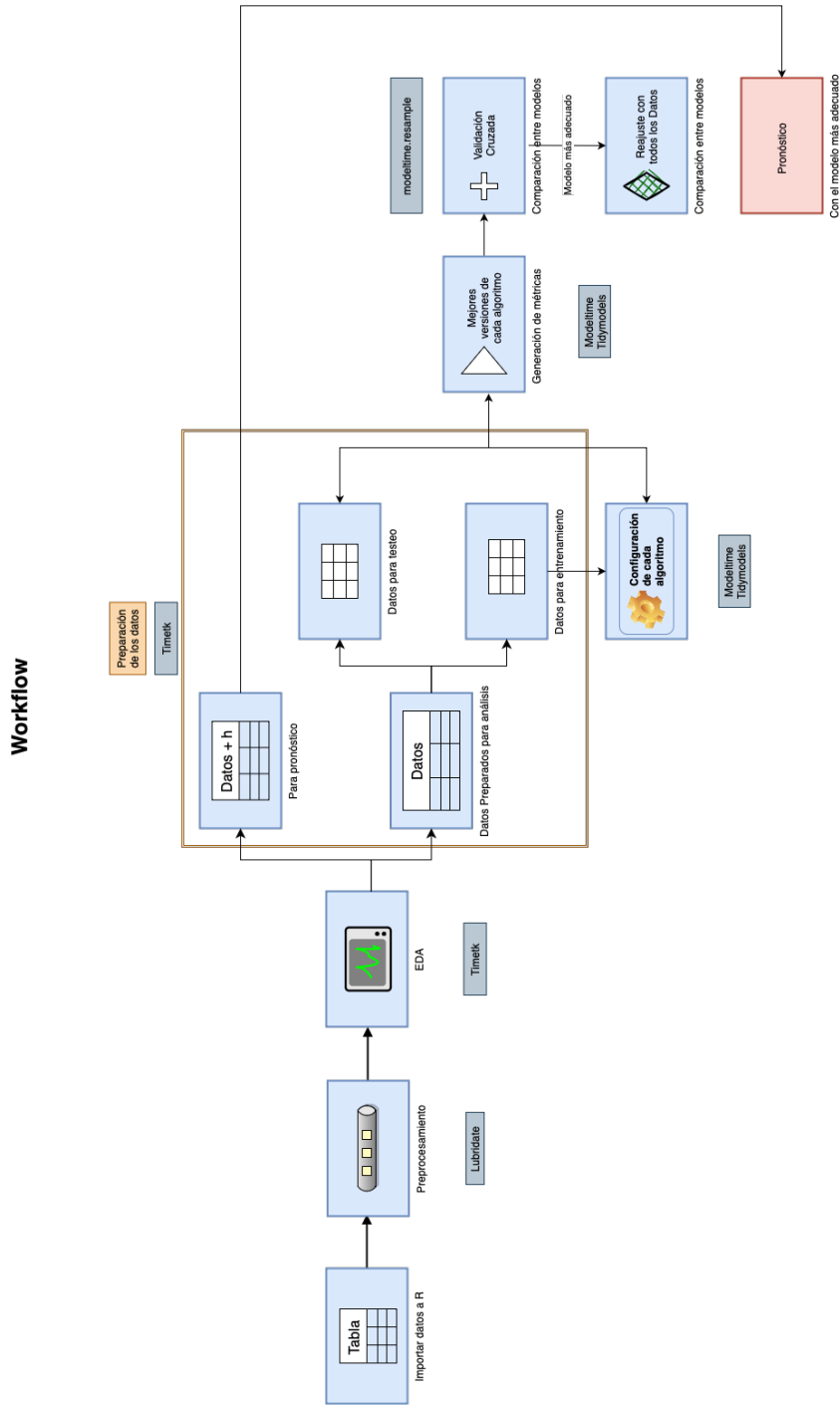


Figura 3.1 : Flujo de Trabajo para la Metodología de Pronóstico de Precio aplicada al Cordero Patagónico

CAPÍTULO 4

Resultados

En esta sección se presenta los resultados obtenidos en base a los objetivos propuestos en la presente tesis. Estos se centraron en el desarrollo de una metodología de pronóstico adecuada para el precio del cordero patagónico, que incluye la identificación de patrones estacionales y tendencias, la evaluación de diferentes modelos de pronóstico, el desarrollo de un marco metodológico para la construcción y ajuste de los modelos que incluye la aplicación de la validación cruzada para la comparación y selección del modelo óptimo, y finalmente, el reajuste y pronóstico a futuro.

4.1. Análisis Exploratorio de los Datos

La variable de interés de esta investigación, el precio promedio del cordero patagónico, *PPCord*, mostró una tendencia exponencial ascendente, como se aprecia en la Figura 4.1, lo que indica que el precio del cordero patagónico ha experimentado aumentos interanuales con una tasa creciente a lo largo del tiempo.

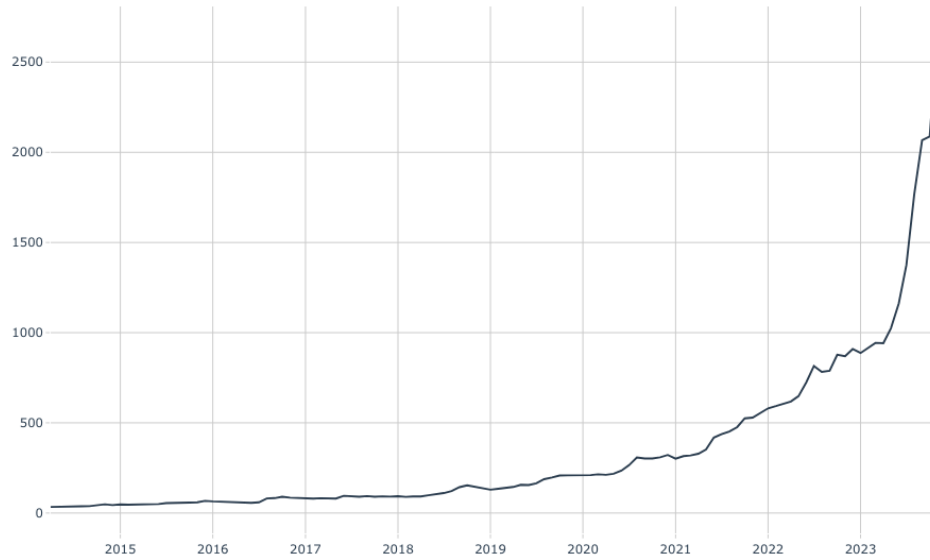


Figura 4.1: Precio Promedio Nominal Mensual del Cordero Patagónico desde Abril 2014 hasta Noviembre 2023.



Figura 4.2: Logaritmo del Precio Promedio Nominal Mensual del Cordero Patagónico desde Abril 2014 hasta Noviembre 2023.

4.1. ANÁLISIS EXPLORATORIO DE LOS DATOS

Para un EDA, la Figura 4.1, es un punto de partida útil, ya que proporciona una visualización clara de la evolución de los precios a lo largo del tiempo. La tendencia exponencial observada justifica una transformación logarítmica para estabilizar la varianza y linealizar la serie, lo que puede simplificar el análisis y la aplicación de modelos de pronóstico. En base a esto, se realizó la transformación logarítmica. La Figura 4.2 ilustra la serie de tiempo del logaritmo del precio promedio mensual del cordero patagónico en el período bajo estudio. La serie de la variable transformada, *Log_PP_Cord*, muestra un crecimiento más constante y lineal.

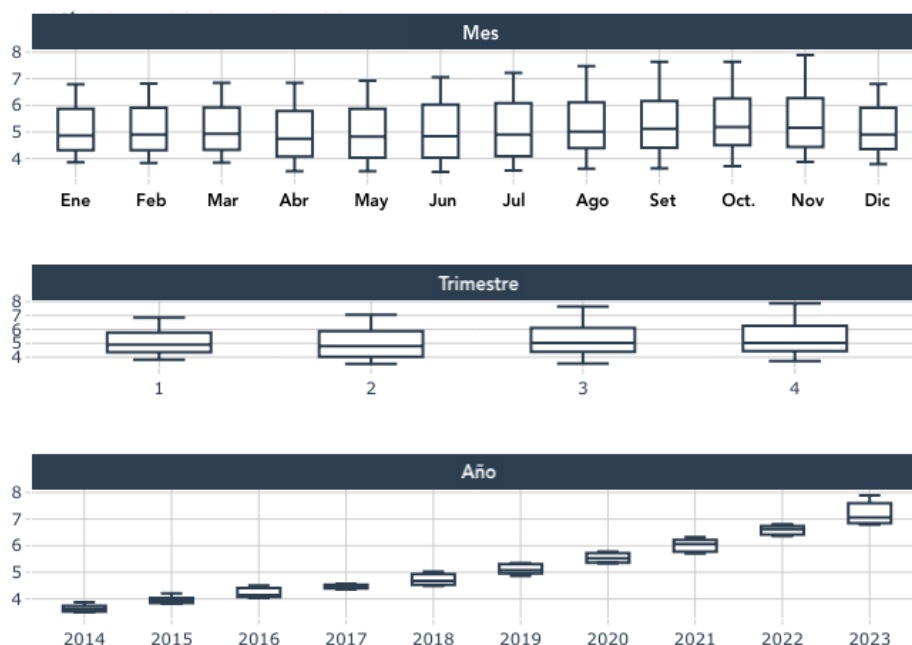


Figura 4.3: Estacionalidad Mensual del Logaritmo del Precio Promedio del Cordero Patagónico

Como siguiente paso, se estudió la estacionalidad del *Log_PP_Cord*. Dicho análisis se planteó mediante el uso de diagramas de caja. En la Figura 4.3 se puede observar las características estadísticas clave de los datos en tres escalas de tiempo: mensual, trimestral y anual. Los diagramas de caja

mensuales revelan la estacionalidad inherente a los precios relacionados con la época de servicios, fecha de parición de las ovejas y la demanda (ver Figura 1.1). Estos resultados concuerdan con los antecedentes considerados, donde se informa que el ciclo de producción de corderos en la región patagónica influye en los precios (Goizueta y Iglesias, 2013). Los precios son mas altos en los meses estivales y disminuyen durante los meses fríos. Además, la variabilidad de los precios se mantiene relativamente constante a través de las diferentes escalas de tiempo examinadas. En la evaluación anual, se identifica una tendencia creciente interanual, lo que refleja un incremento sostenido en los precios a lo largo de los años analizados.

