

UNIVERSIDAD TÉCNICA DE AMBATO



FACULTAD DE INGENIERÍA EN SISTEMAS,
ELECTRÓNICA E INDUSTRIAL

MAESTRÍA EN MATEMÁTICA APLICADA

Tema: Modelos estocásticos para el estudio y predicción de
índices de precios de viviendas en el Ecuador

Trabajo de titulación previo a la obtención del grado académico
de Magíster en Matemática Aplicada

Modalidad de titulación: Proyecto de desarrollo

Autora: Ing. Mayra Alexandra Viscaíno Cuzco, Mg.

Director: Ing. Ernesto Antonio Ponsot Balaguer, Ph.D.

Ambato - Ecuador
2022

APROBACIÓN DEL TRABAJO DE TITULACIÓN

A la Unidad Académica de Titulación de la Facultad de Ingeniería en Sistemas, Electrónica e Industrial.

El Tribunal receptor de la defensa del Trabajo de Titulación presidido por la Ingeniera Elsa Pilar Urrutia Urrutia Magíster, e integrado por los señores: Ingeniero Marcelo Vladimir García Sánchez, Ph.D. e Ingeniera Clara Augusta Sánchez Benítez, Magíster, designados por la Unidad Académica de Titulación de la Facultad de Ingeniería en Sistemas, Electrónica e Industrial de la Universidad Técnica de Ambato, para receptor el Trabajo de Titulación con el tema: “Modelos estocásticos para el estudio y predicción de índices de precios de viviendas en el Ecuador ”, elaborado y presentado por la señorita Ingeniera Mayra Alexandra Viscaíno Cuzco, para optar por el Grado Académico de Magíster en Matemática Aplicada; una vez escuchada la defensa oral del Trabajo de Titulación el Tribunal aprueba y remite el trabajo para uso y custodia en las bibliotecas de la Universidad Técnica de Ambato.

Ing. Elsa Pilar Urrutia Urrutia, Mg.
Presidente del Tribunal

Ing. Marcelo Vladimir García Sánchez, Ph.D.
Miembro del Tribunal

Ing. Clara Augusta Sánchez Benítez, Mg.
Miembro del Tribunal

AUTORÍA DEL TRABAJO DE TITULACIÓN

La responsabilidad de las opiniones, comentarios y críticas emitidas en el trabajo de titulación presentado con el tema: “Modelos estocásticos para el estudio y predicción de índices de precios de viviendas en el Ecuador”, le corresponde exclusivamente a: Ingeniera Mayra Alexandra Viscaíno Cuzco, autora; bajo la dirección del Ingeniero Ernesto Antonio Ponsot Balaguer, PhD, director del trabajo de titulación; y el patrimonio intelectual a la Universidad Técnica de Ambato.

Ing. Mayra Alexandra Viscaíno Cuzco, Mg.
AUTORA

Ing. Ernesto Antonio Ponsot Balaguer, Ph.D.
DIRECTOR

DERECHOS DE AUTOR

Autorizo a la Universidad Técnica de Ambato, para que el trabajo de titulación, sirva como un documento disponible para su lectura, consulta y procesos de investigación, según las normas de la Institución.

Cedo los derechos de mi trabajo de titulación, con fines de difusión pública, además apruebo la reproducción de este, dentro de las regulaciones de la Universidad Técnica de Ambato.

Ing. Mayra Alexandra Viscaíno Cuzco, Mg.
C.C.: 140054630-3

ÍNDICE GENERAL

Contenido

PORTADA.	i
APROBACIÓN DEL TRABAJO DE TITULACIÓN.	ii
AUTORÍA DEL TRABAJO DE TITULACIÓN.	iii
DERECHOS DE AUTOR.	iv
ÍNDICE DE TABLAS.	viii
ÍNDICE DE FIGURAS.	ix
AGRADECIMIENTO.	x
DEDICATORIA.	xi
RESUMEN EJECUTIVO.	xiii
EXECUTIVE SUMMARY.	xv
CAPÍTULO I. EL PROBLEMA DE INVESTIGACIÓN	1
1.1 Introducción	1
1.2. Justificación	2
1.3. Objetivos	3
1.3.1 Objetivo General	3
1.3.2 Objetivos Específicos	3
CAPÍTULO II. ANTECEDENTES INVESTIGATIVOS	4
2.1 Métodos	4
2.1.1. Planificación	4
2.1.2. Ejecución de la revisión	7
2.2 Resultados	8
2.2.1. PI1:¿Qué modelos se han aplicado para la predicción del ICC?	10
2.2.2. PI2:¿Qué variables explicativas se han empleado en los modelos de predicción del ICC?	12
2.2.3. PI3:¿Qué métricas se han utilizado para evaluar los modelos de predicción del ICC?	13
2.2.4. PI4:¿Cuál es el horizonte de predicción de los modelos en los estudios seleccionados?	13
2.3 Discusión	14

2.4 Conclusiones preliminares	16
2.5 Marco Teórico	16
2.5.1. Series Temporales	16
2.5.2. La media en las series temporales	16
2.5.3. Modelo Naïve	17
2.5.4. Modelo Drift	17
2.5.5. Modelo aditivo y multiplicativo	17
2.5.6. Modelo de Suavizado Exponencial Simple	18
2.5.7. Modelo Holt	18
2.5.8. Modelo Holt-Winters	19
2.5.9. Modelo Regresión Lineal	20
2.5.10. Proceso autorregresivo (AR)	23
2.5.11. Proceso de media móvil (MA)	24
2.5.12. Modelo ARMA	24
2.5.13. Modelo ARIMA	25
2.5.14. Métricas del error	26
2.1.5. Una técnica para evaluar los modelos	27
CAPÍTULO III. MARCO METODOLÓGICO	28
3.1 Ubicación	28
3.2. Equipos y materiales	28
3.3. Tipo de investigación	28
3.4. Hipótesis y pregunta de investigación	29
3.4.1. Planteamiento de la hipótesis	29
3.4.2. Planteamiento de la pregunta de investigación	29
3.5. Población o muestra	29
3.6. Recolección de información	29
3.7. Procesamiento de la información y análisis estadístico	31
3.7.1. Metodología	31
3.8. Resultados alcanzados	34
CAPÍTULO IV. RESULTADOS Y DISCUSIÓN	35
4.1. Selección de variables respuesta y predictores candidatos	35
4.2. Análisis descriptivo de las variables respuesta	36
4.2.1. Estadística descriptiva del IPCOU	36
4.2.2. Estadística descriptiva del IPCOM	39
4.3. Selección de modelos candidatos	41
4.3.1. Modelos univariados	41
4.3.2. Modelos multivariados	42
4.4. Evaluación del rendimiento de predicción de los modelos	44
4.4.1. Selección de modelos con mejor rendimiento	44
4.4.2. Generación de pronósticos para las pruebas	45
4.5. Evaluación de la capacidad predictiva de los modelos	45
4.5.1. Modelo para el IPCOU	45
4.5.2. Modelo para el IPCOM	48
4.6. Prueba de hipótesis y respuesta a pregunta de investigación	49
4.7. Generación de predicciones para el año 2023	52
4.7.1. Predicciones de valores del IPCOU	52
4.7.2. Predicciones de valores del IPCOM	54

4.8. Discusión	55
CAPÍTULO V. CONCLUSIONES Y REFERENCIAS	
BIBLIOGRÁFICAS	57
5.1. Conclusiones	57
5.2. Referencias bibliográficas	58
ANEXOS	62
Anexo 1: Extracción de información de estudios revisados.	62
Anexo 2: Cantidad de datos de las variables que intervienen en el modelado.	64
Anexo 3: Código R para prueba de modelos estocásticos.	64
Anexo 4: Código R para prueba de modelos de regresión múltiple. .	70
Anexo 5: Predicciones de variables respuesta y explicativas usadas en las pruebas.	72

ÍNDICE DE TABLAS

1	Fuentes consultadas.	5
2	Criterios para incluir o excluir estudios.	6
3	Repositorios y cadenas de búsqueda.	8
4	Estudios seleccionados.	9
5	Estudios que usaron variables exógenas para predicción del ICC	13
6	Métricas del error para evaluar modelos predictivos del ICC .	14
7	Horizontes de predicción de ICC	15
8	Variables y sus fuentes de información.	30
9	Cantidad de datos de las variables que intervienen en el modelado.	32
10	Variables y tipo de variables que intervienen en este estudio. .	35
11	Análisis de varianzas de coeficientes del modelo del IPCOU. .	43
12	Análisis de varianzas de coeficientes del modelo del IPCOM. .	44
13	Modelos univariantes seleccionados.	45
14	Evaluación de la capacidad predictiva de modelos para el IP- COU.	47
15	Resultados de RMSE y MAPE para modelos del IPCOU. . .	47
16	Evaluación de la capacidad predictiva de modelos para el IPCOM.	50
17	Resultados de RMSE y MAPE para modelos del IPCOM. . . .	50
18	Análisis de varianzas de coeficientes del modelo del IPCOU. .	51
19	Análisis de varianzas de coeficientes del modelo del IPCOM. .	52
20	Estimaciones del IPCOU hasta diciembre-2023.	53
21	Estimaciones del IPCOM hasta diciembre-2023.	54
22	Cantidad de datos de las variables que intervienen en el modelado.	64
23	Predicciones de variables respuesta y explicativas usadas en las pruebas.	73

ÍNDICE DE FIGURAS

1	Proceso de selección de los estudios.	10
2	Fuentes y cantidad de estudios seleccionados.	10
3	Año de publicación de los estudios seleccionados.	11
4	Modelos predictivos usados por los estudios.	12
5	Metodología empleada en este estudio.	33
6	Conjunto de modelos candidatos.	34
7	Serie de tiempo del IPCOU.	36
8	Gráfico de estacionalidad: Serie IPCOU	36
9	Gráfico de subserie estacional: Serie IPCOU	37
10	Función de autocorrelación del IPCOU	37
11	Matriz de diagramas de dispersión y correlaciones	38
12	Serie de tiempo del IPCOM.	39
13	Gráfico de estacionalidad: Serie IPCOM	39
14	Matriz de diagramas de dispersión y correlaciones	40
15	Gráfico de subserie estacional: Serie IPCOM	41
16	Función de autocorrelación del IPCOM	41
17	Conjunto de modelos predictores potenciales.	42
18	Comparación de las predicciones del $IPC\hat{O}U$ y la serie real.	48
19	Comparación de las predicciones del $IPC\hat{O}M$ y la serie real.	49
20	Gráfico de predicciones del IPCOU, h= 18 meses.	53
21	Gráfico de predicciones del IPCOM, h= 18 meses.	54
22	Extracción de información de estudios revisados.	63

AGRADECIMIENTO

Mi infinita gratitud:

A Dios, mi amado Salvador, por su gracia, misericordia y fidelidad, dador de
toda buena dádiva y todo don perfecto.

A mi preciosa familia, siempre el mejor refugio.

A mi buen amigo Sergio, gracias por su apoyo.

Al Dr. Ernesto Ponsot, por su valiosa colaboración.

A la empresa C&V Ingeniería Cía. Ltda. por su respaldo.

A la Universidad Técnica de Ambato por abrirme sus puertas.

Mayra A. Viscaíno C.

DEDICATORIA

Con todo mi afecto para quienes son mi mayor inspiración, Dios y mi familia: Ángel, Rosita, Edgar, Bayron, Alex, Jean, Paola, Paolita y Rafaela.

Mayra A. Viscaíno C.

UNIVERSIDAD TÉCNICA DE AMBATO

FACULTAD DE INGENIERÍA EN SISTEMAS,
ELECTRÓNICA E INDUSTRIAL

MAESTRÍA EN MATEMÁTICA APLICADA

TEMA:

MODELOS ESTOCÁSTICOS PARA EL ESTUDIO Y PREDICCIÓN DE
ÍNDICES DE PRECIOS DE VIVIENDAS EN EL ECUADOR

AUTORA: Ing. Mayra Alexandra Viscaíno Cuzco

DIRECTOR: Ing. Ernesto Antonio Ponsot Balaguer, Ph.D.

LÍNEA DE INVESTIGACIÓN:

- Tecnología de la información y sistemas de control

FECHA: 29 de septiembre de 2022

RESUMEN EJECUTIVO

La industria de la construcción tiene una influencia dinamizadora en la economía. En este campo, es conocido que los proyectos de construcción conllevan un riesgo asociado a las fluctuaciones de los precios de los materiales de construcción. Estimar valores futuros de índices de precios de la construcción es importante porque contratantes y contratistas emplean estos valores para la elaboración de presupuestos. El objetivo de esta investigación fue diseñar modelos estocásticos para explicar y predecir los índices de precios de construcción en viviendas unifamiliares (IPCOU) y multifamiliares (IPCOM) en el Ecuador. Se contempló el diseño de modelos univariantes, y multivariantes que consideran variables exógenas. Se identificó que nueve y ocho predictores resultaron estadísticamente significativos para predecir el IPCOU e IPCOM, respectivamente. Un conjunto de catorce modelos potenciales fueron sometidos a entrenamiento y validación, aplicando la técnica de validación cruzada para series de tiempo. En la etapa de pruebas resultó que el modelo Autorregresivo Integrado de Media Móvil conocido como ARIMA(1,1,2) y el modelo de regresión múltiple para predecir valores del IPCOU, y el modelo ARIMA(2,1,2) para predecir el IPCOM, tienen buena capacidad predictiva para el horizonte de pronóstico que contempla este estudio. Se generaron intervalos de predicción del IPCOU e IPCOM hasta el año 2023, con un nivel de confianza del 95%.

Palabras Claves: Índice de costos de construcción, predicción, modelos estocásticos, modelo autorregresivo integrado de media móvil, regresión lineal múltiple.