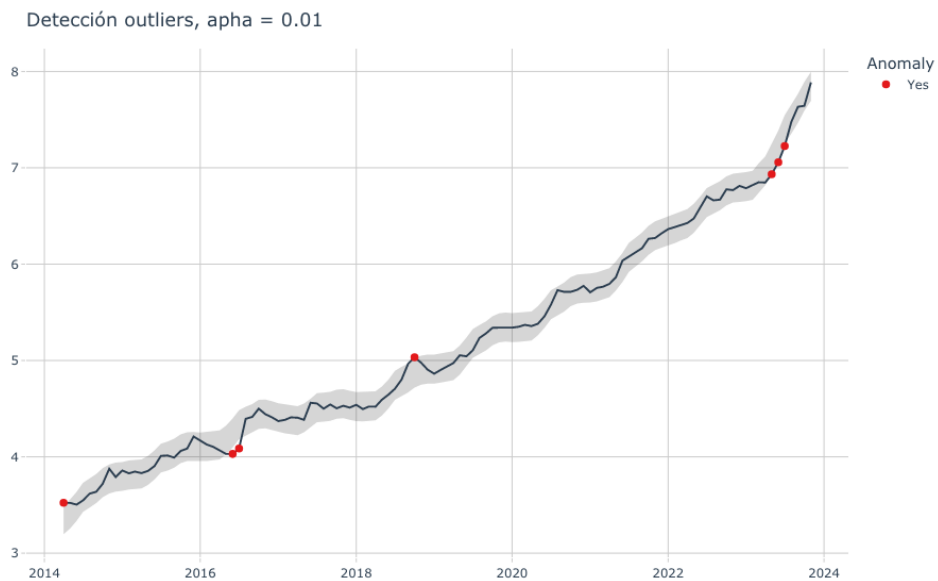


Figura 4.4: Detección de Anomalías en el Logaritmo del Precio Promedio del Cordero Patagónico

Mediante la Figura 4.4, se planteó el estudio de datos anómalos en la serie. Es visible que las anomalías detectadas son mínimas y se encuentran próximas a la banda de confianza establecida que, para identificar los valores atípicos, considera aquellos datos que caen por debajo del cuartil

4.1. ANÁLISIS EXPLORATORIO DE LOS DATOS

uno (Q1) o por encima del cuartil tres (Q3) en una magnitud equivalente a tres veces el rango intercuartílico. Este hallazgo indica que la presencia de datos anómalos en la serie de tiempo es limitada y no representan desviaciones significativas con respecto a la tendencia general. Este comportamiento refuerza la consistencia y calidad de los datos del *Log_PP_Cord* y sugiere que los modelos predictivos no requerirán una adaptación extensiva para corregir por la presencia de valores atípicos, simplificando así la metodología de pronóstico.

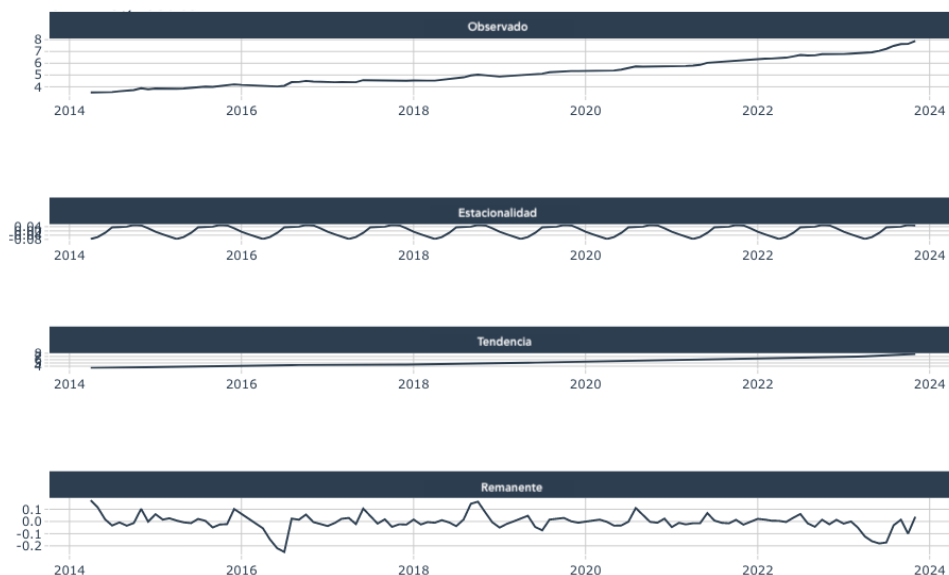


Figura 4.5: Descomposición STL del Logaritmo del Precio Promedio del Cordero Patagónico

Otro diagnóstico visual valioso es el relacionado con la descomposición de la serie bajo estudio. La Figura 4.5 presenta la descomposición STL (*Seasonal and Trend decomposition using Loess*) de la serie temporal del logaritmo del precio promedio del cordero patagónico, revelando las dinámicas subyacentes de la serie. Estas se muestran en los cuatro paneles de la figura 4.5:

1. Observado: Representa la serie temporal original del logaritmo del precio promedio.
2. Estacionalidad: Muestra el componente estacional extraído de la serie, indicando patrones periódicos dentro de cada año.
3. Tendencia: Ilustra el componente de tendencia a largo plazo, resaltando cómo evolucionan los precios a lo largo del tiempo.
4. Remanente: Contiene los residuos de la serie temporal una vez eliminados los efectos estacionales y de tendencia, lo cual puede incluir el ruido aleatorio y cualquier otra variabilidad no explicada por los componentes estacionales y de tendencia.

La descomposición se ha realizado asumiendo una estacionalidad mensual. El componente estacional muestra claros patrones dentro del año, lo que corrobora la presencia de la estacionalidad identificada previamente a través de los diagramas de caja. La tendencia indica un crecimiento sostenido a lo largo del tiempo, lo que sugiere un aumento continuo del *Log_PP*Cord. Los residuos, que son relativamente pequeños y aleatorios, sugieren que la estacionalidad y la tendencia han capturado adecuadamente la mayor parte de la estructura informativa de la serie, y que la variabilidad restante es mínima y aleatoria.

Finalmente, la Figura 4.6 muestra los correlogramas de la función de autocorrelación y la función de autocorrelación parcial del *Log_PP*Cord. La principal característica del correlograma de la serie del *Log_PP*Prom, es que los coeficientes de autocorrelación (FAC), inician con valores altos y disminuye lentamente conforme se prolonga el rezago, un comportamiento característico de las series no estacionaria. La línea de puntos muestra los límites de significación de ruido blanco. La función de autocorrelación parcial, por su parte, muestra un corte después del primer retraso, lo cual podría indicar un modelo AR(1) si se diferencia la serie, para abordar la no estacionariedad.

El análisis exploratorio de datos realizado proporciona una comprensión profunda de las características subyacentes de la serie temporal del

4.2. LA SELECCIÓN DE LOS MODELOS

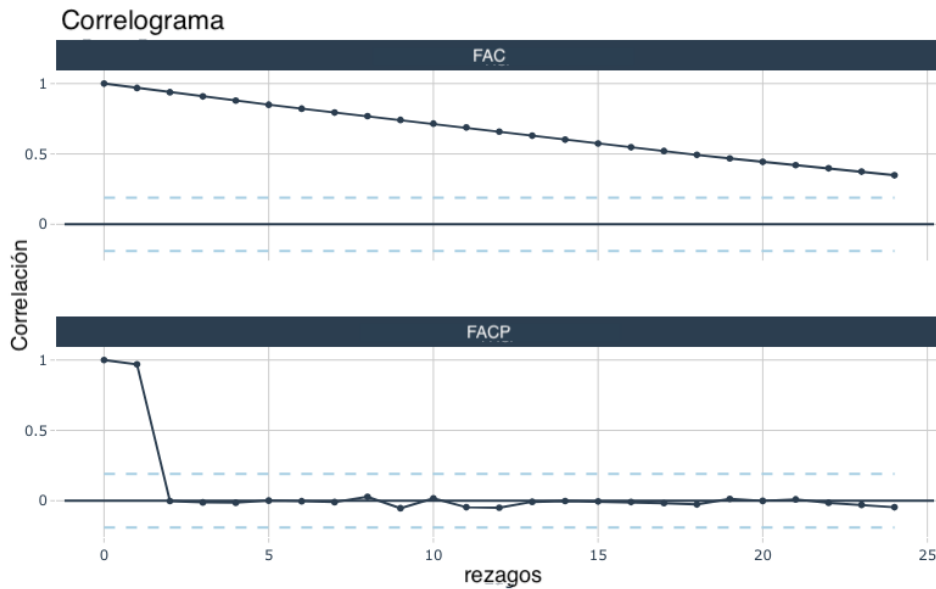


Figura 4.6: Correlograma de la Función de Autocorrelación y Autocorrelación Parcial del Logaritmo del Precio Promedio del Cordero Patagónico

logaritmo del precio promedio del cordero patagónico. Los patrones estacionales, la tendencia creciente y la ausencia de anomalías significativas resaltan la adecuación de los datos para el modelado predictivo. La presencia de autocorrelaciones decrecientes y el corte en la función de autocorrelación parcial después del primer retraso sugieren la aplicabilidad de un modelo AR(1) tras diferenciar la serie. Estos hallazgos son fundamentales para asegurar que la serie de tiempo esté preparada adecuadamente para el análisis y el pronóstico. La información obtenida establece una base sólida para la selección de modelos estadísticos apropiados, cuyos resultados se analizarán en la sección siguiente.

4.2. La selección de los modelos

En la fase inicial de selección de modelos, se realizó un proceso de exploración y evaluación de diversas configuraciones para cada uno de

los algoritmos considerados: Suavizamiento Exponencial (ETS), ARIMA y PROPHET. Este procedimiento crítico permitió identificar las configuraciones más prometedoras para cada tipo de modelo.

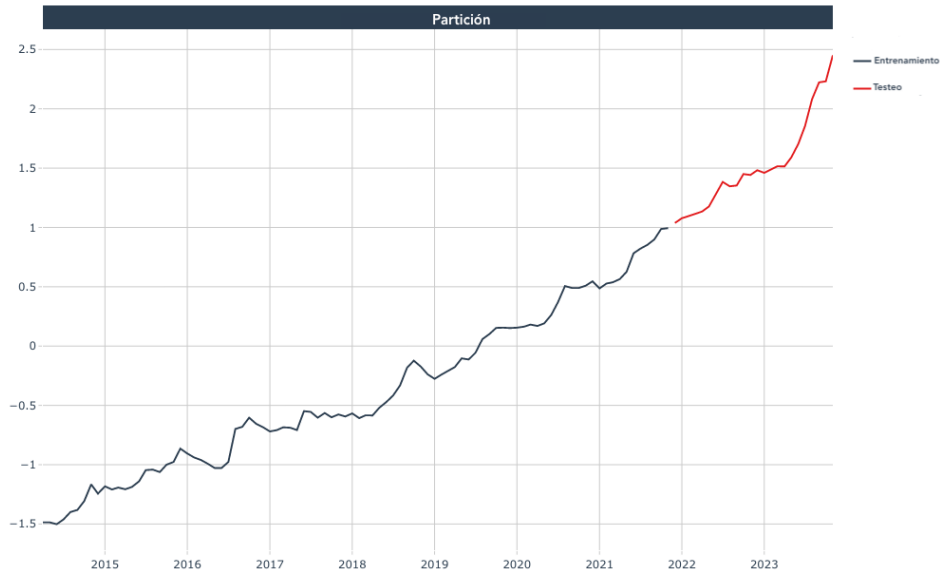


Figura 4.7: Partición de la serie en grupo de entrenamiento y grupo de testeo

4.2.1. Ajuste del Modelo de Suavizamiento Exponencial

Modelo	MAE	MAPE	RMSE
ETS(A,N,N)	0.564	33.5	0.682
ETS(A,A,N)	0.194	10.6	0.291
ETS(A,A,A)	0.182	10.7	0.244

Cuadro 4.1: Comparación de Métricas de Rendimiento en el Grupo de Testeo para los Modelos ETS

La Figura 4.8 y el cuadro 4.1 muestran los resultados de la evaluación de tres configuraciones del modelo ETS. Se han probado las variantes

4.2. LA SELECCIÓN DE LOS MODELOS

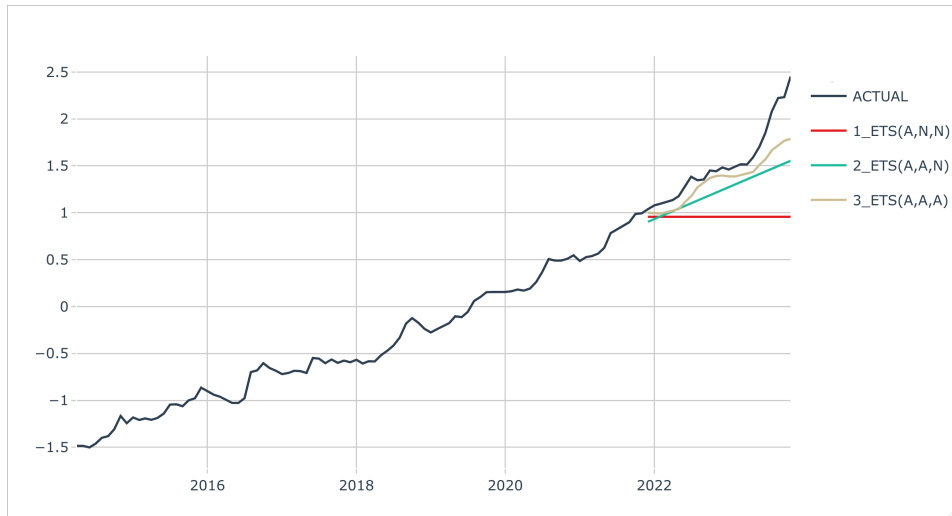


Figura 4.8: Comparación de Pronósticos de Modelos ETS para el Logaritmo del Precio Promedio del Cordero Patagónico

clásicas de suavizamiento exponencial, así como las extensiones que incorporan tendencia y estacionalidad. Las métricas resultantes de este proceso de selección, presentadas en el cuadro 4.1, indican que el modelo $ETS(A,A,A)$, con tendencia y estacionalidad, supera a sus contrapartes en términos de precisión de pronóstico, con los valores más bajos en **MAE**, **MAPE** y **RMSE**. La evaluación se llevó a cabo utilizando únicamente el conjunto de testeo, asegurando así que el rendimiento medido refleja la capacidad del modelo para generalizar a datos nuevos. Por lo tanto, el modelo $ETS(A,A,A)$ se identifica como la mejor opción dentro de la familia de modelos ETS para el pronóstico del precio del cordero patagónico.

4.2.2. Ajuste del Modelo ARIMA

La evaluación del rendimiento de los modelos **ARIMA** seleccionados se presenta en el Cuadro 4.2. El modelo $ARIMA(0,1,0)$ con deriva¹, aun-

¹La denominación “con deriva” se refiere a la inclusión de un término constante en la ecuación del modelo y significa que la serie temporal aumenta (o disminuye) de forma constante en el tiempo Gujarati y Porter, 2011, pág 743.

que proporcionó resultados aceptables, fue superado por el modelo SARIMA: ARIMA(1,1,0)(1,1,0)[12], el cual demostró una mayor precisión en la predicción de los precios del cordero patagónico. Como lo evidencian las métricas de rendimiento, el modelo SARIMA logró un menor error absoluto medio (MAE), un menor error porcentual absoluto medio (MAPE), y un menor error cuadrático medio (RMSE). Estos resultados subrayan la capacidad del modelo SARIMA para capturar la estacionalidad inherente de la serie y proveer pronósticos más cercanos a los valores reales observados, lo que se refleja en un ajuste significativamente mejor en el conjunto de testeo.

Modelo	MAE	MAPE	RMSE
ARIMA(0,1,0) Con Deriva	0.186	10.0	0.286
ARIMA(1,1,0)(1,1,0)[12]	0.136	7.88	0.199

Cuadro 4.2: Comparación de Métricas de Rendimiento en el grupo de testeo para Modelos ARIMA y SARIMA

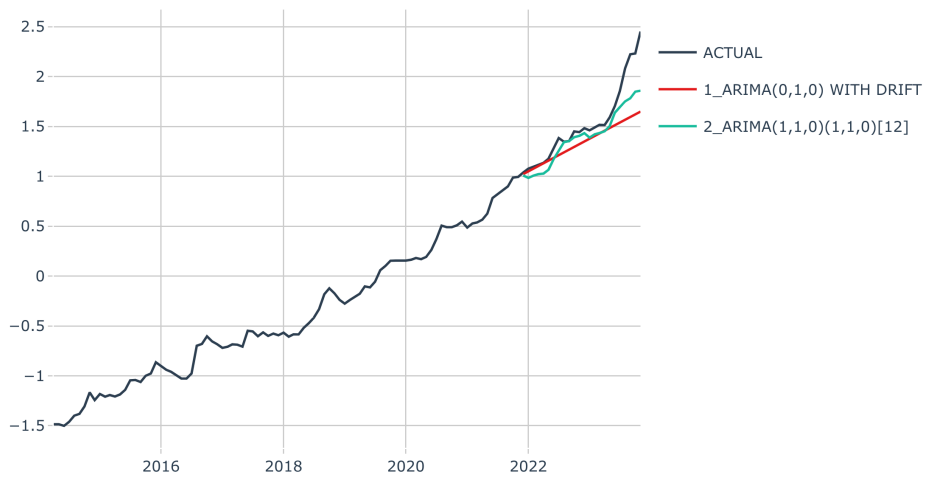


Figura 4.9: Comparación de Pronósticos del Modelo ARIMA con Deriva y Modelo SARIMA

La Figura 4.9 ilustra la comparación entre los pronósticos de los mo-

delos ARIMA(0,1,0) con deriva y el modelo ARIMA(1,1,0)(1,1,0)[12], en relación a los valores reales de la serie temporal. El modelo ARIMA(0,1,0) está representado por la línea roja y muestra una predicción lineal, indicando la inclusión de una tendencia constante pero fallando en capturar la estacionalidad mensual evidente en los datos. En contraste, el modelo ARIMA(1,1,0)(1,1,0)[12], representado por la línea verde, sigue de cerca la trayectoria real de los precios, indicada por la línea negra, ajustándose a las fluctuaciones estacionales y reflejando un entendimiento más profundo de la dinámica de la serie. Esta visualización corrobora la superioridad del modelo ARIMA(1,1,0)(1,1,0)[12] sobre el ARIMA(0,1,0) con deriva para la serie de tiempo en estudio, como también se refleja en las métricas de rendimiento del Cuadro 4.2.

4.2.3. Ajuste del Modelo PROPHET

Finalmente se implementó el modelo PROPHET, un procedimiento flexible que se adapta bien a las series temporales con estacionalidades fuertes y tendencias no lineales.

La Figura 4.10 presenta las predicciones del modelo PROPHET en los datos de testeo. Se muestra la serie original y las predicciones realizadas por dos configuraciones diferentes del modelo PROPHET. La primera configuración, indicada por la línea roja, corresponde a un ajuste estándar del modelo PROPHET, mientras que la segunda configuración, mostrada en verde, incluye ajustes en los puntos de cambio y en la estacionalidad anual para mejorar el rendimiento del modelo. Estos ajustes, se realizaron sobre el número de puntos de cambio que pasaron de 25 a 4, el rango de cambio que pasó de 0,80 a 0,90 y `prior_scale_changepoints` que pasó de 0,05 a 0,1, como se explicó en la metodología (ver Metodología 3.1). Los resultados muestran que el modelo PROPHET con ajustes ha mejorado en todas las métricas de rendimiento en comparación con la configuración estándar ver el Cuadro 4.3. El MAE disminuyó de 0.241 a 0.198, el MAPE se redujo de 14.7% a 11.9%, y el RMSE bajó de 0.293 a 0.255. Estas mejoras indican que los ajustes realizados han permitido que el modelo

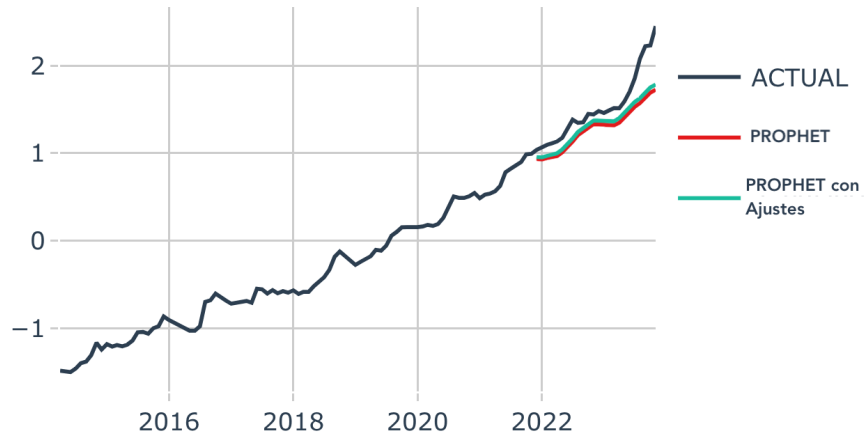


Figura 4.10: Comparación de Pronósticos del Modelo PROPHET y PROPHET con Ajustes

capte con mayor precisión las tendencias y patrones estacionales de la serie de tiempo. En resumen, las configuraciones manuales realizadas en

Modelo	MAE	MAPE	RMSE
PROPHET	0.241	14.7	0.293
PROPHET con Ajustes	0.198	11.9	0.255

Cuadro 4.3: Comparación de Métricas de Rendimiento en el grupo de tes-teo para Modelos PROPHET y PROPHET con Ajustes

el modelo PROPHET resultaron en mejoras significativas en la precisión del pronóstico. La reducción del número de `change_points` y el aumento del `range_points` y `prior_scale_changepoints`, permitieron una adaptación más adecuada del modelo a la serie de tiempo analizada, reduciendo problemas de sobreajuste e inestabilidad.

4.2.4. Validación cruzada

Posteriormente, las mejores versiones de los modelos se sometieron a un proceso de validación cruzada para series temporales a fin de determinar el modelo más adecuado con el cual realizar el pronósticos del precio del cordero patagónico.