UNIVERSIDAD TÉCNICA DE AMBATO

**FACULTAD DE INGENIERÍA EN SISTEMAS,
ELECTRÓNICA E INDUSTRIAL**

MAESTRÍA EN MATEMÁTICA APLICADA

THEME:

**”STOCHASTIC MODELS FOR THE STUDY AND PREDICTION OF
HAUSING PRICE INDEXES IN ECUADOR”**

AUTHOR: Ing. Mayra Alexandra Viscaíno Cuzco

DIRECTED BY: Ing. Ernesto Antonio Ponsot Balaguer, Ph.D.

LINE OF RESEARCH:

- Information technology and control systems

DATE: 29 de septiembre de 2022

EXECUTIVE SUMMARY

The construction industry has a dynamic influence on the economy. In this field, it is known that construction projects carry a risk associated with fluctuations in the prices of construction materials. Estimating future values of construction price indices is important because contractors and builders use these values for budgeting. The aim of this research was to design stochastic models to explain and predict the construction price indices in single-family homes (IPCOU) and multi-family homes (IPCOM) in Ecuador. The design of univariate and multivariate models that consider exogenous variables was contemplated. Nine and eight predictors were found to be statistically significant in predicting the IPCOU and IPCOM, respectively. A set of fourteen potential models were subjected to training and validation, applying the cross-validation technique for time series. In the testing stage, it resulted that the Autoregressive Integrated Moving Average model known as ARIMA(1,1,2) and the multiple regression model to predict IPCOU values, and the ARIMA(2,1,2) model to predict the IPCOM, have good predictive capacity for the forecast horizon contemplated in this study. IPCOU and IPCOM prediction intervals were generated up to the year 2023, with a confidence level of 95%.

Keywords: Construction cost index, forecasting, stochastic models, autoregressive integrated moving average model, multiple linear regression.

CAPÍTULO I

EL PROBLEMA DE INVESTIGACIÓN

1.1 Introducción

El sector de la construcción es una de las fuerzas dinamizadoras de la economía [1], [2], [3], por su capacidad de estimular otras actividades económicas [4]. Para el Ecuador, es uno de los cinco sectores más importantes en la dinámica económica y que más aporta al Producto Interno Bruto (PIB) del país. Su aporte es en términos de atracción de inversión extranjera, generación de valor agregado, producción y creación de plazas de empleo [5].

La Organización Internacional de Trabajo (OIT) consideró que el sector de la construcción puede contribuir desde varias aristas a fomentar la recuperación económica luego de la COVID-19 [6]. De hecho el 26 de abril de 2020, el sector de la construcción fue uno de los primeros sectores en reactivarse [7] en el Ecuador. Para el cuarto trimestre del año 2021 el sector de la construcción fue el quinto sector más importante de entre 15, al aportar el 8,3% al Valor Agregado Bruto (VAB) [2]. Todo esto indica que la contribución de este sector es fundamental para la economía ecuatoriana.

Sin embargo, los costos de construcción son dinámicos [8], por lo que esta industria es sensible a los ciclos económicos [6]. Un indicador cuantitativo que mide las fluctuaciones en los costos de construcción es el Índice de Costos de la Construcción (ICC) [9]. Este valor indica el precio relativo de compra de bienes en comparación a un año base; y es muy usado en la toma de decisiones en las diferentes etapas de los proyectos de construcción. Por ejemplo, es usado en la preparación de presupuestos, es una base para la elaboración de cotizaciones para procesos de licitación y en los procesos de liquidación económica de proyectos [10].

En el caso de Estados Unidos, este indicador es generado mensualmente por *Engineering News-Record* (ENR), responsable de reportar el precio promedio del costo de mano de obra y de algunos materiales como: acero estructural, cemento portland y madera, de 20 ciudades [11], tomando a 1913 como año de referencia [12]. Otros países como: Vietnam [1], Nueva Zelanda [3], Egipto [8] y China [9], también generan bases de datos del ICC, que es un conjunto de datos ampliamente usado en el desarrollo de modelos de pronóstico [3], [10].

En Ecuador, el indicador que mide la evolución mensual de los precios de los insumos más importantes para la construcción local, es el Índice de Precios de la Construcción (IPCO), que considera al año 2000 como base. La entidad

responsable de emitir el IPCO es el Instituto Nacional de Estadística y Censos (INEC), en cumplimiento a lo dispuesto en el artículo 83 de la Ley Orgánica del Sistema Nacional de Contratación Pública [13]. Dentro del conjunto de indicadores que reporta, se encuentra el índice de precios de construcción de viviendas unifamiliares y multifamiliares.

Por otra parte, los proyectos constructivos conllevan riesgos asociados a los costos. Se han identificado como factores de riesgo a la subestimación de los costos de los proyectos y a la fluctuación en los precios de los materiales [14]. Es conocido que las obras de construcción pueden durar meses en ejecutarse y requieren inversiones económicas altas [15]; por lo que los cambios frecuentes en los costos de los insumos de construcción tienen un impacto relevante en la rentabilidad de los proyectos.

En este contexto, el desafío tanto de contratantes como constrictas es la elaboración de presupuestos más certeros que eviten la sobre o subestimación de costos. Este reto ha sido enfrentado con el uso de modelos de predicción del ICC que tengan un margen de error aceptable [9], [15], [16]. Por lo tanto, la predicción de valores del ICC cumple un rol clave en la gestión de costos de los proyectos de construcción. Su aporte se relaciona con la estimación de costos de construcción, la preparación del presupuesto y el control de contingencias [10]. Así, el objetivo de este trabajo es diseñar modelos estocásticos para explicar y predecir los índices de precios de la construcción de viviendas unifamiliares y multifamiliares en el Ecuador.

Este trabajo se compone de cinco capítulos, organizados de la siguiente manera: comienza contextualizando el tema y especificando los objetivos del estudio. En el segundo capítulo, se realiza un estudio de la literatura para conocer los antecedentes en el tema de pronóstico del índice de costos de construcción y se elabora un marco teórico de los modelos utilizados. En el tercer capítulo, se propone la hipótesis y la pregunta de investigación, así como una descripción de la metodología empleada para alcanzar los objetivos. En el cuarto capítulo, se presentan los resultados del trabajo, se da respuesta a la pregunta de investigación y prueba de hipótesis. Finalmente, en el capítulo cinco, se desarrollan las conclusiones y se detalla la bibliografía.

1.2. Justificación

El ICC es un indicador eficaz para capturar tendencias de variabilidad de los costos. Se ha evidenciado que el conjunto de datos del ICC es ampliamente usado para hacer predicciones de costos de construcción en varios países. Para ello se emplean diferentes técnicas de pronóstico de series temporales, sin embargo, aún no se han desarrollado modelos de pronóstico del índice de precios de construcción de viviendas unifamiliares (IPCOU) y multifamiliares (IPCOM) en el Ecuador.

Esta información es valiosa para contratantes, quienes pueden emplear los IPCO pronosticados, en la etapa de planificación de proyectos, específicamente

para la preparación de presupuestos y en los procesos de liquidación económica. En el sector público evitar la subestimación de costos es importante, particularmente porque la asignación de recursos económicos tiene estrictas restricciones. El IPCO, es usado en los contratos cuya forma de pago corresponda al sistema de precios unitarios que celebran las entidades del sector público, para ejecución de obras de infraestructura, que usan fondos del Estado. Basados en ellos, dichos contratos pueden aplicar el sistema de reajuste de precios, que en cierta medida absorbe la incertidumbre de la variabilidad de costos y reduce el riesgo que asume el oferente en la elaboración de cotizaciones para procesos de licitación.

Sin embargo, esto no sucede en el caso de la construcción de obras en el sector privado, en donde no se puede apelar al sistema de reajuste de precios y por tanto el contratista está obligado a asumir las pérdidas de rentabilidad que generen los incrementos de precios de los insumos de construcción. Los errores en la sub estimación de costos, surgen evidentemente, por el desconocimiento de los cambios en las tendencias de los costos de insumos de construcción. Por tal motivo, se considera importante el desarrollo de modelos estocásticos para explicar y predecir los índices de precios de viviendas unifamiliares y multifamiliares, que son el tipo de obras más construidas en el ámbito privado, en el Ecuador.

Los resultados de este trabajo podrán ser utilizados por los profesionales constructores, como modelos de predicción del IPCOU e IPCOM, y considerar sus resultados como un indicador de alerta a tomar en cuenta en la elaboración de sus ofertas económicas; de manera que el riesgo de pérdida de rentabilidad se reduzca en cierta medida. También, los resultados obtenidos se constituyen en una guía para estudios futuros, así como en un fundamento para investigadores interesados en esta área.

1.3. Objetivos

1.3.1. Objetivo General

Diseñar modelos estocásticos para explicar y predecir los índices de precios de viviendas en el Ecuador.

1.3.2. Objetivos Específicos

- Analizar estadísticamente la serie de tiempo de valores mensuales de índices de precios de viviendas unifamiliares y multifamiliares correspondiente al período enero-2003 a marzo-2022.
- Desarrollar modelos estocásticos de series cronológicas de valores mensuales del índice de precios de viviendas unifamiliares y multifamiliares para el periodo enero-2003 a marzo-2022.
- Predecir los índices de precios de viviendas unifamiliares y multifamiliares para el año 2023 y evaluarlos usando técnicas de aprendizaje de máquina.

CAPÍTULO II ANTECEDENTES INVESTIGATIVOS

El ICC ha sido ampliamente usado para medir las fluctuaciones futuras de los costos de construcción. Con este propósito se han desarrollado diferentes modelos que abordan este objetivo. Existen tres enfoques bien diferenciados, a saber: por un lado los modelos estadísticos que tratan datos de series de tiempo, tanto univaridos como multivariados; por otra parte los modelos que emplean algoritmos de “*Machine Learning*” (ML); y por último, otros modelos. En este contexto, realizar una revisión de la literatura (RL) existente, sobre los modelos que explican y predicen valores del ICC, permitirá:

- Determinar los modelos aplicados para la predicción de índices de costos de construcción.
- Identificar las variables explicativas empleadas en los modelos de predicción del ICC.
- Conocer las métricas utilizadas para evaluar los modelos de predicción del ICC.
- Identificar los horizontes de predicción del ICC de los modelos.

2.1 Métodos

Con el fin de garantizar la replicabilidad de los resultados que se obtendrán en la sección de antecedentes investigativos, se aplicará la metodología propuesta por Kitchenham [17], que consta de tres fases. En primera instancia la planificación, etapa en donde se definen las preguntas de investigación, los procesos de investigación, los criterios de inclusión y exclusión, la extracción de datos y su síntesis. En la segunda etapa, la ejecución de la revisión y finalmente la presentación de resultados, discusión y conclusiones que resultaron de la RL.

2.1.1. Planificación

La especificación de las preguntas de investigación que se pretenden responder a través de la RL, es el aspecto más importante, ya que dirigirán el estudio. En este sentido, se han definido las siguientes preguntas de investigación:

- **Preguntas de investigación**

PI: ¿Qué modelos han sido empleados para explicar y predecir valores del ICC?

- **PI1:** ¿Qué modelos o combinación de modelos se han aplicado para la predicción del ICC?
- **PI2:** ¿Qué variables explicativas se han empleado en los modelos de predicción del ICC?
- **PI3:** ¿Qué métricas se han utilizado para evaluar los modelos de predicción del ICC?
- **PI4:** ¿Cuál es el horizonte de predicción de los modelos estudiados?

- **Procesos de investigación**

La Tabla 1, indica los repositorios que fueron consultados con el propósito de encontrar estudios primarios que provean información para responder a las preguntas de investigación; en todos los casos se empleó la opción de búsqueda avanzada.

Table 1: Fuentes consultadas.

Repositorio	Acrónimo
Web of Science	WoS
ASCE Library	ASCE
Taylor & Francis Online	Taylor & Francis
Springer Link	Springer
SAGE journals	SAGE
Science Direct	Science Direct
MDPI	MDPI

Fuente: Mayra Viscaíno (2022)

- **Criterios de inclusión y exclusión**

Para seleccionar estudios que sean de interés y de aporte relevante para este trabajo, se adoptó el protocolo propuesto por Kitchenham [17]. Se establecieron cuatro criterios de inclusión (CI) y tres criterios de exclusión (CE) de estudios que se muestran en la Tabla 2, éstos ayudan a definir los límites de la RL, y finalmente determinarán el número de estudios que serán analizados. Un estudio será incluido en la RL siempre y cuando, éste cumpla todos los criterios de inclusión, y no sea descartado por ningún criterio de exclusión.

Los criterios fueron formulados en base a las preguntas de investigación y en el ámbito temporal de este trabajo; basado en ello se incluirán únicamente los estudios publicados en los últimos 5 años. Con el propósito de seleccionar estudios relevantes se incluirán únicamente estudios publicados en idioma inglés y que hayan sido citados al menos una vez; que explícitamente indiquen en el título que se emplearon modelos para

predicción o explicación del ICC.

Quedarán excluidos artículos de revisión, encuestas, notas técnicas o capítulos de libros, así como artículos duplicados. Una vez que los estudios hayan atravesado los criterios descritos anteriormente, se procederá a la lectura del resumen, si éste no proporciona información relevante que contribuyan a dar respuesta a las preguntas de investigación el estudio será descartado.

Table 2: Criterios para incluir o excluir estudios.

Orden de aplicación	Criterio de inclusión (CI) / Criterio de exclusión (CE)
1	CI1: Estudios publicados en el periodo 01-enero-2017 - 18-julio-2022.
2	CI2: Artículos o conferencias publicadas en idioma inglés.
3	CE1: Artículos de revisión, encuestas, notas técnicas, capítulos de libros.
4	CI3: Artículos citados al menos una vez.
5	CI4: Artículos de ICC que en el título indique que usan modelos para predicción o explicación.
6	CE2: Artículos duplicados (indentificados en otras fuentes)
7	CE3: El resumen no proporciona información relevante para el objeto de este estudio.