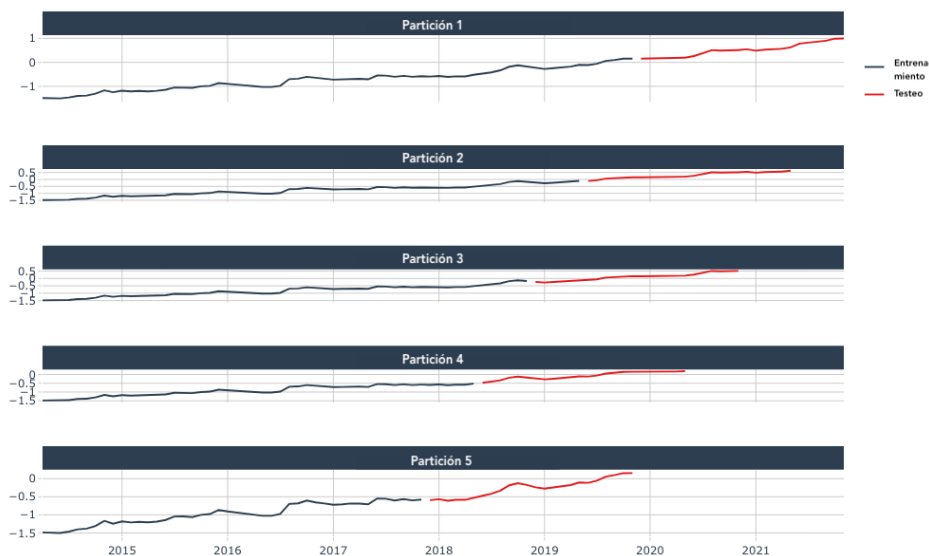


Figura 4.11: Plan de Validación Cruzada en Serie Temporal para los Modelos de Pronóstico del logaritmo del Precio Promedio del Cordero Patagónico

La Figura 4.11 ilustra el plan de validación cruzada en serie temporal aplicado a la serie de precios del cordero patagónico. Se utilizaron cinco particiones (*slices*) con una metodología de origen fijo, donde cada partición sucesiva amplía el conjunto de entrenamiento por seis meses adicionales mientras mantiene un período de prueba constante de 24 meses. Esta técnica permite evaluar la capacidad del modelo para generar pronósticos a dos años vista en diferentes momentos del tiempo, proporcionando una visión más completa de su rendimiento a lo largo de la serie temporal. Es importante mencionar que en esta fase no solo se consideró el valor

medio de cada métrica correspondiente a las 5 rondas o *lices* sino que también se tuvo en cuenta su dispersión.

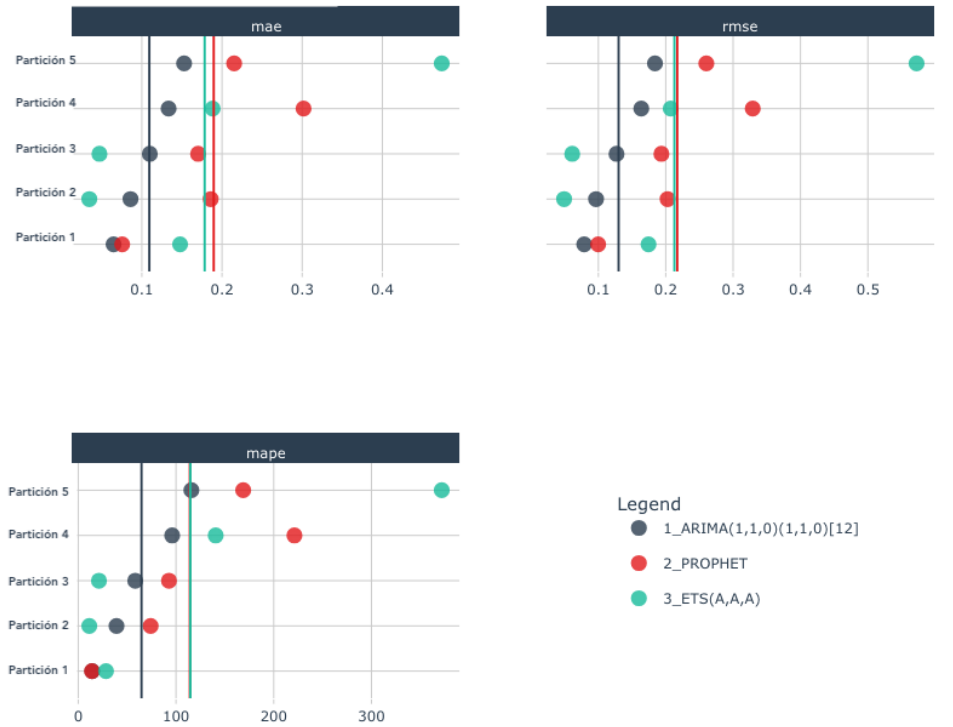


Figura 4.12: Resultados de las Métricas de Rendimiento en la Validación Cruzada en Serie Temporal del Log_PPCord

La Figura 4.12 muestra la dispersión y los valores medios de las métricas de rendimiento para los tres modelos de pronóstico en cinco rondas de validación cruzada. Cada punto representa el resultado de una partición específica, y la línea vertical representa el valor medio de cada métrica identificada por color para cada modelo. Se puede identificar claramente que el modelo ARIMA tiene un mejor rendimiento medio. La variabilidad de las métricas se observa en el alejamiento de los puntos respecto a la línea media para cada modelo en cada métrica. Los puntos más alejados indican una mayor variación en el rendimiento del modelo a través de las

4.2. LA SELECCIÓN DE LOS MODELOS

diferentes particiones de la validación cruzada, que puede sugerir menor estabilidad en las predicciones del modelo bajo diferentes condiciones de los datos.

Los resultados gráficos se acompañan con los Cuadros 4.4 y 4.5.

Modelo	MAE	MAPE	RMSE
ARIMA(1,1,0)(1,1,0)[12]	0.109	64.7	0.130
ETS(A,A,A)	0.178	115	0.213
PROPHET con Ajustes	0.190	114	0.217

Cuadro 4.4: Medias de las métricas de rendimiento en la validación cruzada

El cuadro 4.4 muestra las medias de las métricas de rendimiento para los modelos de pronóstico evaluados mediante validación cruzada en serie temporal. El modelo ARIMA(1,1,0)(1,1,0)[12] obtuvo la media más baja en todas las métricas, con un MAE de 0.191, un MAPE de 64.7 y un RMSE de 0.130. Estos valores indican que este modelo, en promedio, tiene un mejor desempeño en la predicción del precio del cordero patagónico. El cuadro 4.5 exhibe la desviación estándar de las métricas de rendimiento para los modelos en la validación cruzada. El modelo ARIMA(1,1,0)(1,1,0)[12] muestra la menor variabilidad con un desvío estándar de 0.035 en MAE, 41.3 en MAPE y 0.044 en RMSE. El modelo PROPHET Y ETS (A,A,A) tienen mayor variabilidad en todas las métricas. Estos resultados sugirieron que el modelo ARIMA(1,1,0)(1,1,0)[12] es más consistente en sus pronósticos en función de la dispersión de las métricas.

Modelo	mae	mape	rmse
ARIMA(1,1,0)(1,1,0)[12]	0.035	41.3	0.044
ETS(A,A,A)	0.178	153	0.212
PROPHET con Ajustes	0.082	81.4	0.085

Cuadro 4.5: Desvío Estándar de las métricas de rendimiento en la validación cruzada

Basándose en los valores medios y la desviación estándar de las métricas, el modelo ARIMA(1,1,0)(1,1,0)[12] parece ser el más adecuado para

el pronóstico del *Log_PP*Cord.

Por último, la Figura 4.12 muestra cómo el rendimiento de los modelos **ARIMA**, **PROPHET** y **ETS** fluctúa, a lo largo de las diferentes particiones de tiempo. Esta fluctuación resalta que la superioridad de cualquier modelo es intrínsecamente condicional y susceptible a cambios conforme se introducen nuevos datos. En particular, el modelo **ARIMA** se muestra más eficiente en las particiones más antiguas, 4 y 5, así como en la más actual, partición 1. En las particiones 2 y 3 el modelo con mejores métricas es el modelo **ETS**. De esta forma, se ilustra claramente cómo el rendimiento de cada modelo puede variar de un período a otro, lo que subraya la importancia de un enfoque de pronóstico dinámico y capaz de adaptarse a las evoluciones del mercado.

4.2.5. Pronóstico

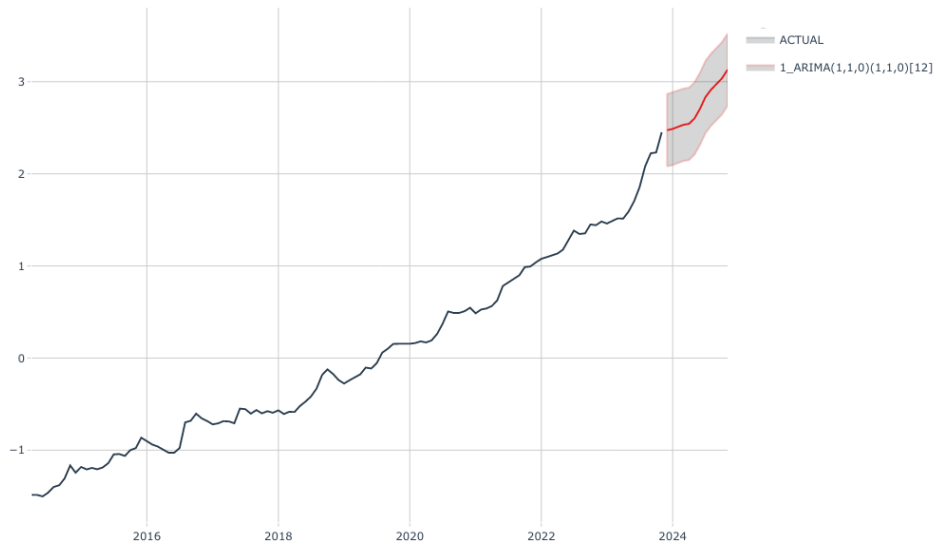


Figura 4.13: Pronóstico de 12 Meses mediante el Modelo ARIMA(1,1,0)(1,1,0)[12], Reajustado con Datos Completos

A la vista de los resultados planteados en los cuadros 4.4 y 4.5, que res-

4.2. LA SELECCIÓN DE LOS MODELOS

paldaron la opción **ARIMA**, se reajustó el modelo $ARIMA(1,1,0)(1,1,0)[12]$, con el 100 % de las observaciones y se realizó el pronóstico para una temporada de 12 meses a futuro. Reajustar el modelo $ARIMA(1,1,0)(1,1,0)[12]$ con el 100% de los datos es un paso crucial en el análisis, ya que utiliza toda la información disponible para capturar la estructura subyacente y las tendencias en los datos. Este proceso no solo mejora la precisión del modelo al incluir la información más reciente, sino que también refina las estimaciones de los componentes de tendencia y estacionalidad.

El pronóstico resultante, que proyecta 12 meses hacia el futuro, se beneficia de la mayor integridad de los datos, ofreciendo potencialmente predicciones más confiables y útiles para la toma de decisiones. La Figura 4.13 adjunta, muestra el ajuste del modelo y el intervalo de pronóstico para la trayectoria futura anticipada del modelo calibrado con la totalidad de las observaciones.

Período	Precio Promedio del Cordero Patagónico En pesos	Límite Inferior del Intervalo de Pronóstico	Límite Superior del Intervalo de Pronóstico
2023-12	2729	1764	4221
2024-01	2772	1792	4288
2024-02	2846.	1840	4402
2024-03	2913	1883	4505
2024-04	2948	1906	4560
2024-05	3141	2031	4859
2024-06	3536	2286	5470
2024-07	4066	2628	6289
2024-08	4442	2872	6872
2024-09	4782	3092	7397
2024-10	5098	3296	7886
2024-11	5653	3655	8744

Cuadro 4.6: Pronóstico del PPCord del Cordero Patagónico para una temporada de 12 meses a futuro y sus intervalos de pronóstico, en pesos.