Fuente: Mayra Viscaíno (2022)

- **Extracción de datos**

En esta etapa se diseñó un formulario para la recopilación de la información de los estudios primarios. Los campos que componen el formulario contribuyen a responder a las preguntas de revisión y a la aplicación de los criterios de calidad del estudio [17]. A continuación, se detalla la información que debe ser extraída y registrada de cada estudio revisado:

- **Datos de la revisión:**

- * Nombre del revisor
- * Fecha de la revisión

- **Información básica del estudio:**

- * Título del artículo
- * Nombres de los autores
- * Fecha de publicación
- * Revista y repositorio
- * Link de acceso

- **Información relevante:**

- * Objetivo del estudio

- * Ámbito de aplicación
- * Nombre del modelo (explicativo o predictivo)
- * Descripción de las variables explicativas
- * Métricas de evaluación del modelo
- * Horizontes de predicción
- * Observaciones

- **Síntesis de los datos**

Según Kitchenham, la síntesis de datos implica recopilar y resumir los resultados de los estudios primarios que resultaron seleccionados. La síntesis de la información recopilada será descriptiva (no cuantitativa), en este caso. La información extraída sobre los estudios se tabularon de manera consistente con las preguntas de investigación. Las tablas están estructuradas de manera que permiten resaltar similitudes y diferencias entre los resultados de los estudios. Es importante identificar si los resultados de los estudios son consistentes entre sí (homogéneos) o inconsistentes (heterogéneos) [17].

2.1.2. Ejecución de la revisión

La siguiente etapa, una vez definido el protocolo, consistió en realizar la revisión, a través de la ejecución de las siguientes actividades: búsqueda de los estudios, para ello se emplearon las palabras clave en los repositorios seleccionados; aplicación del filtro criterios de inclusión y exclusión, para obtener los estudios seleccionados y finalmente, la extracción de la información.

- **Identificación de la investigación**

El objetivo de la investigación fue la directriz del enfoque de esta etapa. La estrategia adoptada para la selección de palabras clave, se fundamentó en la revisión previa de artículos científicos y artículos de revisión publicados en idioma inglés sobre el tema de interés; identificándose las siguientes palabras clave para la búsqueda de información: *construction costs index, time series, forecasting model, engineering news-record (ENR), cost estimation, house price index*.

- **Selección de estudios**

La estructura de las cadenas de búsqueda que fueron empleadas y las fechas de consulta se indican en la Tabla 3. Por otra parte, la Figura 1, esquematiza los resultados del proceso de selección. Finalmente, fueron seleccionados 20 estudios. La Figura 2, muestra las fuentes y el número de estudios encontrados. Las fuentes que aportaron con la mayor cantidad de estudios, fueron: ASCE, Elsevier y Spriger, con el 30%, 30% y 15%, respectivamente. La Tabla 4, presenta los estudios que resultaron seleccionados. Todos los estudios fueron publicados en los últimos cinco años (ver Figura 3.)

- **Extracción de datos**

La extracción de la información de cada estudio revisado fue documentada en una hoja de Excel y puede revisarse en el Anexo 1.

Table 3: Repositorios y cadenas de búsqueda.

Repositorio	Cadena de búsqueda	Consultado
WoS	(ALL=(Construction costs index)) AND ALL=(forecasting model) OR ((TI=(construction cost index)) AND ALL=(cost prediction)) AND ALL=(time series analysis) OR (((TI=(construction cost index)) OR ALL=(ENR)) OR ALL=(Engineering news-record)) AND ALL=(forecasting)	18-07-2022
ASCE	(TI=(Construction costs index)) AND ALL=(forecasting model) OR (TI=(Construction costs index)) AND ALL=(Engineering news-record) AND ALL=(cost prediction) OR (TI=(Construction costs index)) AND ALL=(cost estimation) AND ALL=(time series)	18-07-2022
Taylor & Francis	(title: “construction cost index”) OR (title: (“ENR”) AND (anywhere: “construction cost index”) OR (anywhere: “house price index”) AND (keywords: “forecasting”)) OR (title: (“construction cost index”) AND (anywhere: “time series”) AND (anywhere: “cost estimation”))	18-07-2022
Springer	(“construction cost index”) AND (forecasting OR model) OR (title: “house price index” AND (forecasting OR model)) OR (“engineering news-record”) AND (forecasting OR prediction OR estimation)	18-07-2022
SAGE	[[Abstract construction cost index] AND [All forecasting model]] OR [[Abstract house price index] AND [Abstract cost estimation] AND [All prediction model]] OR [[Abstract construction cost index] AND [All arima model] AND [Abstract forecasting]]	19-07-2022
Science Direct	((construction cost index cci) AND forecasting) OR ((Engineering news-record (ENR)) AND (forecasting OR prediction)) OR ((ENR OR (“Construction Cost Index”)) AND ((forecasting model) OR (cost estimation)))	19-07-2022
MDPI	Search for Articles: ((abstract) “construction cost index”) OR ((abstract) “Engineering news-record (ENR)”) AND (all) “time series”)	19-07-2022

Fuente: Mayra Viscaíno (2022)

2.2 Resultados

Luego de haber realizado la extracción de la información, se procede a responder a las preguntas de investigación.

Table 4: Estudios seleccionados.

Cód.	Autor	Descripción del artículo	Año	Ref.
P1	F. Jiang et al.	Analysis of Construction Cost and Investment Planning Using Time Series Data.	2022	[16]
P2	M. Mir et al.	Neural network-based interval forecasting of construction material prices.	2021	[18]
P3	S. Faghih et al.	Time Series Analysis Framework for Forecasting the Construction Labor Costs.	2021	[19]
P4	C.-Y. Choi et al.	Predicting City-Level Construction Cost Index Using Linear Forecasting Models.	2021	[20]
P5	Y. Cao et al.	Predicting the Volatility of Highway Construction Cost Index Using LSTM.	2020	[21]
P6	O. Swei	Forecasting Infidelity-Why Current Methods for Predicting Costs Miss the Mark.	2020	[22]
P7	J. Ahn et al.	Performance evaluation of normalization-based CBR models for improving construction cost estimation.	2020	[23]
P8	J. Dong et al.	Cost Index Predictions for Construction Engineering Based on LSTM Neural Networks.	2020	[9]
P9	P. Pakalapati et al.	Using moving-window cross-validation algorithm to optimize bid-based cost estimating data usage.	2020	[24]
P10	D. Chakraborty et al.	A novel construction cost prediction model using hybrid natural and light gradient boosting.	2020	[25]
P11	Y. Elfahham	Estimation and prediction of construction cost index using neural networks, time series, and regression.	2019	[8]
P12	L. Zhao et al.	Forecasting residential building costs in New Zealand using a univariate approach.	2019	[3]
P13	S. Mao et al.	A novel method for forecasting Construction Cost Index based on complex network.	2019	[26]
P14	S. Moon et al.	Predicting Construction Cost Index Using the Autoregressive Fractionally Integrated Moving Average Model.	2018	[10]
P15	S. Faghih et al.	Forecasting Construction Material Prices Using Vector Error Correction Model.	2018	[27]
P16	M. Rafiei et al.	Novel Machine-Learning Model for Estimating Construction Costs Considering Economic Variables and Indexes.	2018	[28]
P17	R. Zhang et al.	Forecasting Construction Cost Index Based on Visibility Graph: A Network Approach.	2018	[29]
P18	T. Moon et al.	Forecasting Model of Construction Cost Index Based on VECM with Search Query.	2017	[30]
P19	T. Moon et al.	Forecasting Construction Cost Index Using Interrupted Time-Series.	2017	[12]
P20	O. Oshodi et al.	Comparing univariate techniques for tender price index forecasting- Box-Jenkins and neural network model.	2017	[31]

Fuente: Mayra Viscaíno (2022)

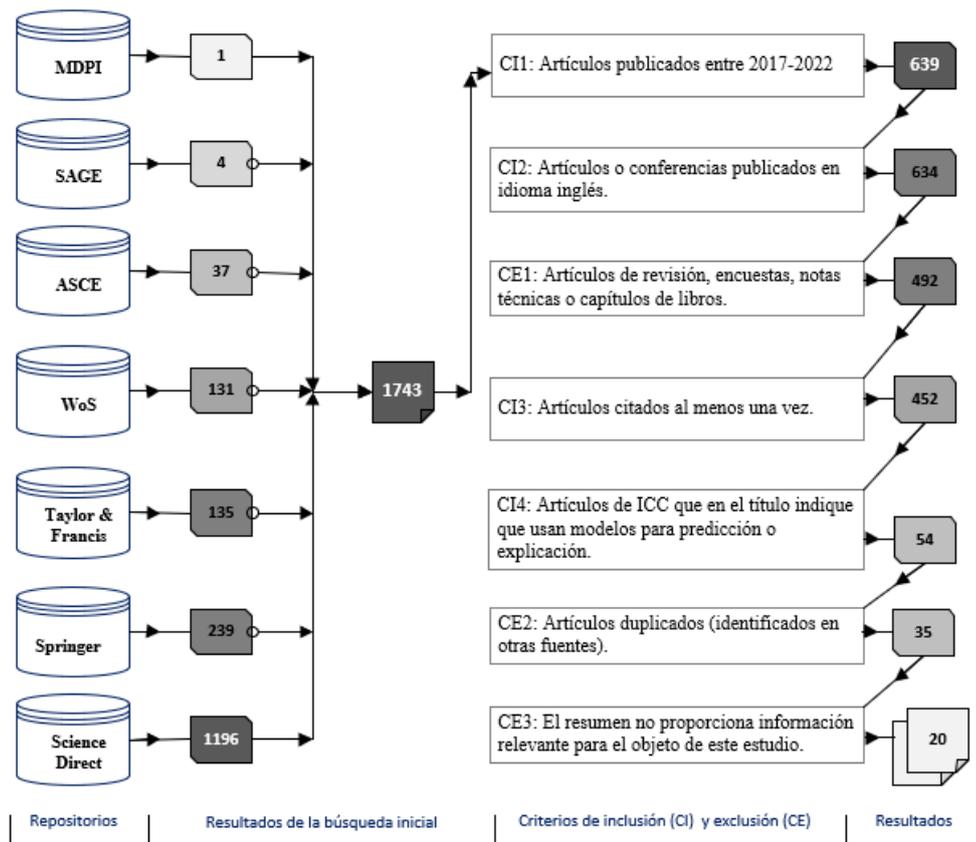


Figure 1: Proceso de selección de los estudios.
Fuente: Mayra Viscaíno (2022)

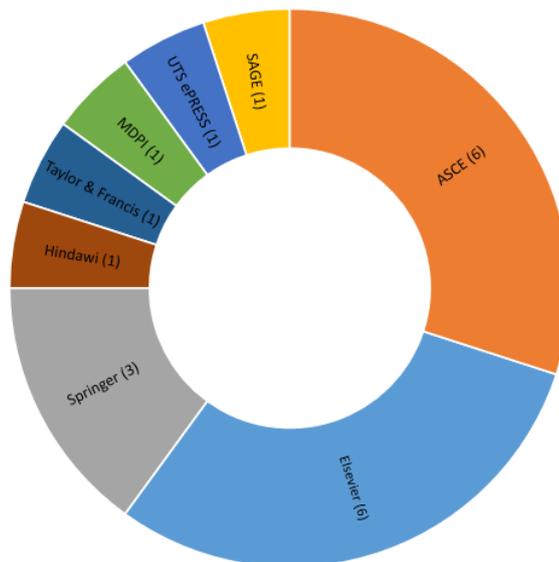


Figure 2: Fuentes y cantidad de estudios seleccionados.
Fuente: Mayra Viscaíno (2022)

2.2.1. PI1: ¿Qué modelos se han aplicado para la predicción del ICC?

Los responsables de preparar presupuestos y licitaciones en el área de la construcción emplean los valores del ICC para tomar decisiones, lo cual genera

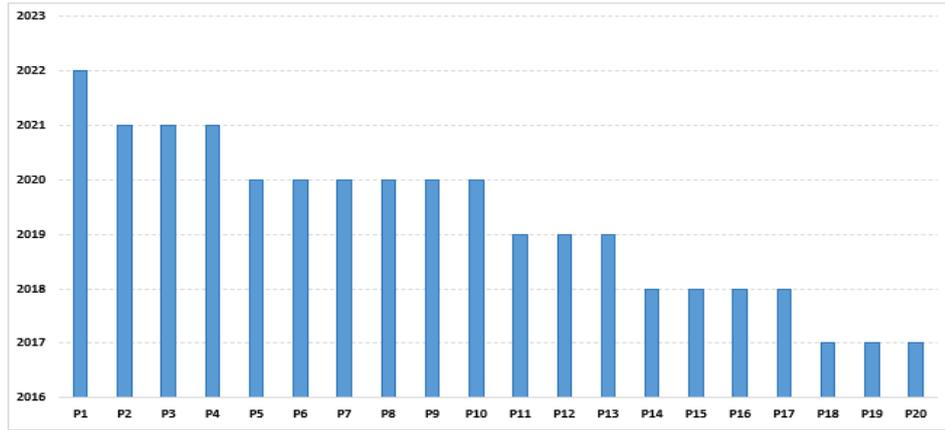


Figure 3: Año de publicación de los estudios seleccionados.
Fuente: Mayra Viscaíno (2022)

la necesidad de realizar predicciones del ICC tan precisas como sea posible, esto ha motivado a investigadores a diseñar y probar la capacidad predictiva de varios modelos.

A partir de la revisión de la literatura, se categoriza en tres grupos los modelos usados para predecir valores del ICC, éstos se clasifican en: modelos estadísticos, modelos que emplean algoritmos de *Machine Learning* y modelos que emplean otras técnicas, como: análisis de redes complejas y razonamiento basado en casos (CBR).

Dentro de los modelos estadísticos están los análisis de regresión y las series de tiempo. En el primer caso, estos modelos asumen que la variable respuesta está relacionada linealmente con variables independientes explicativas [32]. Estudios previos, usaron modelos de regresión lineal múltiple con el propósito de comparar su desempeño con otros métodos; es el caso de [8] y [16] que emplearon estos modelos para la predicción del ICC, mientras que [25], los usó para la estimación de costos de construcción de elementos estructurales.

Por otra parte, en series de tiempo se pueden diseñar modelos univariados y multivariados. Modelos univariados como: *Winter's additive*, *Simple Seasonal* fueron probados por [16] y métodos de suavizado exponencial de Holt-Winters por [3].

En esta categoría, los modelos que mejor capacidad para predecir valores del ICC, según [3], [16] y [20], fueron los modelos autorregresivos integrados de media móvil (ARIMA) y sus variaciones como ARIMA Interrumpido [12] y el modelo autorregresivo fraccionalmente integrado de media móvil ARFIMA (*Autoregressive fractionally integrated moving average*) [10]. Estos modelos también han sido usados para comparar su desempeño con algoritmos de *Machine Learning* (ML), como las redes neuronales artificiales [8], [31].

Dentro de los modelos de series de tiempo multivariados, el modelo Vecto-

rial de Corrección de Errores (VEC), ha sido probado para la estimación de costos laborales en la construcción [19], el pronóstico del precio individualizado de materiales (asfalto, acero y cemento) [28] y la predicción del ICC [30].

Dentro de las técnicas de ML, se han probado diversos tipos de Redes Neuronales Artificiales [8], [9], [18], [21], [25], [28], [31]. Otras técnicas como el análisis de las redes complejas [26], [29] y razonamiento basado en casos (CBR) [23], también fueron probados con estos propósitos.

Dentro de la revisión de la literatura, se identificó que el 55% de los estudios usaron modelos estadísticos, el 30% empleó algoritmos de ML y el 15% empleó otros tipos de modelos. La Figura 4, muestra un resumen de los modelos que mejor desempeño tuvieron dentro del conjunto de modelos probados en los estudios revisados, y que finalmente fueron catalogados por los investigadores como útiles para la estimación de costos de construcción.

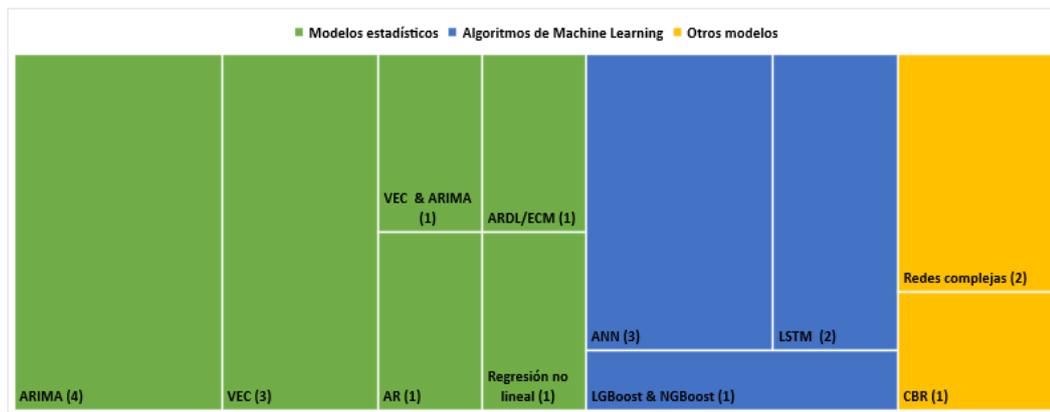


Figure 4: Modelos predictivos usados por los estudios.

Fuente: Mayra Viscaíno (2022)

2.2.2. PI2: ¿Qué variables explicativas se han empleado en los modelos de predicción del ICC?

Los investigadores que probaron modelos que requieren variables exógenas, tuvieron que afrontar el reto de definir un conjunto de variables independientes candidatas, de hecho este es un asunto cuestionado por la dificultad que representa. Un sin número de variables han sido propuestas por los diferentes autores, y no hay unanimidad sobre ellas.

La revisión de la literatura permitió identificar a factores: macroeconómicos, como el Índice de Precios al Consumidor (IPC), como la variable más usada; condiciones del mercado de la construcción como los precios de materiales específicos: concreto, hierro, cemento, materiales pétreos, vidrio, madera y otros materiales [8], [9], [25], [27]; y factores que caracterizan aspectos físicos de la obra, como: área del lote, área bruta construida, número de pisos, tipo de sistema estructural, entre otros [23], [25], [28].

Algunos investigadores también consideraron las condiciones económicas de la población, como: los ingresos y gastos personales, tasas de empleo y desempleo [16], [20], [27]. La Tabla 5, muestra un conjunto de variables exógenas y las investigaciones que las han usado.

Table 5: Estudios que usaron variables exógenas para predicción del ICC

Factores exógenos	Referencias
Índice de precios al consumidor	[9], [16], [18], [20],[27], [28]
Tasa de desempleo general	[16]
Tasa de empleo	[16], [18]
Índice de precios al productor	[16], [27]
Precios del petróleo crudo	[9], [16], [20]
Producto Interno Bruto	[9], [16]
Índice de precios de importación	[16]
Oferta monetaria	[16], [20]
Precio del petróleo crudo West Texas Intermediate	[18], [27]
Tasa de interés para préstamo	[9], [28]
Número de permisos de construcción emitidos	[16], [18], [27], [28]
Nuevas viviendas	[18], [27]

Fuente: Mayra Viscaíno (2022)

2.2.3. PI3: ¿Qué métricas se han utilizado para evaluar los modelos de predicción del ICC?

La evaluación del desempeño de los modelos probados es determinante para saber qué modelo tiene mejor capacidad predictiva; se han usado varias métricas con este fin (ver Tabla 6); en todos los casos, valores bajos de las métricas indican un mejor rendimiento predictivo del modelo [19].

Se identificó que los indicadores que con mayor frecuencia fueron usados son: el Error Porcentual Absoluto Medio (MAPE), Error Absoluto Medio (MAE) y Raíz del Error Cuadrático Medio (RMSE). Según [19],[27] y [16], valores inferiores al 10% indican buen desempeño predictivo de un modelo.

Investigadores que probaron modelos multivariados, calcularon el valor de R^2 , para valorar la correlación de la variable respuesta con las variables explicativas candidatas [16], [25], [27].

2.2.4. PI4: ¿Cuál es el horizonte de predicción de los modelos en los estudios seleccionados?

Según [21], un buen modelo de predicción debería pronosticar en diferentes horizontes temporales, o al menos se debería indicar el horizonte para el cual el modelo presenta un desempeño confiable. En algunos casos este aspecto no es mencionado [25], [26], [28]. De la revisión de los estudios, se identificó que

Table 6: Métricas del error para evaluar modelos predictivos del ICC

Siglas	Métricas	Referencias
MAPE	Error porcentual absoluto medio	[3], [9], [10], [12], [16],[18], [19], [21], [22],[24], [26], [27], [29], [30], [31]
MAE	Error absoluto medio	[3], [8], [9], [12], [16], [19], [21], [27], [29], [30]
RMSE	Raíz del error cuadrático medio	[3], [8], [12], [16], [18], [19], [20], [25], [27], [30]
MSE	Error cuadrático medio	[9], [22], [29], [28]
U Theil	Coeficiente de desigualdad Theil	[19], [27], [31]
BIC	Criterio de información bayesiano	[3], [20]
MAD	Desviación absoluta media	[23], [26]
AIC	Criterio de información de Akaike	[20]
SMAPE	Error porcentual absoluto medio simétrico	[26]
MBE	Error medio de sesgo	[25]

Fuente: Mayra Viscaíno (2022)

existen tres horizontes: corto, mediano y largo plazo; en los cuales los investigadores han probado sus modelos y afirman que tienen un buen desempeño, sea en alguno de ellos o en todos los casos.

Es importante destacar, que no hay consenso en establecer el periodo de tiempo que corresponde a cada categoría, para algunos investigadores un periodo de tiempo de 6 o 12 meses, significa corto plazo; mientras que un largo plazo puede ir de 3 a 8 años. Sin embargo, varios investigadores [10], [20] y [29] consideraron 12, 24 y 36 meses como corto, mediano y largo plazo respectivamente. En la Tabla 7, se presentan los modelos y el horizonte de pronóstico para el cual los modelos tienen un desempeño confiable, de acuerdo a cada estudio.

2.3 Discusión

En esta sección se presenta una breve discusión sobre los hallazgos relacionados con las preguntas de investigación que motivaron esta revisión de la literatura.

PI1: Modelos predictivos de ICC.

Varios modelos han sido probados, en los diferentes ámbitos de la estimación de costos de construcción. De entre ellos destacan los modelos autorregresivos integrados de media móvil (ARIMA), por su buena capacidad predictiva, según [3], [16] y [20]; sin embargo, algunos investigadores aclaran que esto es cierto para periodos de tiempo cortos. El buen desempeño de estos modelos ha hecho que sean considerados como modelos referenciales, a ser superados por otros modelos.

Table 7: Horizontes de predicción de ICC

Estudio	Modelo	Corto plazo	Mediano plazo	Largo plazo
P1	ARIMA		16 meses	
P2	LUBE-ANN	5 meses		
P3	VEC	X		X
P4	VEC o ARIMA	12 meses	24 meses	36 meses
P5	LSTM	12 meses	24 meses	96 meses
P6	ARDL/ECM	6 meses		
P8	LSTM	X		
P9	Regresión no lineal	12 meses		60 meses
P11	AR			84 meses
P12	ARIMA	X		
P14	ARFIMA	12 meses	24 meses	36 meses
P15	VEC	X		X
P17	Redes complejas	12 meses	24 meses	36 meses
P19	ARIMA Inter-rumpido	12 meses		
P20	ANN			36 meses

NOTA: “X”, indica que el estudio no especifica el tiempo del horizonte de pronóstico.

Fuente: Mayra Viscaíno (2022)

Una ventaja de los modelos autorregresivos, es que determinan tendencias futuras empleando el conocimiento de valores pasados de los datos; y por tanto, se requiere únicamente un conjunto de datos del ICC para diseñar el modelo y generar las predicciones.

PI2: Variables explicativas o predictoras.

Variables independientes de tipo macroeconómicas han sido usadas por algunos investigadores y se ha verificado que existe influencia de ellas sobre los costos de construcción [28].

PI3: Métricas.

En cuanto a las métricas usadas para evaluar los modelos, a pesar de que se identificaron varias de ellas; se observó que hay unanimidad en usar el MAPE, MAE, RMSE como indicadores para evaluar la capacidad predictiva de los modelos, son deseables valores bajos en todos los casos.

PI4: Horizontes de predicción.

Se evidenció que no existe consenso en el establecimiento de los períodos de tiempo, para los diferentes horizontes de pronóstico. En los estudios revisados, el 30% afirma que el corto plazo corresponde a 12 meses, el 20% afirmó que 24 y 36 meses corresponde a corto y mediano plazo respectivamente. El 5% de los estudios, sólo reportó el horizonte de manera cualitativa, sin especificar el lapso de tiempo. Mientras que, el 25% de los estudios no indicó el horizonte del predicción de sus modelos.

2.4 Conclusiones preliminares

- El modelado estadístico de series de tiempo sigue siendo uno de los métodos cuantitativos más usados en este campo de investigación.
- Las variables exógenas que más fueron empleadas para predecir los valores del ICC son: el índice de precios al consumidor, precio del petróleo crudo, número de permisos de construcción.
- La métrica más usada para evaluar el desempeño predictivo de un modelo es el MAPE, seguido del RMSE.
- La mayoría de los modelos tienen un buen desempeño predictivo a corto plazo (12 meses).

2.5 Marco Teórico

En esta sección se presenta la teoría sobre la cual se fundamentó el desarrollo de este trabajo. Se presentan algunas definiciones importantes, así como la descripción de los modelos empleados.

2.5.1. Series temporales

Dado que la información que se empleó en este trabajo corresponde a series cronológicas, es relevante definir las. Según [33], *“Una serie temporal o cronológica es una secuencia de datos, observaciones o valores, medidos en determinados momentos del tiempo, ordenados cronológicamente y, normalmente, espaciados entre sí de manera uniforme”*. Una serie de tiempo, puede ser expresada de diferentes maneras, una de ellas es según la ecuación 1:

$$Y_t : t = 1, \dots, N \quad (1)$$

En donde Y_t , es la observación t ($1 \leq t \leq N$) de la serie y N , es el número (tamaño o longitud) de la serie completa.

2.5.2. La media en las series temporales

Según [32], algunos métodos de pronóstico son extremadamente simples y sorprendentemente efectivos. En el caso del método de la media, los pronósticos de todos los valores futuros son iguales al promedio (o “media”) de los datos históricos. Si Y_1, \dots, Y_t , representan los datos históricos, los pronósticos pueden escribirse como sigue:

$$\hat{Y}_{t+h|t} = \bar{Y} = \frac{Y_1 + \dots + Y_t}{t}$$

Donde:

- $\hat{Y}_{t+h|t}$: forma abreviada de la estimación de Y_{t+h} basada en los datos Y_1, \dots, Y_t .
- \bar{Y} : promedio de los datos históricos.

2.5.3. Modelo Naïve

Para los pronósticos ingenuos, simplemente se configuran todos los pronósticos para que su valor sea el valor de la última observación. Es decir:

$$\hat{Y}_{t+h|t} = Y_t$$

Este método funciona muy bien para muchas series temporales económicas y financieras [32].

2.5.4. Modelo Drift

Una variación del método ingenuo es permitir que los pronósticos aumenten o disminuyan con el tiempo, donde la cantidad de cambio con el tiempo (llamada deriva) se establece como el cambio promedio observado en los datos históricos [32]. Así, el pronóstico para el tiempo viene dado por:

$$\hat{Y}_{t+h|t} = Y_t + \frac{h}{t-1} \sum_{i=2}^t (Y_i - Y_{i-1}) = Y_t + h \left(\frac{Y_t - Y_1}{t-1} \right)$$

Donde:

- $\hat{Y}_{t+h|t}$: forma abreviada de la estimación de Y_{t+h} basada en los datos Y_1, \dots, Y_t .
- Y_1 : primer dato histórico.
- Y_t : último dato histórico.
- h : horizonte de pronóstico.