Los resultados del pronóstico se encuentran en el Cuadro 4.6, que muestra en la primer columna el periodo a futuro y en la columna siguien-

te, el precio promedio proyectado para cada mes en la escala original de pesos nominales, es decir, luego de quitar las transformaciones aplicadas: estandarización y transformación logarítmica. La representación de la serie completa del Precio Promedio del Cordero Patagónico, junto a su pronóstico, se exhibe en la Figura 4.14.

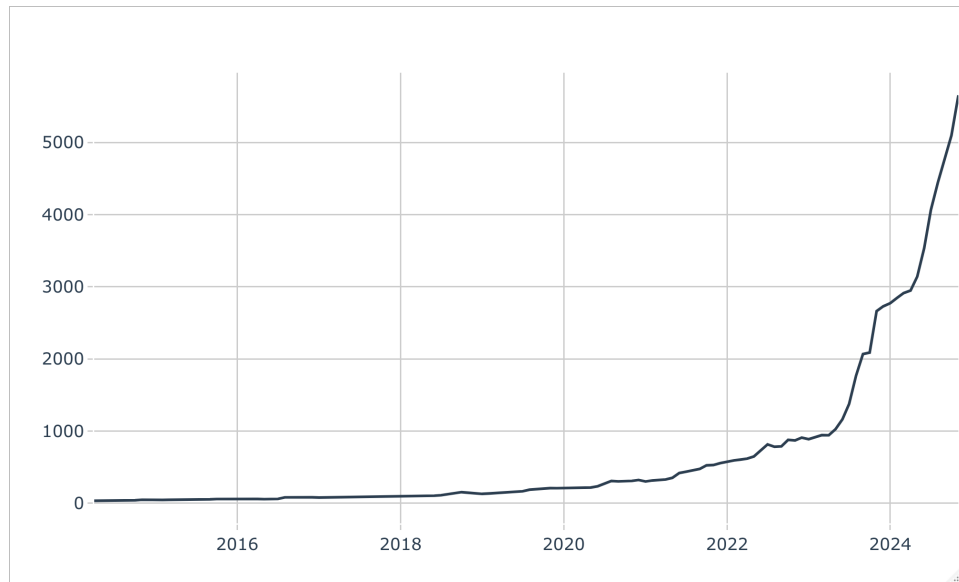


Figura 4.14: PPCord y su Pronóstico de 12 Meses mediante el Modelo ARIMA(1,1,0)(1,1,0)[12], Reajustado con Datos Completos, en pesos.

Estas estrategias metodológicas aseguraron la rigurosidad y robustez del análisis, proporcionando resultados confiables y fundamentados para pronosticar el Precio del Cordero Patagónico.

4.3. Resultados en relación al Objetivo General

El objetivo principal de esta investigación fue desarrollar una metodología de pronóstico adecuado para el precio del cordero patagónico, que permita prever con anticipación su evolución en el mercado. Para alcanzar este propósito, se establecieron y se lograron una serie de objetivos específicos que respaldaron la construcción y validación de la metodología de

4.3. RESULTADOS EN RELACIÓN AL OBJETIVO GENERAL

pronóstico.

La presente investigación se propuso desarrollar una metodología robusta y eficaz para el pronóstico del precio del cordero patagónico. Los resultados obtenidos a través de la implementación de modelos estadísticos avanzados, incluyendo el suavizamiento exponencial, modelos ARIMA y PROPHET, han demostrado una notable precisión y fiabilidad en las predicciones, basados en los valores bajos de los errores calculados en el grupo de testeo. La validación cruzada aplicada ha reforzado la robustez de estos modelos. Además, la metodología desarrollada se presenta como práctica y aplicable para los actores relevantes del sector, incluyendo productores y organismos de investigación. Por lo tanto, se ha dado cumplimiento al objetivo principal de esta tesis, al proporcionar una herramienta efectiva para anticipar la evolución de los precios en el mercado del cordero patagónico, así como a los objetivos específicos como se describe a continuación:

1. Se realizó una revisión exhaustiva de la documentación científica en el campo del pronóstico de precios, adquiriendo los fundamentos teóricos necesarios para el desarrollo del método.
2. Se llevó a cabo un análisis descriptivo de la serie histórica de precios del cordero patagónico, identificando patrones estacionales, tendencias y posibles cambios estructurales. Se destacó la ausencia de anomalías importantes, aspecto crucial para respaldar la robustez del análisis estadístico.
3. Se evaluaron y compararon diferentes modelos de pronóstico. El modelo ARIMA(1,1,0)(1,1,0)[12], se destacó como el más efectivo y confiable en términos de precisión y estabilidad, en el contexto de los datos disponibles.
4. Se aplicó una metodología de validación cruzada para series de tiempo para evaluar y comparar el rendimiento de los modelos de pronóstico, proporcionando una base sólida para seleccionar el modelo más adecuado.

5. Se desarrolló un marco metodológico sólido para la construcción y ajuste de los modelos de pronóstico seleccionados, teniendo en cuenta las características específicas de la serie de tiempo del precio del cordero patagónico. Esto garantiza pronósticos precisos y realistas.
6. Se realizó un pronóstico para una temporada futura de 12 meses, brindando proyecciones confiables y fundamentadas en la historia de la serie de tiempo.

Para concluir esta sección, se puede decir que la metodología desarrollada no solo cumplió con los objetivos propuestos, sino que también demostró ser una herramienta versátil y adaptable para la predicción de precios en mercados con series de tiempo comparables.

CAPÍTULO 5

Discusión de Resultados

En esta tesis se utilizó modelos como **ARIMA**, **ETS** y **PROPHET** para predecir el precio promedio del Cordero Patagónico. Estos modelos comúnmente utilizados sobre bienes primarios, han demostrado su utilidad en diversas aplicaciones. La revisión de la literatura revela una tendencia a abordar el pronóstico de los precios de bienes agropecuarios, mediante el uso de un modelo en particular, en general, relacionado a la familia **ARIMA**. El artículo de **Rojas y cols. (2022)** que estudia el precio de la miel, propone un método que combina modelos **ARIMA** con y sin intervención y opta por la alternativa mas adecuada en base a los valores de **ME** y **MAPE** calculados sobre un único grupo de validación. El trabajo sobre la carne de cerdo en Polonia **Zielińska-Sitkiewicz y Chrzanowska (2021)**, compara dos métodos: tendencia local y **ARIMA**, realiza evaluaciones retrospectivas, aunque carece de un enfoque sistemático de validación cruzada. El trabajo sobre el pronóstico del precio de la manzana de **Ruiz Hernández y cols. (2019)**, presenta una aplicación de la metodología Box-Jenkins, siguiendo rigurosamente sus etapas de identificación, estimación y diagnóstico. Sin embargo, al enfocarse en un solo algoritmo, no permite comparar

diferentes enfoques.

Estos antecedentes reflejan la tendencia a confiar en un conjunto particular de modelos para pronosticar precios, como en el caso del precio de la miel y la carne de cerdo. Sin embargo, a pesar de su prevalencia, estos estudios no siempre contemplan la adaptabilidad de los modelos a contextos de datos variables, lo que puede limitar su uso sistemático a largo plazo. La presente tesis argumenta la necesidad de una metodología más flexible que no solo identifique el “mejor” modelo en un conjunto de datos dado, sino que también permita adaptarse a otros conjuntos de datos, manteniendo la capacidad de identificar el modelo más adecuado en función de las circunstancias cambiantes y las dinámicas del mercado. A diferencia de estos estudios, que se comprometen en la búsqueda de un solo modelo, la metodología propuesta, implicó un enfoque más holístico, permitiendo comparar varios modelos y utilizando la validación cruzada para determinar el más adecuado. Esto refleja una evolución en la práctica del pronóstico de precios, donde la adaptabilidad y la comparación exhaustiva se vuelven esenciales para entender los patrones de comportamiento de los precios en mercados específicos como el del cordero patagónico.

Este trabajo de investigación cumple en presentar una metodología integral de pronóstico, desde el análisis inicial hasta la validación y selección del modelo óptimo, También abre la posibilidad de incorporar otros modelos que aborden contextos más complejos que el analizado en este trabajo. En este sentido, los avances en aprendizaje estadístico y aprendizaje de máquina facilitan la implementación de una amplia variedad de algoritmos para el pronóstico de series de tiempo. Según [Bergmeir y cols. \(2018\)](#), es crucial disponer de métodos robustos para identificar el modelo más adecuado. Los resultados obtenidos mediante la implementación de diversas técnicas para series de tiempo y la comparación de modelos, han demostrado que es posible aplicar una metodología eficaz para el análisis y el pronóstico de precios como es el caso del precio del cordero patagónico. Se debe mencionar la posibilidad de que el tamaño de la muestra utilizada en el estudio pueda considerarse pequeño. Si bien el término “conjunto de datos pequeño” no tiene una definición precisa, en función de los resulta-

dos encontrados por [Cerqueira, Torgo, y Soares \(2019\)](#), se puede inferir que se refiere a conjuntos de datos con alrededor de 100-150 observaciones o menos. Es importante tener en cuenta esta limitación en el estudio, ya que el rendimiento de los modelos de pronóstico puede verse afectado por el tamaño de la muestra. Aunque se han utilizado técnicas robustas como la validación cruzada para evaluar y comparar el rendimiento de los modelos, es relevante considerar que un tamaño de muestra pequeño puede limitar el alcance de los resultados. El primer paso fue caracterizar el comportamiento temporal del precio promedio mensual del cordero patagónico (*PPCord*). Como destacan [Hyndman y Athanasopoulos \(2021\)](#), esta es una fase esencial para comprender la naturaleza de la serie de tiempo, permitiendo identificar patrones y características relevantes antes de aplicar técnicas de modelado más avanzadas. Sin embargo, no todos los estudios revisados abordan esta exploración inicial. Trabajos como los de [Chan \(2020\)](#), [Fernández Lavalle \(2019\)](#) y [Lugo y Porto \(2020\)](#) no incorporan esta etapa crucial, mientras que otros investigadores, como [Menculini y cols. \(2021\)](#), [Rojas y cols. \(2022\)](#) y [Zielińska-Sitkiewicz y Chrzanowska \(2021\)](#), sacan provecho de esta etapa del análisis para asegurar que la serie de tiempo esté en condiciones apropiadas para aplicar los procedimientos de pronóstico.

[Bergmeir y cols. \(2018\)](#) plantea dos enfoques para identificar el mejor modelo: comparando el rendimiento de un modelo alternativo contra un modelo de referencia, o bien, permitiendo que los modelos “compitan” entre sí. Este último enfoque, adoptado tanto por [Menculini y cols. \(2021\)](#) como en este estudio, permite una evaluación más flexible de las opciones disponibles.

Identificadas las mejores versiones de cada modelo, se procedió a evaluarlos mediante herramientas de validación cruzada. En esta investigación se utilizó la técnica de *Validación Cruzada con Ventana Deslizante e Inicio Fijo* aunque esta no es la única opción, como describe [Bergmeir y Benítez \(2012\)](#) en su artículo. Esta elección es superior principalmente cuando es necesario el reentrenamiento periódico de los modelos, y es muy útil cuando el modelo necesita ajustarse a cambios o tendencias en los datos

a medida que avanza en el tiempo. Aunque debe señalarse que computacionalmente puede ser exigente debido al frecuente reentrenamiento de los modelos.

La práctica del reentrenamiento periódico es fundamental para mantener la relevancia y precisión de los modelos en un entorno económico en constante cambio. Es crucial reconocer las variaciones en el rendimiento que pueden surgir al utilizar diferentes modelos en competencia. Una observación clave en el diseño de la metodología aquí propuesta, es que el rendimiento de los modelos puede variar significativamente con las diferentes ventanas de pronóstico. Esta variabilidad subraya que, aunque un modelo puede ser superior en un momento dado, su efectividad puede disminuir o ser superada por otros modelos en momentos diferentes o bajo diferentes condiciones del contexto.

Hyndman y Athanasopoulos (2021), enfatizan este punto señalando que “la precisión de los pronósticos puede degradarse con el tiempo si el modelo no se actualiza regularmente” (p. 103), lo que implica la necesidad de una evaluación continua. El reentrenamiento regular de los modelos ofrece una ventaja significativa sobre enfoques más estáticos, pero esta ventaja debe ser evaluada constantemente a la luz de la variabilidad en el rendimiento de los modelos. Esta evaluación crítica y dinámica asegura que se utilice el modelo más adecuado para las condiciones vigentes en el mercado, permitiendo una mejor adaptación a los patrones cambiantes de precios y promoviendo decisiones informadas y precisas.

En la aplicación concreta de la metodología para el pronóstico del precio del cordero patagónico, el modelo $ARIMA(1,1,0)(1,1,0)[12]$ se destacó como la opción más adecuada para realizar predicciones más precisas y estables. Ya que, por un lado, este obtuvo la media más baja en todas las métricas: MAE, MAPE y RMSE y por otra parte, presentó menor variabilidad en todas estas. La efectividad del modelo ARIMA en esta tesis, donde supera a alternativas como PROPHET y ETS, subraya su eficacia como herramienta de pronóstico. Sin embargo, es importante reconocer que esta superioridad se manifiesta en el marco de los datos disponibles y en el contexto específico de los precios del cordero patagónico. La literatura pre-

via ha demostrado consistentemente la nobleza de los modelos **ARIMA** en una amplia gama de contextos de bienes primarios, como lo evidencian los estudios como el de **Rojas y cols. (2022)** en el mercado de la miel y el análisis de **Ruiz Hernández y cols. (2019)** en el mercado de la manzana. Estas investigaciones resaltan la capacidad de los modelos de la familia **ARIMA** para adaptarse y proporcionar pronósticos precisos en diversos escenarios de mercado. No obstante, es crucial considerar que la eficacia de cualquier modelo de pronóstico, incluido **ARIMA**, puede variar con la incorporación de nuevos datos o en diferentes contextos de mercado. Por lo tanto, aunque **ARIMA** haya demostrado ser el modelo más efectivo en el actual conjunto de datos para el cordero patagónico, esta superioridad debe entenderse como condicional y sujeta a cambios provocados por la evolución del mercado y la disponibilidad de nueva información. Esta comprensión enfatiza la necesidad de un enfoque dinámico en el pronóstico, donde los modelos deben ser continuamente evaluados y reentrenados para mantener su relevancia y precisión en un entorno económico fluctuante.

Un aspecto planteado por **Hyndman y Athanasopoulos (2021)** es la conveniencia de trabajar con un enfoque autorregresivo univariado, como el adoptado en esta tesis, en lugar de incorporar variables como el precio de los insumos o el valor del dólar. Esta decisión se fundamenta en varias consideraciones. Primeramente, la naturaleza intrínseca de la serie estudiada, muestra patrones y tendencias que pueden ser efectivamente capturados mediante modelos autorregresivos. Estos modelos, se centran en la información histórica de la propia serie para predecir su comportamiento futuro, aprovechando la autocorrelación presente en los datos tal cual se describe en el libro de **Gujarati y Porter (2011)**. En segundo lugar, la inclusión de variables exógenas requiere no solo la disponibilidad de datos históricos precisos y con una periodicidad adecuada, sino también un entendimiento profundo de la relación dinámica entre estas variables y el precio objetivo (**Hyndman y Athanasopoulos, 2021**). Dicha complejidad aumenta el riesgo de inflar los errores y puede complicar la interpretación de los resultados, especialmente en contextos donde las interacciones económicas son volátiles y pueden estar sujetas a cambios abruptos debido

a factores geopolíticos, cambios en políticas comerciales, o fluctuaciones en los mercados internacionales. Por otra parte, siguiendo el principio de parsimonia, la metodología propuesta trató de privilegiar la simplicidad y robustez. Los modelos autorregresivos univariados planteados en esta tesis, si bien no son los únicos posibles, se alinean con la literatura existente, que los reconoce como adecuados para pronosticar precios de bienes primarios. Finalmente, la decisión de enfocarse en este tipo de modelos no excluye la posibilidad de integrar enfoques más complejos en el futuro, especialmente con el avance de las técnicas de aprendizaje automático y la disponibilidad de datos pertinentes.

Por último, en el camino de futuras investigaciones, se destaca la necesidad de explorar nuevos algoritmos, más modernos y potencialmente más eficientes, para mejorar aún más los pronósticos, especialmente en series de tiempo con características más complejas que el Precio Promedio del Cordero Patagónico. La búsqueda constante de metodologías y técnicas de pronóstico más avanzadas no solo contribuirá a fortalecer la robustez y aplicabilidad de la metodología desarrollada, sino que también impulsará el progreso en el campo de la predicción de precios en el sector agropecuario. Este esfuerzo por mejorar las herramientas de pronóstico beneficiará no solo al mercado ganadero, sino a otros sectores del mercado agropecuario, optimizando la toma de decisiones informadas y facilitando una gestión más efectiva de los recursos y estrategias de cada mercado.

CAPÍTULO 6

Conclusiones

La presente tesis ha logrado desarrollar una metodología adecuada de pronóstico aplicada al precio del cordero patagónico. Fue diseñada para anticipar la evolución del precio en este mercado e integra eficazmente tanto técnicas estadísticas y econométricas tradicionales como enfoques más modernos vinculados al aprendizaje automático.

La investigación comenzó con un análisis exhaustivo de la serie histórica de precios del cordero patagónico, recopilada por el INTA. Este análisis detallado permitió identificar patrones estacionales, tendencias y cambios estructurales potenciales en los datos, sentando las bases para la aplicación de modelos de pronóstico más precisos y contextualizados. Posteriormente, se examinaron y evaluaron diversos modelos de pronóstico, incluyendo ARIMA y ETS, tradicionales en el campo de la econometría, así como el modelo PROPHET, más reciente y vinculado al aprendizaje automático. Este examen reveló que, si bien cada modelo tiene sus fortalezas, el ARIMA demostró ser particularmente efectivo en el contexto actual de los datos disponibles.

La implementación de una metodología de validación cruzada adap-

tada a series temporales ha sido fundamental en esta tesis para evaluar de manera integral el rendimiento de varios modelos de pronóstico: ARI-MA, PROPHET y ETS. Este proceso no solo permitió determinar cuál era el modelo más eficaz con el conjunto actual de datos, sino que también destacó la importancia crítica del reentrenamiento periódico. A través de la validación cruzada, se pudo observar cómo la eficacia de los diferentes modelos puede fluctuar con la incorporación de nuevos datos, subrayando que la superioridad de un modelo particular es condicional y puede variar con el tiempo y los cambios del mercado.

Esta comprensión llevó a este estudio a una conclusión clave: la necesidad de un enfoque dinámico y adaptable en el pronóstico de precios. La repetición periódica de la metodología propuesta, fundamentada en las enseñanzas de la validación cruzada, se propone, en consecuencia, como una práctica esencial, a fin de garantizar que los pronósticos se mantengan actualizados y alineados con las últimas tendencias y condiciones del mercado. Además, esta metodología dejó abierta la puerta a la experimentación con nuevos modelos, ampliando el espectro de herramientas disponibles para enfrentar la complejidad de las series temporales. Por lo tanto, este enfoque dinámico de validación y reentrenamiento no solo ratificó la efectividad de los modelos empleados, sino que también sentó las bases para pronósticos más confiables y precisos en el futuro, asegurando su aplicabilidad y relevancia en un entorno económico que está en constante evolución.

6.1. Sugerencias para Investigación Futura

En esta tesis, no solo se estableció una metodología sólida para el pronóstico del precio del cordero patagónico, sino que se abrieron diversas vías para investigaciones futuras que pueden ampliar y enriquecer este campo de estudio.

- Una de las áreas a explorar se relaciona al análisis de técnicas de validación cruzada alternativas para series temporales. Este estudio

6.1. SUGERENCIAS PARA INVESTIGACIÓN FUTURA

podría proporcionar resultados valiosos sobre cómo mejorar la precisión de los modelos de pronóstico, adaptándolos a las particularidades de diferentes tipos de series temporales.

- Validación cruzada aplicada a series temporales multivariadas: Investigar cómo diferentes esquemas de validación cruzada impactan en la precisión de los modelos multivariados podría arrojar luz sobre estrategias óptimas de validación en contextos más complejos.
- Redes Neuronales Recurrentes (RNNs) y Long Short-Term Memory (LSTM): Estos modelos son particularmente adecuados para el análisis de series temporales debido a su capacidad para recordar información pasada a largo plazo, lo que es crucial para predecir secuencias de datos temporales. La LSTM, una variante de las RNN, es eficaz en el manejo de dependencias a largo plazo y puede capturar patrones complejos en los datos que modelos más simples, podrían no detectar.
- Modelos Híbridos: Combinar modelos estadísticos con algoritmos de aprendizaje automático puede ofrecer un equilibrio entre la interpretabilidad de los primeros y la capacidad de modelado flexible de los segundos. Por ejemplo, un modelo híbrido que combine ARIMA para capturar la estructura lineal de la serie temporal y una red neuronal para modelar las no linealidades podría proporcionar pronósticos más precisos y robustos. Esta aproximación permite explotar las fortalezas de ambos enfoques, mejorando la precisión de los pronósticos en series temporales complejas.
- Modelos Basados en Ensemble: Los métodos de ensemble, como el Bagging y Boosting, combinan las predicciones de varios modelos para mejorar la precisión y la robustez del pronóstico. Por ejemplo, un ensemble de modelos ARIMA, LSTM y PROPHET podría aprovechar las fortalezas individuales de cada modelo, mitigando al mismo tiempo sus debilidades. Esta estrategia es particularmente valiosa en contextos con alta incertidumbre y volatilidad.

Estas líneas de investigación futuras, tienen el potencial de avanzar significativamente en el conocimiento y la aplicación del pronóstico de precios en el sector agropecuario y contribuir a la eficacia y precisión de las decisiones estratégicas en una variedad de contextos de mercado. Por ello se propone continuar con el estudio de la validación cruzada y las series temporales multivariadas interrelacionadas, como el paso natural siguiente al actual trabajo. Y así enfocar futuros esfuerzos en el uso de técnicas más avanzadas y en la aplicación de la metodología propuesta a contextos más amplios y diversos.