2.5.5. Modelo aditivo y multiplicativo

Al igual que en otros campos de la estadística, en las series temporales se busca identificar patrones, como eventos estacionales o atípicos, que aporten información sobre eventos pasados y permitan su predicción. Estos patrones, pueden clasificarse en estructurales y no estructurales [34].

- **Patrones estructurales:** Los tres tipos de patrones estructurales, conocidos como las componentes de la serie, son: la tendencia (T_t), el ciclo (C_t) y la estacionalidad (S_t); las series pueden tener algunas o todas ellas. La tendencia, representa la dirección general de la serie en el tiempo, ya sea creciente, decreciente, lineal o exponencial. Un ciclo puede describirse como una secuencia de eventos repetibles a lo largo del tiempo, donde el punto de inicio de un ciclo está en un mínimo local de la serie y el punto final está en el siguiente. La estacionalidad indica una variación repetida en la serie, que está relacionada con las unidades de frecuencia de la serie (por ejemplo, los meses del año para una serie mensual).
- **Patrones no estructurales:** Se refiere a cualquier otro patrón que no esté relacionado con las componentes de la serie, también se le conoce como la componente irregular (I_t).

En base a esta categorización, las series cronológicas pueden expresarse mediante las siguientes ecuaciones:

- **Serie con estructura aditiva:**

$$Y_t = T_t + S_t + C_t + I_t$$

- **Serie con estructura multiplicativa:**

$$Y_t = T_t \times S_t \times C_t \times I_t$$

2.5.6. Modelo de Suavizado Exponencial Simple (SES)

Según [34], el principal atributo del modelo SES es la técnica del promedio ponderado, que se basa en el decaimiento exponencial de los pesos de observación según su distancia cronológica (es decir, índice de serie o marca de tiempo) desde los primeros valores pronosticados. La tasa de caída de los pesos de observación se establece mediante el parámetro de suavizado del modelo. La función SES es una función escalonada, donde el valor n pronosticado del modelo se convierte en la entrada del próximo pronóstico, de las próximas observaciones, $n + 1$. Formalmente, esta relación se expresa de la siguiente manera:

$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha) \hat{Y}_t$$

Donde:

- \hat{Y}_{t+1} : es el valor pronosticado de la observación $t + 1$, para una serie de n observaciones ($t=1, \dots, n$)
- Y_t : es la observación t , de la serie.
- α : es el parámetro de suavizado del modelo, donde $0 < \alpha \leq 1$.
- \hat{Y}_t : es el valor pronosticado de la observación t al paso $t - 1$.

2.5.7. Modelo Holt

El método de Holt, también conocido como modelo de suavizado exponencial doble, es una versión ampliada del modelo SES. Se basa en estimar el nivel y la tendencia más recientes con el uso de dos parámetros de suavizado. Una vez que el modelo estima el nivel y la tendencia más recientes, los utiliza para construir el pronóstico de la serie [34], utilizando la ecuación (2) y ecuación (3), dependiendo el caso.

- Para series con tendencia aditiva:

$$\hat{Y}_{t+h} = L_t + hT_t \tag{2}$$

- Para series con tendencia multiplicativa:

$$\hat{Y}_{t+h} = L_t \times hT_t \tag{3}$$

Donde:

- \hat{Y}_{t+h} : es el valor pronosticado del h valor pronosticado de una serie con t observaciones.
- L_t : es la estimación del nivel de la serie en el tiempo t .
- T_t : es la estimación del impacto marginal de la tendencia de la serie en el momento t .
- h : es el número de pasos de pronóstico desde el tiempo t .

Similar al modelo SES, el cálculo del nivel y la tendencia de la serie por parte del modelo Holt, se basa en un promedio ponderado con el uso de dos parámetros de suavizado α y β . Para una serie con tendencia aditiva, el cálculo del nivel y tendencia más reciente de la serie se puede obtener usando las ecuaciones (4) y (5):

$$L_t = \alpha Y_t + \alpha(1 - \alpha)(L_{t-1} + T_{t-1}) \quad (4)$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (5)$$

Para series con tendencia multiplicativa, deben usarse las ecuaciones (6) y (7).

$$L_t = \alpha Y_t + \alpha(1 - \alpha)(L_{t-1} \times T_{t-1}) \quad (6)$$

$$T_t = \beta \left(\frac{L_t}{L_{t-1}} \right) + (1 - \beta)T_{t-1} \quad (7)$$

El modelo de Holt estima los valores de L_t y T_t , el nivel y la tendencia de la serie en el tiempo “ t ”, utilizando un promedio ponderado de todas las observaciones de la serie. De manera similar al proceso de estimación del modelo SES, este proceso recursivo comienza con el pronóstico de la segunda observación del modelo y está dado por las ecuaciones (8) y (9).

$$\hat{L}_2 = \alpha Y_1 + \alpha(1 - \alpha)(\hat{L}_1 + \hat{T}_1) \quad (8)$$

$$\hat{T}_2 = \beta(\hat{L}_2 - \hat{L}_1) + (1 - \beta)\hat{T}_1 \quad (9)$$

2.5.8. Modelo Holt-Winters

El modelo Holt-Winters (HW), es el modelo más avanzado de los modelos de suavizado exponencial; es una versión extendida del modelo Holt y puede manejar datos de series de tiempo con componentes de tendencia y estacionales. Pronosticar el componente estacional requería un tercer parámetro y ecuación más suaves, además de los de nivel y tendencia [34].

Tanto los componentes de tendencia como los estacionales podrían tener una estructura aditiva o de multiplicidad, lo que agrega cierta complejidad al modelo, ya que existen múltiples combinaciones posibles, como las que se indican:

- Tendencia aditiva y componentes estacionales.
- Tendencia aditiva y componentes estacionales multiplicativos.
- Tendencia multiplicativa y componentes estacionales aditivos.
- Tendencia multiplicativa y componentes estacionales.

Por lo tanto, antes de construir un modelo HW, se necesita identificar la estructura de la tendencia y los componentes estacionales. Las ecuaciones 10, 11, 12 y 13, describen el modelo HW para una serie con componente estacional aditivo.

$$\hat{Y}_{t+1} = L_t + hT_t + S_{t+h-m} \quad (10)$$

$$L_t = \alpha(Y_t - S_{t-m}) + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (11)$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta) + (1 - \beta)T_{t-1} \quad (12)$$

$$S_t = \gamma(Y_t - L_t) + (1 - \gamma)S_{t-m} \quad (13)$$

Mientras que las ecuaciones 14, 15, 16 y 17, describen el modelo HW para una serie con una estructura estacional multiplicativa:

$$\hat{Y}_{t+1} = (L_t + hT_t)S_{t+h-m} \quad (14)$$

$$L_t = \frac{\alpha Y_t}{S_{t-m}} + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (15)$$

$$T_t = \beta(L_t - L_{t-1} - 1) + (1 - \beta)T_{t-1} \quad (16)$$

$$S_t = \gamma \frac{Y_t}{L_t} + (1 - \gamma)S_{t-m} \quad (17)$$

Donde m , representa la longitud de la ventana móvil ($m \leq t$).

2.5.9. Modelo Regresión Lineal

El uso principal del modelo de regresión lineal es cuantificar la relación entre la variable dependiente “Y” (también conocida como variable respuesta) y la(s) variable(s) independiente(s) “X” (también conocida(s) como variable predictor(a)s, impulsora(s) o regresora(s)) de manera lineal. En otras palabras, el modelo expresa la variable dependiente como una combinación lineal de las variables independientes [34]. La relación entre las variables dependientes e independientes, se puede generalizar mediante las ecuaciones 18 y 19:

- **Con una variable independiente:**

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \epsilon_i \quad (18)$$

- **Con n variables independientes:**

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \dots + \beta_n X_{n,i} + \epsilon_i \quad (19)$$

Donde:

- i : representa el índice de las observaciones, $i = 1, \dots, N$
- Y_i : representa la i – *ésima* observación de la variable dependiente.
- $X_{j,i}$: representa el valor i – *ésimo* de la variable independiente j , donde $j = 1, \dots, n$
- β_0 : representa el valor del término constante (o intercepto).
- β_j : representa los parámetros correspondientes (o coeficientes) de las j variables independientes.
- ϵ_i : define el término de error, que no es más que toda la información que no fue capturada por las variables independientes para la i – *ésima* observación.

El modelo de regresión lineal, a diferencia de los modelos tradicionales de series temporales como ARIMA o Holt-Winters, no se diseñó explícitamente para manejar y pronosticar datos de series temporales. Sin embargo, es un modelo genérico con una amplia gama de aplicaciones, desde la inferencia causal hasta el análisis predictivo [34].

Suponiendo que las ecuaciones 18 y 19, representan la verdadera naturaleza de la relación entre las variables dependientes e independientes, entonces el modelo de regresión lineal proporciona una estimación de esos coeficientes, que se puede formalizar mediante las ecuaciones 20 y 21:

- **Modelo de regresión lineal univariante:**

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1,i} \quad (20)$$

- **Modelo de regresión lineal multivariante:**

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1,i} + \hat{\beta}_2 X_{2,i} + \dots + \hat{\beta}_n X_{n,i} \quad (21)$$

Donde:

- i : representa el índice de las observaciones, $i = 1, \dots, N$
- \hat{Y}_i : representa la estimación de la i – *ésima* observación de la variable dependiente.

- $X_{j,i}$: representa el valor i –ésimo de la variable independiente j , donde $j = 1, \dots, n$
- $\hat{\beta}_0$: representa la estimación del valor del término constante.
- $\hat{\beta}_1, \dots, \hat{\beta}_n$: son las estimaciones de los parámetros correspondientes (o coeficientes) de las n variables independientes.

La estimación de los coeficientes del modelo se basa en definir una función de costo (conocida como función de pérdida), estableciendo alguna métrica de error y optimización matemática para minimizar la función de costo.

Estimación de los coeficientes con el método OLS

El enfoque de estimación más común es aplicar los mínimos cuadrados ordinarios, OLS por sus siglas en inglés (*Ordinary Least Squares*). El objetivo de OLS es identificar los coeficientes que minimizan la suma de cuadrados de los residuos. Supóngase que el residuo de la i –ésima observación se define como sigue:

$$\hat{\epsilon}_i = Y_i - \hat{Y}_i$$

La función de costo, puede expresarse mediante la ecuación 22:

$$\sum_{i=1}^N \hat{\epsilon}_i^2 = (Y_1 - \hat{Y}_1)^2 + (Y_2 - \hat{Y}_2)^2 + \dots + (Y_n - \hat{Y}_n)^2 \quad (22)$$

Por simplicidad, se representa la función de costo en forma matricial:

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ \vdots \\ Y_N \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & X_{1,1} & \cdot & \cdot & \cdot & X_{1,n} \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ 1 & X_{N,1} & \cdot & \cdot & \cdot & X_{N,n} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \beta_n \end{bmatrix}, \boldsymbol{\epsilon} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdot \\ \cdot \\ \epsilon_N \end{bmatrix}$$

Donde:

- \mathbf{Y} : representa la variable dependiente con N observaciones.
- \mathbf{X} : representa las correspondientes n variables independientes y los escalares 1 corresponden al intercepto β_0 .
- $\boldsymbol{\beta}$: representan los coeficientes del modelo.
- $\boldsymbol{\epsilon}$: representa el error y se obtiene de $\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$

La suma de cuadrados del error (**SSE**) en términos matriciales es entonces:

$$\mathbf{SSE} = \boldsymbol{\epsilon}^T \boldsymbol{\epsilon} \quad (23)$$

Expandiendo la expresión 23:

$$\boldsymbol{\epsilon}^T \boldsymbol{\epsilon} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \quad (24)$$

Aplicando la propiedad distributiva en la ecuación 24:

$$\boldsymbol{\epsilon}^T \boldsymbol{\epsilon} = \mathbf{Y}^T \mathbf{Y} - 2\mathbf{Y}^T \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X}\boldsymbol{\beta} \quad (25)$$

Como el objetivo es encontrar el $\boldsymbol{\beta}$ que minimice la ecuación 25, ésta se deriva respecto a $\boldsymbol{\beta}$ y luego se la iguala a cero:

$$\frac{\partial \boldsymbol{\epsilon}^T \boldsymbol{\epsilon}}{\partial \boldsymbol{\beta}} = \frac{\partial (\mathbf{Y}^T \mathbf{Y} - 2\mathbf{Y}^T \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X}\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0 \quad (26)$$

Resolviendo la ecuación 26, se tiene:

$$\mathbf{X}^T \mathbf{X}\boldsymbol{\beta} = \mathbf{X}^T \mathbf{Y} \quad (27)$$

Operando en la ecuación 27, se tiene la estimación del vector de coeficientes $\hat{\boldsymbol{\beta}}$.

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

2.5.10. Proceso autorregresivo (AR)

El proceso autorregresivo (AR), define el valor actual de la serie Y_t , como una combinación lineal de los p rezagos anteriores de la serie [34], y se puede formalizar como sigue:

$$AR(p) : Y_t = c + \sum_{i=1}^p \phi_i Y_{t-i} + \epsilon_t$$

Donde:

- $AR(p)$: es la notación para un proceso AR con orden p .
- c : representa una constante (o deriva).
- p : define el número de rezagos de la regresión contra Y_t .
- ϕ_i : es el coeficiente del i –ésimo rezago de la serie (debe estar entre -1 y 1, de lo contrario, la serie tendría una tendencia alta o baja y, por lo tanto, no puede ser estacionaria en el tiempo).
- Y_{t-i} : es el i –ésimo rezago de la serie.
- ϵ_t : representa el término de error, el cual es ruido blanco (un proceso estocástico con media 0 y varianza constante).

Se puede usar un proceso AR en datos de series de tiempo sí, y solo sí, la serie es estacionaria. Por lo tanto, antes de aplicar un proceso AR en una serie, deberá verificarse que la serie sea estacionaria. De lo contrario, deberá aplicarse algún método de transformación (como diferenciación, transformación logarítmica, etc.) para llevar la serie a un estado estacionario. El modelo ARIMA, puede manejar datos de series de tiempo no estacionarios [34].