APÉNDICE A

Apéndice

A.1. Código de R

A.1.1. Librerías

```
1 library(tidymodels)
2 library(modeltime)
3 library(tidyverse)
4 library(lubridate)
5 library(timetk)
6 library(forecast)
7 library(ggplot2)
8 library(prophet)
9 library(rstantools)
10 library(plotly)
11 library(modeltime.resample)
```

A.1.2. Datos

```
1 data_tbl <- read_csv2("data_4_22.csv")
```

```

2
3 data_tbl <- data_tbl %>%
4   mutate(Fecha = dmy(Mes)) %>%
5   select(-Mes)
6
7 data_tbl %>% head()

```

A.1.3. EDA

```

1 data_tbl %>%
2   plot_time_series(.date_var   = Fecha,
3                   .value      = PPCord,
4                   .smooth     = FALSE,
5                   .title      = "Precio Promedio Mensual del
6                               Cordero Patagonico"
7   )
8 data_tbl %>%
9   plot_time_series(.date_var   = Fecha,
10                  .value      = log(PPCord),
11                  .smooth     = FALSE,
12                  .title      = "Log PPCord Patagonico"
13   )
14
15 data_tbl %>%
16   plot_seasonal_diagnostics(.date_var   = Fecha,
17                            .value      = log(PPCord),
18                            .title      = "Log del PPCord -
19                                    BoxPlot"
20   )
21 data_tbl %>%
22   plot_anomaly_diagnostics(.date_var   = Fecha,
23                          .value      = log(PPCord),
24                          .alpha      = 0.05,
25                          .max_anomalies = 0.2,
26                          .title      = "Deteccion de
27                                    outliers"

```

```

28
29 data_tbl %>%
30   plot_acf_diagnostics(.date_var      = Fecha,
31                       .value         = log(PPCord),
32                       .lags           = 24,
33                       .show_white_noise_bars = TRUE
34   )
35
36 data_tbl %>%
37   plot_stl_diagnostics(.date_var = Fecha,
38                       .value     = log(PPCord),
39                       .frequency = "12 months",
40                       .trend     = "1 year",
41                       .feature_set = c("observed", "season", "
42                                     trend", "remainder")

```

A.1.4. Preparación de Datos

```

1 data_transformed_tbl <- data_tbl %>%
2   mutate(log_PPCCord = log(PPCCord)) %>%
3   select(-PPCCord)
4
5 data_transformed_tbl <- data_transformed_tbl %>%
6   mutate(PPCCord_trans = standardize_vec(log_PPCCord)) %>%
7   select(-log_PPCCord)
8
9 #Standardization Parameters
10 st_mean<- 5.17
11 st_sd<- 1.109
12
13 data_transformed_tbl %>% head()
14
15 horizon <- 12
16 data_prepared_full_tbl <- data_transformed_tbl %>%
17   bind_rows(
18     future_frame(.data = ., .date_var = Fecha, .length_out =
19                   horizon)

```

```

20
21 data_prepared_full_tbl %>% tail(13) #forecasting data
22
23 data_prepared_full_tbl %>%
24   plot_time_series(Fecha, PPCord_trans, .title = "Para
      Pronosticar", .smooth = FALSE)
25
26 data_prepared_tbl <- data_prepared_full_tbl %>%
27   filter(!is.na(PPCord_trans))
28
29 data_prepared_tbl %>% tail()
30
31 forecast_tbl <- data_prepared_full_tbl %>%
32   filter(is.na(PPCord_trans))
33 forecast_tbl #fechas que se pronosticaran

```

A.1.5. Splits: Entrenamiento vs Validación

```

1 splits <- time_series_split(data_prepared_tbl,
2                             assess      = "24 months",
3                             cumulative = TRUE # para que trabaje con
4                             origen fijo
5                             )
6
7 splits
8
9 splits %>%
10  tk_time_series_cv_plan() %>%
11  plot_time_series_cv_plan(Fecha, PPCord_trans)

```

A.1.6. ETS

```

1 mod_fit_ets <- exp_smoothing(
2   error = "additive",
3   trend = "none",
4   season = "none",
5   smooth_level = 0.5
6 ) %>%

```

```
7   set_engine("ets") %>%
8   fit(PPCord_trans ~ Fecha, data = training(splits))
9
10  modeltime_table(
11    mod_fit_ets
12  ) %>%
13    modeltime_calibrate(testing(splits)) %>%
14    modeltime_forecast(
15      new_data = testing(splits),
16      actual_data = data_prepared_tbl
17    ) %>%
18    plot_modeltime_forecast(.conf_interval_show = FALSE)
19
20  mod_fit_H <- exp_smoothing(
21    error = "additive",
22    trend = "additive",
23    season = "none",
24    smooth_level = 0.5
25  ) %>%
26    set_engine("ets") %>%
27    fit(PPCord_trans ~ Fecha, data = training(splits))
28
29  modeltime_table(
30    mod_fit_ets,
31    mod_fit_H
32  ) %>%
33    modeltime_calibrate(testing(splits)) %>%
34    modeltime_forecast(
35      new_data = testing(splits),
36      actual_data = data_prepared_tbl
37    ) %>%
38    plot_modeltime_forecast(.conf_interval_show = FALSE)
39
40  mod_fit_HWS <- exp_smoothing(
41    error = "additive",
42    trend = "additive",
43    season = "additive",
44    smooth_level = 0.5
45  ) %>%
46    set_engine("ets") %>%
```

```

47 fit(PPCord_trans ~ Fecha, data = training(splits))
48
49 modeltime_table(
50   mod_fit_ets,
51   mod_fit_H,
52   mod_fit_HWS
53 ) %>%
54 modeltime_calibrate(testing(splits)) %>%
55 modeltime_forecast(
56   new_data = testing(splits),
57   actual_data = data_prepared_tbl
58 ) %>%
59 plot_modeltime_forecast(.conf_interval_show = FALSE)
60
61 modeltime_table(
62   mod_fit_ets,
63   mod_fit_H,
64   mod_fit_HWS
65 ) %>%
66 modeltime_calibrate(testing(splits)) %>%
67 modeltime_accuracy()

```

A.1.7. ARIMA

```

1 mod_fit_autoarima <- arima_reg() %>%
2   set_engine("auto_arima") %>%
3   fit(
4     PPCord_trans ~ Fecha ,
5     data = training(splits)
6   )
7
8 modeltime_table(
9   mod_fit_autoarima
10 ) %>%
11 modeltime_calibrate(testing(splits)) %>%
12 modeltime_forecast(
13   new_data = testing(splits),
14   actual_data = data_prepared_tbl
15 ) %>%

```

```

16 plot_modeltime_forecast(.conf_interval_show = FALSE)
17
18 mod_fit_arima <- arima_reg(
19   non_seasonal_ar = 1,
20   non_seasonal_differences = 1,
21   seasonal_period = 12,
22   seasonal_ar = 1,
23   seasonal_differences = 1,
24   seasonal_ma = 0
25 ) %>%
26 set_engine("arima") %>%
27 fit(PPCord_trans ~ Fecha, training(splits))
28
29 modeltime_table(
30   mod_fit_autoarima,
31   mod_fit_arima
32 ) %>%
33 modeltime_calibrate(testing(splits)) %>%
34 modeltime_forecast(
35   new_data = testing(splits),
36   actual_data = data_prepared_tbl
37 ) %>%
38 plot_modeltime_forecast(.conf_interval_show = FALSE)
39
40 modeltime_table(
41   mod_fit_autoarima,
42   mod_fit_arima
43 ) %>%
44 modeltime_calibrate(testing(splits)) %>%
45 modeltime_accuracy()

```

A.1.8. PROPHET

```

1 mod_fit_prophet <- prophet_reg() %>%
2   set_engine("prophet") %>%
3   fit(PPCord_trans ~ Fecha, data = training(splits))
4
5 modeltime_table(
6   mod_fit_prophet

```



```
7 ) %>%
8 modeltime_calibrate(new_data = testing(splits)) %>%
9 modeltime_forecast(
10   new_data = testing(splits),
11   actual_data = data_prepared_tbl
12 ) %>%
13 plot_modeltime_forecast(.conf_interval_show = FALSE)
14
15 mod_fit_prophetP <- prophet_reg(
16   changepoint_num = 4,
17   changepoint_range = 0.90,
18   seasonality_yearly = TRUE
19 ) %>%
20 set_engine("prophet") %>%
21 fit(PPCord_trans ~ Fecha,
22     data = training(splits))
23
24 modeltime_table(
25   mod_fit_prophet,
26   mod_fit_prophetP
27 ) %>%
28 modeltime_calibrate(new_data = testing(splits)) %>%
29 modeltime_forecast(
30   new_data = testing(splits),
31   actual_data = data_prepared_tbl
32 ) %>%
33 plot_modeltime_forecast(.conf_interval_show = FALSE)
34
35 calibrate_tbl <- modeltime_table(
36   mod_fit_prophet,
37   mod_fit_prophetP
38 ) %>%
39 modeltime_calibrate(testing(splits))
40
41 calibrate_tbl %>%
42 modeltime_accuracy()
```

A.1.9. Resample

```

1 resamples_tscv_lags <- time_series_cv(
2   data = training(splits),
3   cumulative = TRUE,
4   assess = "24 month",
5   skip = "6 month",
6   slice_limit = 5
7 )
8
9 resamples_tscv_lags %>%
10  tk_time_series_cv_plan() %>%
11  plot_time_series_cv_plan(Fecha, PPCord_trans)

```

A.1.10. Validación Cruzada

```

1 model_tbl <- modeltime_table(
2   mod_fit_arima,
3   mod_fit_prophetP,
4   mod_fit_HWS
5 )
6
7 model_resamples_tbl <- model_tbl %>%
8   modeltime_fit_resamples(
9     resamples = resamples_tscv_lags,
10    control = control_resamples(verbose = TRUE, allow_par =
11      TRUE)
12  )
13 model_resamples_tbl %>%
14   modeltime_resample_accuracy(metric_set = metric_set(mae,
15     mape, rmse, rsq),
16     summary_fns = list(mean = mean,
17       sd = sd)
18   ) %>%
19   arrange(mae_mean)
20 model_resamples_tbl %>%
21   plot_modeltime_resamples(

```

```

22   .metric_set = metric_set(mae, mape, rmse, rsq),
23   .point_size = 3,
24   .point_alpha = 0.8,
25   .facet_ncol = 2
26
27 )

```

A.1.11. Refit

```

1  calibration_tbl <- modeltime_table(
2    mod_fit_prophetP) %>%
3    modeltime_calibrate(testing(splits))
4
5  refit_tbl <- calibration_tbl %>%
6    modeltime_refit(data_prepared_tbl)
7
8  refit_tbl %>%
9    modeltime_forecast(
10     new_data      = forecast_tbl,
11     actual_data   = data_prepared_tbl
12   ) %>%
13   plot_modeltime_forecast()