2.5.11. Proceso de media móvil (MA)

En algunos casos, el modelo de pronóstico no puede capturar todos los patrones de la serie y, por lo tanto, queda algo de información en los residuos del modelo (o error de pronóstico) ϵ_t . El objetivo del proceso de promedio móvil es capturar patrones en los residuos, si existen, modelando la relación entre Y_t , el término de error, ϵ_t y los términos de error “ q ” pasados de los modelos (por ejemplo, $\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_{t-q}$). La estructura del proceso MA es bastante similar a la del AR [34]. La siguiente ecuación define un proceso MA con un orden “ q ”:

$$MA(q) : Y_t = \mu + \epsilon_t + \sum_{i=1}^q \theta_i \epsilon_{t-i}$$

Donde:

- $MA(q)$: es la notación para un proceso MA con orden q .
- μ : representa la media de la serie.
- $\epsilon_{t-q}, \dots, \epsilon_t$: son términos de error de ruido blanco.
- θ_i : es el coeficiente correspondiente de ϵ_{t-i} .
- q : define el número de términos de error pasado que se utilizarán en la ecuación.

2.5.12. Modelo ARMA

El modelo ARMA es una combinación de los procesos $AR(p)$ y $MA(q)$; esta combinación permite manejar datos de series temporales más complejas y se expresa de la siguiente manera:

$$ARMA(p, q) : Y_t = c + \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{i=1}^q \theta_i \epsilon_{t-i} + \epsilon_t$$

Donde:

- $ARMA(p, q)$: define un proceso $ARMA$ con un proceso AR de orden p y un proceso MA de orden q .
- Y_t : representa a la serie de tiempo.
- c : representa una constante (o deriva).
- p : define el número de rezagos de la regresión contra Y_t .
- ϕ_i : es el coeficiente del i –ésimo rezago de la serie.
- Y_{t-i} : es el i –ésimo rezago de la serie.
- q : define el número de términos de error pasado que se utilizarán en la ecuación.

- θ_i : es el coeficiente correspondiente de ϵ_{t-i} .
- $\epsilon_{t-q}, \dots, \epsilon_t$: son términos de error de ruido blanco.
- ϵ_t : representa el término de error, el cual es ruido blanco.

2.5.13. Modelo ARIMA

Una de las limitaciones de los modelos AR, MA y ARMA es que no pueden manejar datos de series de tiempo no estacionarios. Por lo tanto, si la serie de entrada no es estacionaria, se requiere un paso de preprocesamiento para transformar la serie de un estado no estacionario a un estado estacionario. El modelo ARIMA proporciona una solución para este problema al agregar el proceso integrado para el modelo ARMA. El proceso Integrado (I) es simplemente diferenciar la serie con sus rezagos, donde el grado de diferenciación está representado por el parámetro “ d ”. Por ejemplo, $Y_t - Y_{t-1}$ representa la primera diferenciación de la serie, mientras que $(Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})$ representa la segunda diferenciación [34]. Se puede generalizar el proceso de diferenciación con la ecuación 28:

$$Y_d = (Y_t - Y_{t-1}) - \dots - (Y_{t-d+1} - Y_{t-d}) \quad (28)$$

Donde, Y_d es la diferenciación “ d ” de la serie. Agregando el componente de diferenciación al modelo ARMA, el modelo ARIMA, se expresa según la ecuación 29 :

$$ARIMA(p, d, q) : Y_d = c + \sum_{i=1}^p \phi_i Y_{d-i} + \sum_{i=1}^q \theta_i \epsilon_{t-i} + \epsilon_t \quad (29)$$

Donde:

- $ARIMA(p, d, q)$: define un proceso ARIMA con un proceso AR de orden “ p ”, un grado de diferenciación “ d ” y un proceso MA de orden “ q ”.
- Y_d : es la diferencia d de la serie Y_t .
- c : representa una constante (o deriva).
- p : define el número de rezagos de la regresión contra Y_t .
- ϕ_i : es el coeficiente del i – ésimo rezago de la serie.
- Y_{d-i} : es el i – ésimo rezago de la serie.
- q : define el número de términos de error pasado que se utilizarán en la ecuación.
- θ_i : es el coeficiente correspondiente de ϵ_{t-i} .
- $\epsilon_{t-q}, \dots, \epsilon_t$: son términos de error de ruido blanco.
- ϵ_t : representa el término de error, el cual es ruido blanco.

Los modelos AR, MA o ARMA pueden representarse con el modelo ARIMA, de la siguiente manera:

- El modelo ARIMA(0, 0, 0) es equivalente al ruido blanco.
- El modelo ARIMA(0, 1, 0) es equivalente a un paseo aleatorio.
- El modelo ARIMA(1, 0, 0) es equivalente a un proceso AR(1).
- El modelo ARIMA(0, 0, 1) es equivalente a un proceso MA(1).
- El modelo ARIMA(1, 0, 1) es equivalente a un proceso ARMA(1,1).

2.5.14. Métricas del error

Una vez que finalice el ajuste de un modelo, es hora de probar la capacidad del modelo para predecir observaciones que el modelo no vio antes (a diferencia de los valores ajustados que el modelo vio durante el proceso de entrenamiento). El método más común para evaluar el éxito del pronóstico con el fin de predecir los valores reales es usar métricas de precisión o error. El método más común para evaluar el éxito del pronóstico es predecir los valores reales con el uso de una métrica de error para cuantificar la precisión general del pronóstico. La selección de una métrica de error específica, depende de los objetivos de precisión del pronóstico [34]. Algunas métricas de error comunes son las siguientes:

- **Error cuadrático medio**

Esto cuantifica la distancia media al cuadrado entre los valores reales y pronosticados:

$$MSE = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2$$

El efecto cuadrático del error evita que los valores positivos y negativos se cancelen entre sí y penaliza la puntuación de error a medida que aumenta la tasa de error.

- **Raíz del error cuadrático medio**

Esta es la raíz de la distancia cuadrada promedio de los valores reales y pronosticados:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2}$$

Al igual que MSE, el RMSE produce errores amplificados debido al efecto cuadrático y, por lo tanto, es sensible a los valores atípicos.

- **Error absoluto medio**

Esto mide la tasa de error absoluta del pronóstico:

$$MAE = \frac{|Y_t - \hat{Y}_t|}{n}$$

De manera similar a MSE y RMSE, este método solo puede tener valores positivos. Esto es para evitar la cancelación de valores positivos y negativos. Por otro lado, no hay penalización de errores y, por lo tanto, este método no es sensible a los valores atípicos.

- **Error absoluto medio porcentual**

Mide el porcentaje de error absoluto promedio:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|$$

2.1.5. Una técnica para evaluar los modelos

- **Validación cruzada**

Una versión más sofisticada de conjuntos de entrenamiento/prueba es la validación cruzada de series de tiempo. En este procedimiento, hay una serie de conjuntos de prueba, cada uno de los cuales consta de una sola observación. El conjunto de entrenamiento correspondiente consta solo de observaciones que ocurrieron antes de la observación que forma el conjunto de prueba. Por lo tanto, no se pueden utilizar observaciones futuras para construir el pronóstico. Dado que no es posible obtener un pronóstico confiable basado en un pequeño conjunto de entrenamiento, las primeras observaciones no se consideran conjuntos de prueba [32].

La precisión del pronóstico se calcula promediando los conjuntos de prueba. Este procedimiento a veces se conoce como “evaluación sobre un origen de pronóstico continuo”, porque el “origen” en el que se basa el pronóstico avanza en el tiempo.

CAPÍTULO III MARCO METODOLÓGICO

3.1. Ubicación:

Los datos que se emplean en esta investigación, corresponden a valores del Índice de Precios de la Construcción (IPCO) de viviendas de unifamiliares y multifamiliares en Ecuador. El IPCO, es emitido mensualmente por el Instituto Nacional de Estadística y Censos del Ecuador (INEC).

Para su cálculo el INEC recopila información de 16 provincias ecuatorianas: Carchi, Imbabura, Pichincha, Esmeraldas, Santo Domingo de los Tsáchilas, Guayas, Los Ríos, Manabí, Santa Elena, Azuay, Cañar, Loja y El Oro [13].

3.2. Equipos y materiales:

Para la construcción del conjunto de datos, se utilizó un computador portátil con acceso a internet para consultar y descargar la información. Para la construcción del conjunto de datos se empleó el software Microsoft Excel. Las fuentes de información y las fechas de consultas de las variables explicativas candidatas y las variables explicadas, pueden revisarse en la Tabla 22, presentada en el Anexo 2. Los modelos estocásticos fueron programados en el software libre R [35], utilizando la interfaz RStudio (versión 2022.7.1.554) [36]. El código fuente desarrollado en esta investigación se presenta en el Anexo 3. El informe final de este trabajo fue realizado en el software Overleaf.

3.3. Tipo de investigación:

El enfoque de la investigación es del tipo cuantitativo, dadas las siguientes características:

- La investigación se fundamentó en la revisión analítica de literatura actual y relevante sobre el tema en cuestión, además ésta constituyó la directriz de este trabajo.
- La investigación está orientada al diseño de un modelo para explicar y predecir el IPCO de viviendas unifamiliares y multifamiliares, lo que indica que el problema es específico y está acotado.
- El modelo a diseñar contempla el involucramiento de variables exógenas y dos variables respuesta.
- El insumo principal para el diseño del modelo, es un conjunto de datos medidos y observados a lo largo de 171 meses.

- La hipótesis planteada requiere ser probada.

3.4. Hipótesis y pregunta de investigación

3.4.1. Planteamiento de la hipótesis

En base al objetivo de este trabajo, que se enfoca en el diseño de un modelo que explique a la variable respuesta (IPCO), se planteó la siguiente hipótesis:

- H0: No hay factores que explican la variación del índice de precios de construcción de viviendas unifamiliares y multifamiliares en Ecuador.
- H1: Existen factores que explican la variación del índice precios de construcción de viviendas unifamiliares y multifamiliares en Ecuador.

3.4.2. Planteamiento de la pregunta de investigación.

De la misma manera, la pregunta de investigación se fundamenta en el objetivo general de este trabajo, aquí se aborda el enfoque predictivo del modelo a diseñar.

- ¿Cuál es el modelo con mejor capacidad para predecir valores del IPCO de viviendas unifamiliares y multifamiliares en Ecuador, para el año 2023?

3.5. Población o muestra:

Dadas las características de este estudio, en el que la información a emplearse corresponde a series de tiempo, es importante indicar que, para el diseño de los modelos que expliquen y predigan valores del IPCO, es deseable contar con la mayor cantidad de datos posible. En tal virtud, para este estudio no corresponde hablar de población o muestra.

Sin embargo, en la Tabla 22, que se presenta en el Anexo 2, se indica la cantidad de datos para las variables explicativas candidatas y variables explicadas.

3.6. Recolección de información

Los datos empleados en este trabajo fueron obtenidos de tres fuentes: El Banco Central del Ecuador (BCE), Instituto Nacional de Estadística y Censos (INEC) y la Agencia Federal de Energía Americana (US-EIA).

En la Tabla 8, se muestran las variables y la fuente de donde se obtuvieron los datos. Este estudio consideró 14 variables, de las cuales los Índices de precios de la construcción de viviendas unifamiliares (IPCOU) y multifamiliares (IPCOM) son las variables respuesta, las doce variables restantes son consideradas como variables explicativas candidatas o predictores candidatos, cuya influencia sobre la variable dependiente será valorada más adelante.

Las variables se clasifican en cinco categorías relacionadas con el sector de la construcción, la economía ecuatoriana y aspectos sociales.

Table 8: Variables y sus fuentes de información.

Categoría	Sub-categoría	Variable	Descripción	Fuente
Estadísticas de edificaciones	Indicadores de la construcción	IPCOU	Índice de precios de la construcción de viviendas unifamiliares	INEC
		IPCOM	Índice de precios de la construcción de viviendas multifamiliares	INEC
		PCONS	Número de permisos de construcción otorgados	INEC
Estadísticas monetarias y financieras	Indicadores monetarios y financieros	OM	Oferta monetaria	BCE
	Tasa de interés	TI	Tasa de interés para el segmento inmobiliario	BCE
Estadísticas del sector exterior	Comercio exterior	IFOB	Importaciones FOB por uso o destino económico - Materias primas para materiales de construcción	BCE
		ICIF	Importaciones CIF por uso o destino económico - Materias primas para materiales de construcción	BCE
Estadísticas del sector real	Estadísticas petroleras	PPWTI	Precios promedio del crudo internacional - petróleo West Texas Intermediate	US-EIA
		PPCE	Precios promedio del crudo ecuatoriano cesta Oriente y Napo	BCE
	Precios y salarios	IPC	Índice de precios al consumidor	INEC
		IPP	Índice de Precios al Productor de Disponibilidad Nacional	INEC
	Producto interno bruto	PIB	Producto interno bruto	BCE
	Índices de confianza	ICC	Índice de confianza del consumidor	BCE
Estadísticas sociodemográficas y sociales	Desempleo	DESEM	Tasa de desempleo	INEC

Fuente: Mayra Viscaíno (2022)

Todas las variables son de tipo cuantitativo. En la Tabla 9, se indica infor-

mación adicional importante sobre las variables que se consideraron en este estudio.

3.7. Procesamiento de la información y análisis estadístico:

3.7.1. Metodología

La metodología empleada en este estudio, consta de las siguientes etapas:

1. Identificación de variables explicativas candidatas y recopilación de datos correspondientes: para este estudio son de interés las variables el índice de costos de la construcción de viviendas unifamiliares (IPCOU) y el índice de costos de la construcción de viviendas multifamiliares (IPCOM). Además, con el propósito de diseñar un modelo que involucre variables exógenas, a través de la literatura se identificaron 12 variables explicativas candidatas.
2. Estadísticos descriptivos de las variables explicadas y predictores candidatos: antes de diseñar los modelos, se considera necesario realizar un breve análisis de las series de interés, esto permitirá identificar valores mínimos, media y máximos de la serie; así como identificar si las series presentan estacionalidad, estacionariedad y qué variables exógenas están correlacionadas con las variables de interés.
3. Selección y construcción del conjunto de modelos candidatos: en base a la revisión de la literatura se identificaron los modelos univariados que han presentado buen desempeño, tanto para las variables respuesta como para las variables explicativas candidatas. Finalmente se conformó un conjunto de 14 modelos potenciales a ser probados.
4. Evaluación del rendimiento de predicción de los modelos candidatos: previo a la evaluación se generaron las predicciones y en un conjunto de datos de prueba de 24 meses se evaluó el rendimiento de los modelos. En base a la revisión de la literatura se seleccionaron a la RMSE y MAPE, como las métricas que permitirán evaluar a los modelos.
5. Selección del modelo con mayor capacidad predictiva de valores del IPCOU e IPCOM.
6. Generación de las predicciones del IPCOU e IPCOM para el año 2023.

En la Figura 5, se esquematizan las diferentes etapas de este estudio y algunos detalles que destacan en cada una de ellas.

Table 9: Cantidad de datos de las variables que intervienen en el modelado.

No	Variable	Unidad	Periodicidad	Observaciones
1	IPCOU	Tasa	Mensual	Año base 2000
2	IPCOM	Tasa	Mensual	Año base 2000
3	PIB	Millones de USD	Anual	Año base 2007. Los valores mensuales de este indicador se obtuvieron mediante interpolación.
4	IPP	Tasa	Mensual	Año base 2015
5	IPC	Porcentaje	Mensual	Año base 2014
6	OM	Millones de USD	Mensual	Ninguna
7	TI	Porcentaje	Mensual	Indicador del segmento de vivienda, sector inmobiliario.
8	ICIF	Miles de USD	Mensual	Se tomó la información de "Materias primas: Materiales de construcción".
9	IFOB	Miles de USD	Mensual	Se tomó la información de "Materias primas: Materiales de construcción".
10	PPEC	USD por barril	Mensual	Para el cálculo de la cesta de los crudos Oriente y Napo exportados por EP Petroecuador, los precios son ponderados, sin que signifique que los petróleos se mezclen para su exportación.
11	PPWTI	USD por barril	Mensual	El seguimiento del precio del WTI (West Texas Intermediate) es importante para la economía ecuatoriana, considerando que la información y análisis más relevante de los mercados especializados en petróleo, se centran principalmente en este tipo de crudo y se ha constituido en la referencia para la determinación de los precios de los crudos ecuatorianos.
12	ICC	Tasa	Mensual	Índice de confianza del consumidor nacional de 5 ciudades: Guayaquil, Quito, Cuenca, Ambato y Machala.
13	DESEM	Porcentaje	Semestral	Los valores mensuales de este indicador se obtuvieron mediante interpolación. Se asume el mismo valor de 11/2007 para 10/2007.
14	PCONS	Índice	Mensual	Ninguna

Fuente: Mayra Viscaíno (2022)

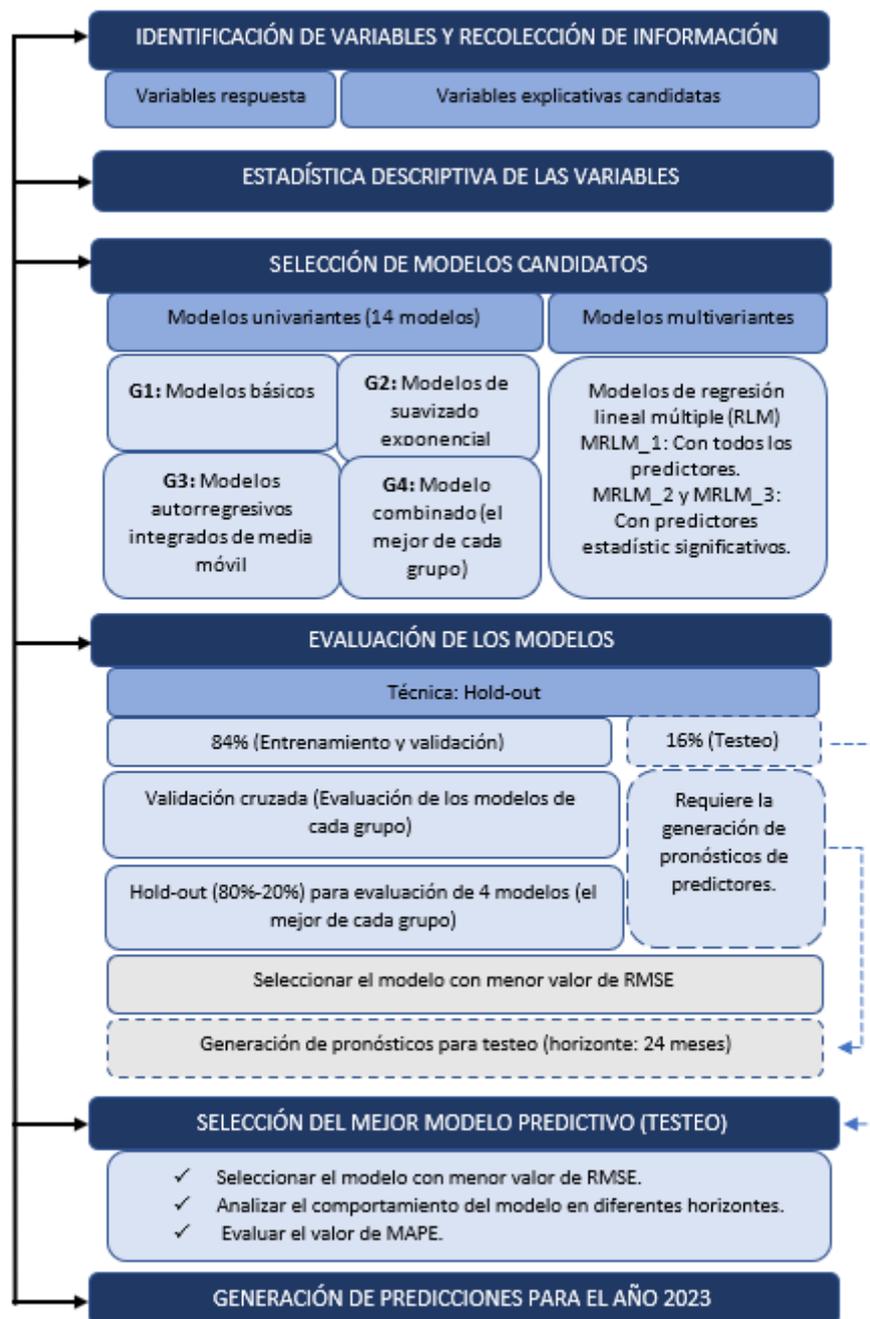


Figure 5: Metodología empleada en este estudio.
Fuente: Mayra Viscaíno (2022)

En la etapa de selección de los modelos candidatos, se conformó un conjunto de modelos potenciales, formado por cuatro grupos (ver Figura 6). Todos los modelos son probados con todas las series temporales de las variables respuesta y predictoras. Se empleó la técnica de validación cruzada para series de tiempo, para la evaluación de los modelos. De cada grupo se seleccionó un modelo, el que tuvo el menor valor de RMSE.

Un grupo particular de modelos es el número 4, compuesto por dos modelos.

El primero, resulta de la combinación de los modelos básicos; y el segundo es un modelo promedio resultante de combinar los modelos seleccionados de cada grupo (m_1, m_2, m_3). El modelo seleccionado es el que presenta el menor valor de RMSE y será el modelo m_4 .

Los cuatro modelos (m_1, m_2, m_3, m_4), son evaluados nuevamente, empleando la técnica de Hold-out. Finalmente, uno de los cuatro modelos, el que tenga el menor valor de RMSE, será seleccionado como modelo predictivo univariante, y será empleado para generar los pronósticos de la variable respuesta. Se siguió el mismo procedimiento para determinar el modelo predictor de cada variable explicativa.

Las predicciones generadas, se usaron para evaluar la capacidad predictiva del modelo de regresión lineal. De este proceso, se obtiene el primer modelo M_1 , que luego será comparado con el modelo de regresión lineal múltiple.

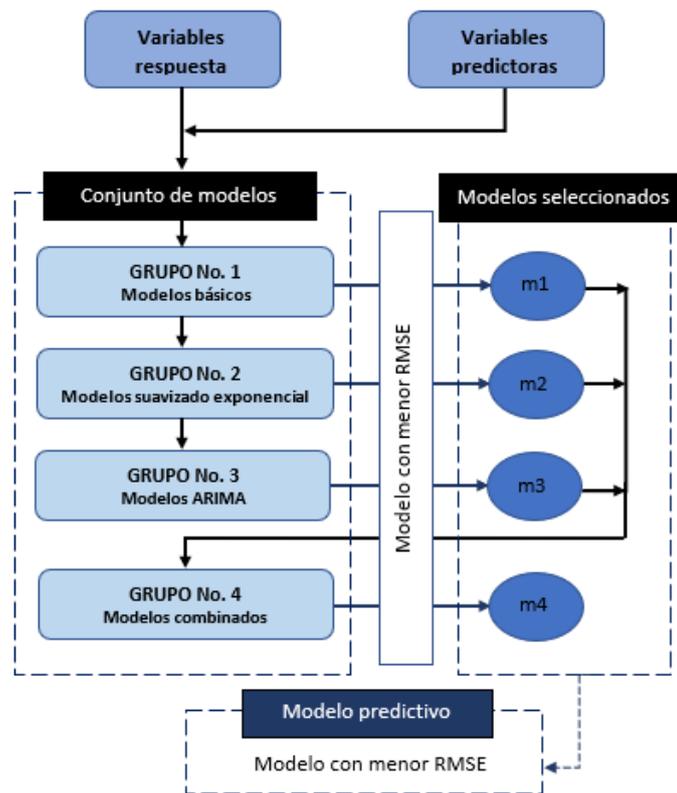


Figure 6: Conjunto de modelos candidatos.
Fuente: Mayra Viscaíno (2022)

3.8. Resultados alcanzados

A través de la aplicación de la metodología descrita se logró diseñar modelos estocásticos para explicar y predecir los índices de precios de viviendas en el Ecuador. Los resultados se presentan con mayor detalle en el capítulo IV RESULTADOS Y DISCUSIÓN.

CAPÍTULO IV. RESULTADOS Y DISCUSIÓN

4.1. Selección de variables respuesta y predictores candidatos

De acuerdo al objetivo de este trabajo, se seleccionaron dos variables respuesta que tienen relación con los precios de construcción de viviendas en el Ecuador y con base a la literatura revisada, se seleccionaron doce variables predictoras candidatas, esta información se detalla en la Tabla 10.

Table 10: Variables y tipo de variables que intervienen en este estudio.

No	Variable	Descripción	Tipo de variable
1	IPCOU	Índice de precios de la construcción de viviendas unifamiliares.	Respuesta
2	IPCOM	Índice de precios de la construcción de viviendas multifamiliares.	Respuesta
1	PIB	Producto interno bruto.	Predictora
2	IPP	Índice de Precios al Productor de Disponibilidad Nacional.	Predictora
3	IPC	Índice de precios al consumidor.	Predictora
4	OM	Oferta monetaria.	Predictora
5	TI	Tasas de interés vigentes para el segmento inmobiliario.	Predictora
6	ICIF	Importaciones CIF por uso o destino económico (Materias primas para materiales de construcción).	Predictora
7	IFOB	Importaciones FOB por uso o destino económico (Materias primas para materiales de construcción).	Predictora
8	PPEC	Precios promedio del crudo ecuatoriano cesta Oriente y Napo.	Predictora
9	PPWTI	Precios promedio del crudo internacional - petróleo West Texas Intermediate.	Predictora
10	ICC	Índice de confianza del consumidor nacional (Quito, Guayaquil, Cuenca, Ambato y Machala).	Predictora
11	DESEM	Tasa de desempleo.	Predictora
12	PCONS	Permisos de construcción otorgados a nivel nacional.	Predictora

Fuente: Mayra Viscaíno (2022)

4.2. Análisis descriptivo de las variables respuesta

4.2.1. Estadística descriptiva del IPCOU

La serie de tiempo del IPCOU (ver la Figura 7), presenta un valor mínimo de 134 al inicio de la serie en enero de 2002, tiene una tendencia creciente, lo que indica que no es una serie estacionaria. La serie alcanza un valor máximo de 280 en mayo de 2022. Su valor medio es de 222.

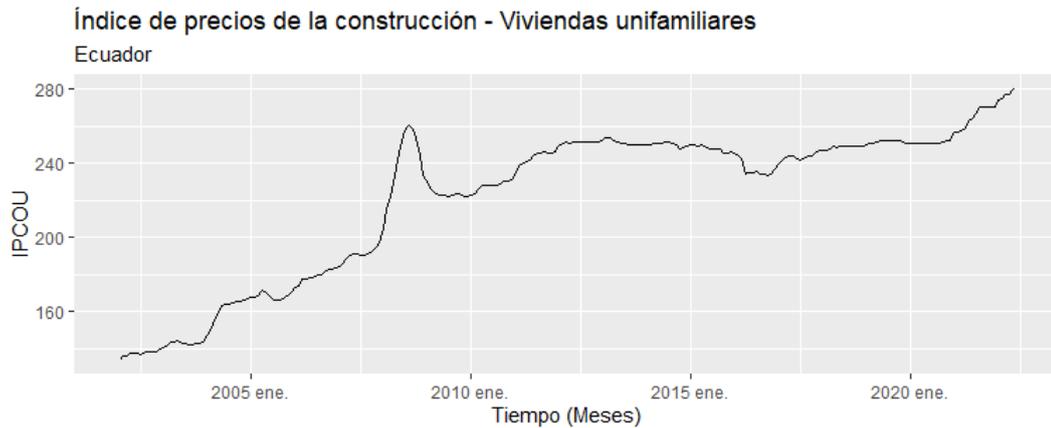


Figure 7: Serie de tiempo del IPCOU.
Fuente: Mayra Viscaíno (2022)

En la Figura 8, se observa que la serie cronológica no presenta estacionalidad. De la misma manera, en la Figura 9 que corresponde a un gráfico de subseries estacionales, se observa que la media (línea azul), permanece constante a lo largo de todo el año, lo cual puede considerarse como un indicativo de que no existe estacionalidad.