```

A.1.12. Pronóstico en la escala original

```

1  forecast_full_tbl <- refit_tbl %>%
2    modeltime_forecast(
3     new_data      = forecast_tbl,
4     actual_data   = data_prepared_tbl
5   )
6
7  forecast_full_tbl %>% tail(13)
8
9  forecast_full_tbl %>%
10   mutate(.value = standardize_inv_vec(.value, mean = st_mean,
11     sd = st_sd)) %>%
12   mutate(.value = exp(.value)) %>%

```

```
12 tail(13) %>%
13   select(.value)
14
15 forecast_full_tbl %>%
16   mutate(.value = standardize_inv_vec(.value, mean = st_mean,
17     sd = st_sd)) %>%
17   mutate(.value = exp(.value)) %>%
18   plot_time_series(.index, .value, .smooth = FALSE)
```

Referencias

- Bergmeir, C., y Benítez, J. (2012, 05). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191, 192–213. doi: 10.1016/j.ins.2011.12.028
- Bergmeir, C., Hyndman, R. J., y Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics Data Analysis*, 120, 70-83. Descargado de <https://www.sciencedirect.com/science/article/pii/S0167947317302384> doi: <https://doi.org/10.1016/j.csda.2017.11.003>
- Cerqueira, V., Torgo, L., y Soares, C. (2019). Machine learning vs statistical methods for time series forecasting: Size matters. *arXiv preprint arXiv:1909.13316*.
- Chan, W. N. (2020). Stock market prediction for myanmar stock price: A comparative study of arima and prophet model. *Journal of Computer Applications and Research*, 1(1), 75-79.
- Dancho, M. (2021). modeltime.resample: Resampling tools for time series forecasting [Manual de software informático]. Descargado de <https://github.com/business-science/modeltime.resample> (R package version 0.2.0)
- Dancho, M. (2023). modeltime: The tidymodels extension for time series modeling [Manual de software informático]. (<https://github.com/business-science/modeltime>, <https://business-science.github.io/modeltime/>)
- Dancho, M., y Vaughan, D. (2022). timetk: A tool kit for working with time series in r [Manual de software informático]. Descargado de <https://github.com/business-science/timetk/>
- Fernández Lavalle, N. (2019). Reservas, deuda externa y precio de commodities: pronósticos para el caso argentino. Descargado de <https://repositorio.udea.edu.ar/jspui/bitstream/10908/18464/1/%5bP%5d%5bW%5d%20T.M.%20Eco.%20Fern%3a1ndez%20Lavalle%2c%20Nicol%3a1s.pdf>
- Goizueta, M., y Iglesias, D. (2013). *Análisis de la cadena de carne ovina en*

- argentina* (Estudios Socioeconómicos de los Sistemas Agroalimentarios y Agroindustriales, ISSN 1852-4605 n.º 11). Ciudad de Buenos Aires, Argentina: Instituto Nacional de Tecnología Agropecuaria (INTA).
- Grolemund, G., y Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, 40(3), 1–25. Descargado de <https://www.jstatsoft.org/v40/i03/>
- Guerrero Melgarejo, J. E. (2020). Propuesta para la implementación de modelos de proyección de compra para medicamentos y dispositivos médicos en una organización del sector salud.
- Gujarati, D. N., y Porter, D. C. (2011). *Econometría básica*. ed. Porto Alegre: AMGH.
- Hajaj, Y. (2023). *Introduction to machine learning*. Descargado de <https://www.baeldung.com/cs/machine-learning-intro>
- Hannah Clarke. (2023). *Gb lamb prices rumble on, culls strengthening*. Website: AHDB. Descargado de <https://ahdb.org.uk/news/gb-lamb-prices-rumble-on-culls-strengthening>
- Hastie, T., Tibshirani, R., y Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). Springer.
- Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20(1), 5-10. Descargado de <https://www.sciencedirect.com/science/article/pii/S0169207003001134> (Reprinted) doi: <https://doi.org/10.1016/j.ijforecast.2003.09.015>
- Hyndman, R. J., y Athanasopoulos, G. (2021). *Forecasting: Principles and practice. 3rd edition*. OText.com, Melbourne, Australia.
- IFOA Data Science Research. (2023). *Comparison of machine learning methods*. IFOA Data Science Research. Descargado de <https://ifoatascienceresearch.github.io/tutorial/comparison/>
- Iglesias, D., y Ghezan, G. (2010). Análisis de la cadena de la carne ovina en argentina. *Estudios socioeconómicos de los sistemas agroalimentarios y agroindustriales*(5).
- INDEC. (2022). *Anuario estadístico de la república argentina 2022* (Vol. 37;

- S. T. M. S. V. Marco Lavagna Pedro Ignacio Lines, Ed.). Instituto Nacional de Estadística y Censos (INDEC).
- James, G., Witten, D., Hastie, T., y Tibshirani, R. (2021). *An introduction to statistical learning with applications in r* (Second Edition ed.). Springer.
- Kuhn, M., Vaughan, D., y Wickham, H. (2021). tidymodels: Tidy machine learning in r [Manual de software informático]. Descargado de <https://www.tidymodels.org/> (R package version 0.1.3)
- Lauria, J., y Uccelli, J. L. (2022, Febrero). *Informe del sector ovino de argentina* (Informe de mercado n.º 0). Argentina: JL Square Consulting.
- Lauria, J., y Uccelli, J. L. (2023, Marzo). *Tiempos de ovinos* (Informe de mercado n.º 13). Argentina: JL Square Consulting.
- Li, H., Cui, Y., Wang, S., Liu, J., Qin, J., y Yang, Y. (2020, 01). Multivariate financial time-series prediction with certified robustness. Descargado de <https://scite.ai/reports/10.1109/access.2020.3001287>
- Lugo, J. M. C., y Porto, M. L. (2020). *Análisis estacional y de tendencia de los precios de cerdo en pie en México 2000-2019*.
- Lyashenko, V. (2023). *Cross-validation in machine learning: How to do it right*. Descargado 2023-07-30, de <https://neptune.ai/blog/cross-validation-in-machine-learning-how-to-do-it-right>
- MAGyP. (2021). *Estudio de mercado de carne ovina* (Inf. Téc.). Ciudad de Buenos Aires, Argentina: Ministerio de Agricultura Ganadería y Pesca.
- Maino, G. H., y Martinez, L. A. (1980). La empresa agropecuaria.
- Menculini, L., Marini, A., Massimiliano Proietti, A. G., Bozza, A., Moretti, C., y Marconi, M. (2021). Comparing prophet and deep learning to arima in forecasting wholesale food prices. *Forecasting, MDPI*(3), 644-662. Descargado de <https://doi.org/10.3390/forecast3030040>
- Murpratiwi, A. A., S. I. (2021). Accuracy analysis of predictive value in transaction data of service company using combination of k-means clustering and time series methods. *J-Cosine*, 5(1), 30–39. Descargado de <https://doi.org/10.29303/jcosine.v5i1.378>

- Niaki, S. T. A., Seifi, F., y Azizi, M. J. (2020, jul). Time series forecasting of bitcoin price based on autoregressive integrated moving average and machine learning approaches. Descargado de <https://scite.ai/reports/10.5829/ije.2020.33.07a.16>
- Paldam, M. (2021). Methods used in economic research: An empirical study of trends and levels. *Economics*, 15(1), 28–42. Descargado 2023-07-13, de <https://doi.org/10.1515/econ-2021-0003> doi:doi:10.1515/econ-2021-0003
- Parra, F. (2019). *Estadística y machine learning con r*. Descargado de <https://bookdown.org/content/2274/series-temporales.html>
- Peralta, T., y Darwin, J. (2019). Pronósticos financieros como herramienta para la toma de decisiones en proyectos de inversión de empresas de producción..
- R Core Team. (2023). R version 4.2.1: A language and environment for statistical computing [Manual de software informático]. Vienna, Austria. Descargado de <https://www.R-project.org/>
- Res 326/2014 de la sagyp. (2014). Documento oficial. Descargado de <https://www.argentina.gob.ar/normativa/nacional/resoluci326-2014-235186/texto> (Publicado en el sitio web de la Secretaría de Agricultura, Ganadería y Pesca)
- Rojas, S. L., Sánchez, R. C. G., Mata, R. G., Coronado, O. A. A., y Valverde, B. R. (2022). Modelo de intervención y pronóstico de precios pagados al productor de la miel de abeja (*apis mellifera l.*) en México. *Agrociencia*, 56(3), 638–668.
- Ruiz, E. G. (2019, 29 de julio). *El cordero patagónico es ecológico, magro y nutritivo, pero en Argentina el consumo es bajísimo*. Descargado de <https://www.infocampo.com.ar/el-cordero-patagonico-es-ecologico-magro-y-nutritivo-pero-en-argentina-el-consumo-es-bajisimo/>
- Ruiz Hernández, J. A., Barrios Puente, G., y Gómez Gómez, A. A. (2019). Análisis del precio de la manzana mediante un modelo sarima. *Revista mexicana de ciencias agrícolas*, 10(2), 225–237.
- SENASA. (2022). *Caracterización de existencias ovinas*. Descarga-

do de https://www.argentina.gob.ar/sites/default/files/110_2-caracterizacion_ovinos_marzo_2022.pdf

- Shumway, R. H., y Stoffer, D. S. (2017). *Time series analysis and its applications: With r examples*. Springer.
- Taylor, S. J., y Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1), 37-45. Descargado de <https://doi.org/10.1080/00031305.2017.1380080> doi: 10.1080/00031305.2017.1380080
- Tejada, E. O., y et al. (2023). Informe de precios de carne y ganado de la patagonia.
- Tom Price. (2023). *Lamb prices supported into easter*. Website: AHDB. Descargado de <https://ahdb.org.uk/news/lamb-prices-supported-into-easter>
- Wickham, H. (2021). tidyverse: Easily install and load the 'tidyverse' [Manual de software informático]. Descargado de <https://tidyverse.tidyverse.org/> (R package version 1.3.1)
- Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, 6(3), 324–342.
- Wooldridge, J. M. (2010). Introducción a la econometría. un enfoque moderno, 4a. edición. *Journal of Chemical Information and Modeling*, 53.
- Zielińska-Sitkiewicz, M., y Chrzanowska, M. (2021). Prediction of pork meat prices by selected methods as an element supporting the decision-making process. *Operations Research and Decisions*, 31(3).

Siglas

AE Aprendizaje Estadístico 18, 21–23, 40

ARIMA Autoregresivo Integrado de Medias Móviles 22–28, 30, 31, 47, 49, 56, 58, 60, 66, 82, 83, 90, 91, 95, 98, 99

CNN Convolutional Neural Network 27

EDA Exploratory Data Analysis 45, 67, 77

ETS Exponential Time Smoothing 30, 31, 43, 47, 58, 60, 66, 82, 90, 95

INTA Instituto Nacional de Tecnología Agropecuaria 19, 24, 59, 60

LSTM Long Short Term Memory 27

MAE Mean Absolute Error 24, 56, 57, 68, 70, 82–84, 86, 98

MAPE Mean Absolute Percentage Error 23, 24, 27, 56, 57, 68, 70, 82–84, 86, 95, 98

ME Mean Error 24, 56, 95

MPE Mean Percentage Error 24

RMSE Root Mean Squared Error 24, 27, 56, 57, 68, 70, 82–84, 86, 98

VAR Vector Auto-Regressive 26

VEC Vector Error Correction 26, 27

[type=acronym]