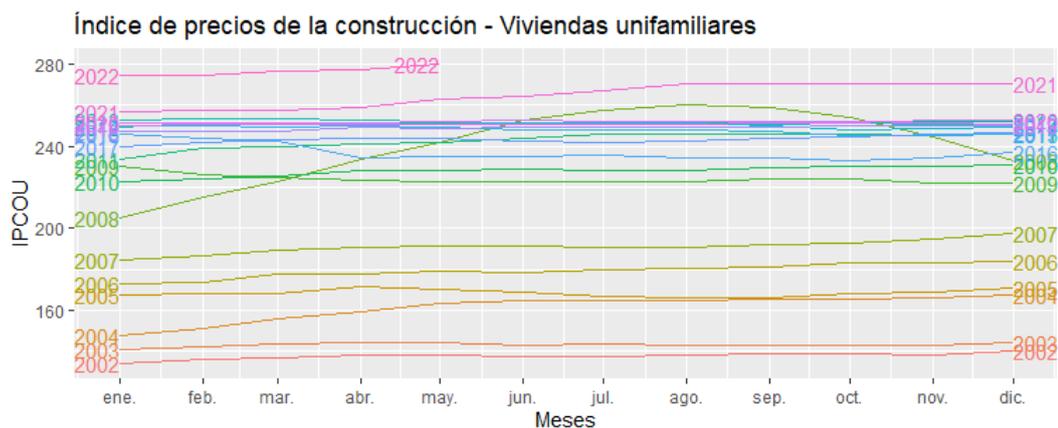


Figure 8: Gráfico de estacionalidad: Serie IPCOU
Fuente: Mayra Viscaíno (2022)

Así como la correlación mide el alcance de una relación lineal entre dos variables, la autocorrelación mide la relación lineal entre los valores rezagados de una serie de tiempo [32], en el autocorrelograma del IPCOU que se presenta

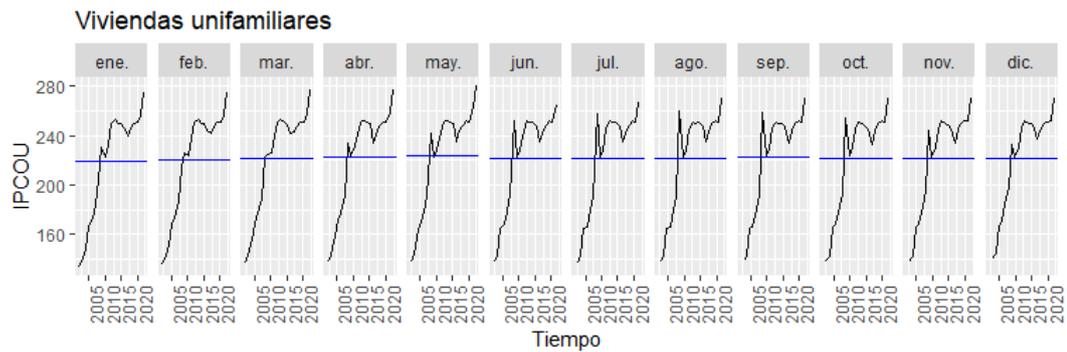


Figure 9: Gráfico de subserie estacional: Serie IPCOU
Fuente: Mayra Viscaíno (2022)

en la Figura 10. Las líneas azules discontinuas indican si las correlaciones son significativamente diferentes de cero, valores fuera de ese rango indican que existen valores autocorrelacionados en la serie más allá del retraso 25, con lo cual la serie no parece ser estacionaria.

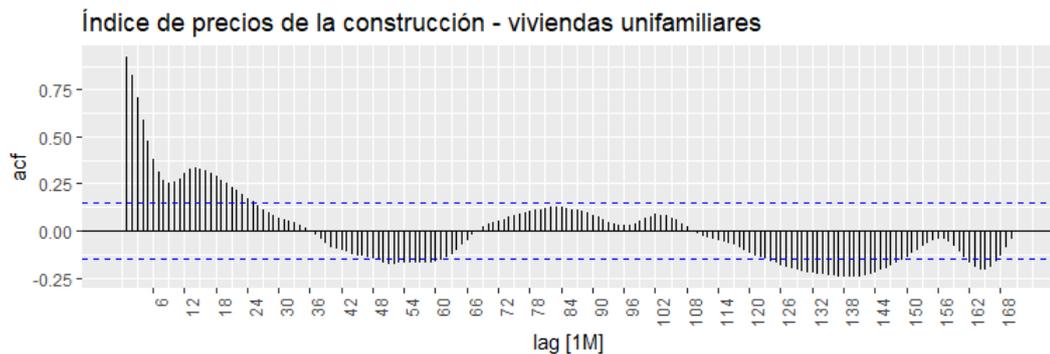


Figure 10: Función de autocorrelación del IPCOU
Fuente: Mayra Viscaíno (2022)

Cuando hay varias variables predictoras potenciales, como es el caso de este trabajo, resulta útil graficar cada variable frente a otra, observar los gráficos de dispersión y valorar la correlación existente. En la Figura 11, se busca observar la relación de las variables predictoras candidatas y la variable respuesta de interés (IPCOU).

De acuerdo a los valores de correlación que se indican en la Figura 11, se observa que el IPCOU tiene un valor de correlación alto (≥ 0.5), con variables exógenas como: PIB, IPP, IPC, OM y TI, y éstas se correlacionan negativamente, lo que indica que la relación es inversamente proporcional.

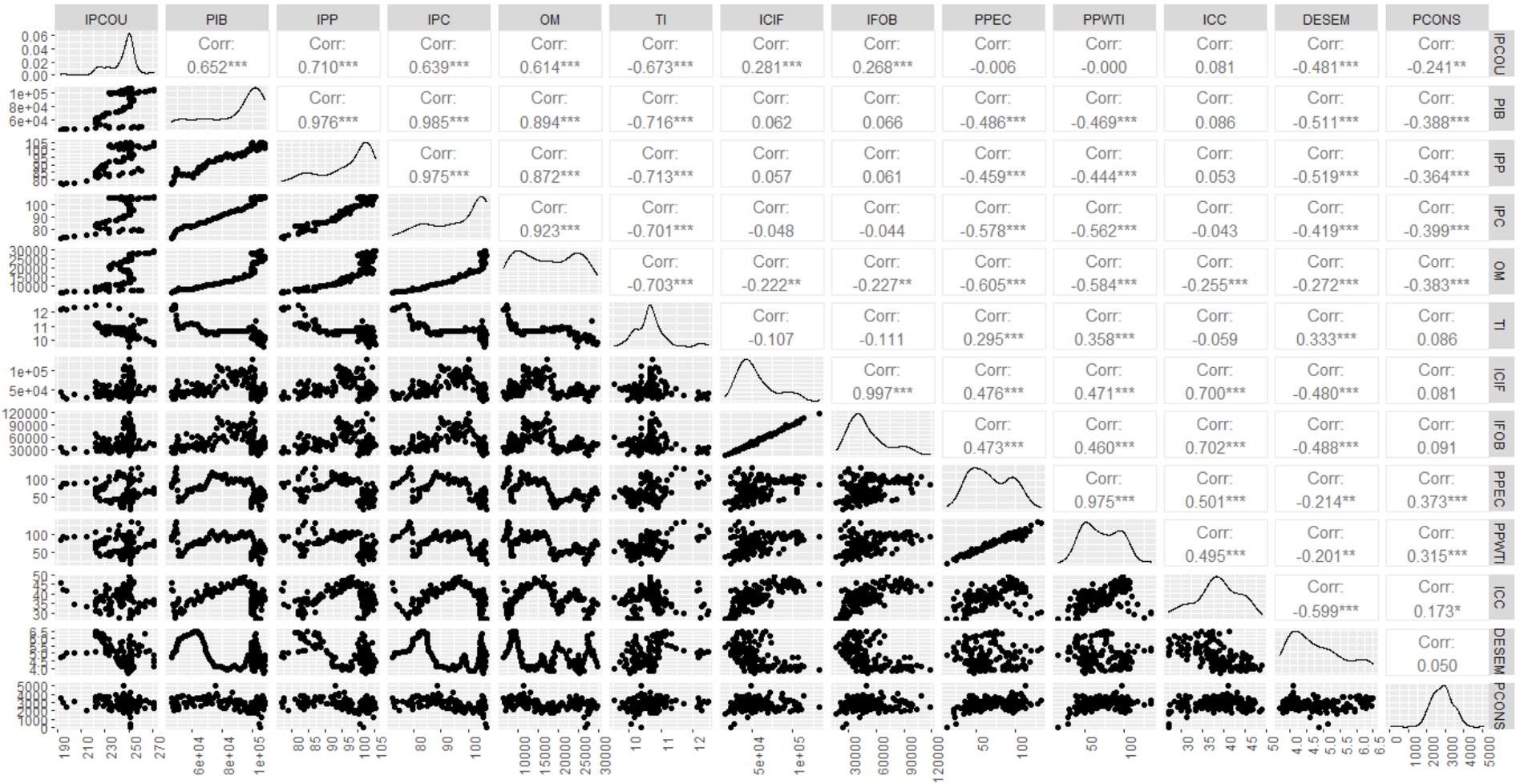


Figure 11: Matriz de diagramas de dispersión y correlaciones
Fuente: Mayra Viscaíno (2022)

4.2.2. Estadística descriptiva del IPCOM

La serie de tiempo IPCOM (Figura 12), presenta un valor mínimo de 127 al inicio de la serie en enero de 2002, tiene una tendencia creciente, alcanzando un máximo de 281 en mayo de 2022. Su valor medio es de 219. Al igual que la serie del IPCOU, esta serie no es estacionaria.

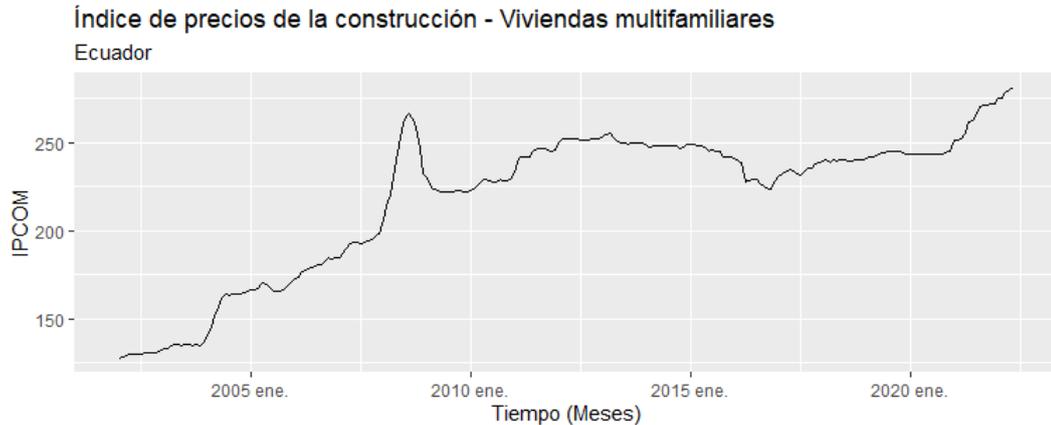


Figure 12: Serie de tiempo del IPCOM.
Fuente: Mayra Viscaíno (2022)

En las figuras 13 y 15, que son figuras que permiten apreciar gráficamente la estacionalidad, se observa que la serie IPCOM, no presenta estacionalidad.

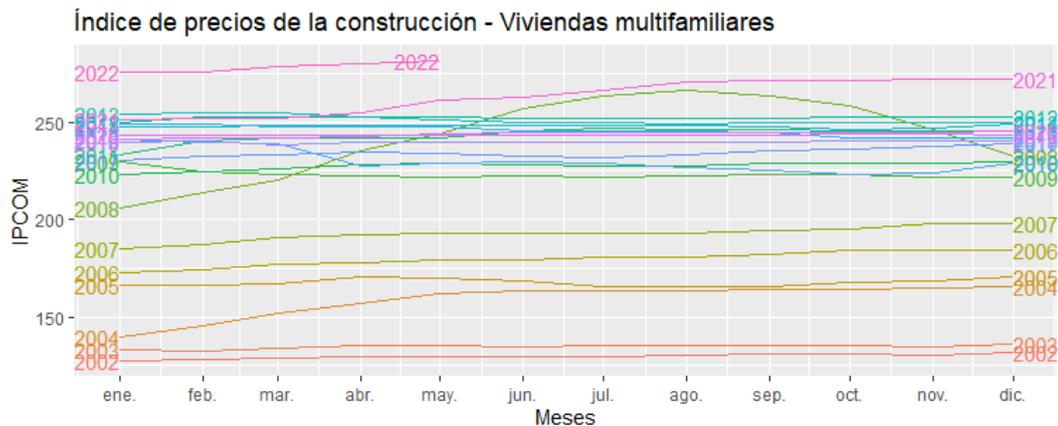


Figure 13: Gráfico de estacionalidad: Serie IPCOM
Fuente: Mayra Viscaíno (2022)

En la Figura 14, se puede observar la relación de las variables predictoras candidatas y la variable respuesta de interés (IPCOM). Según los valores de correlación que se indican en la Figura 14, se observa que el IPCOM tiene un valor de correlación alto (≥ 0.5), con variables exógenas como: IPP y TI cuya relación es negativa, lo que indica que a medida que el índice de precios al productor y la tasa de interés para el sector inmobiliario crece, el IPCOM decrece.

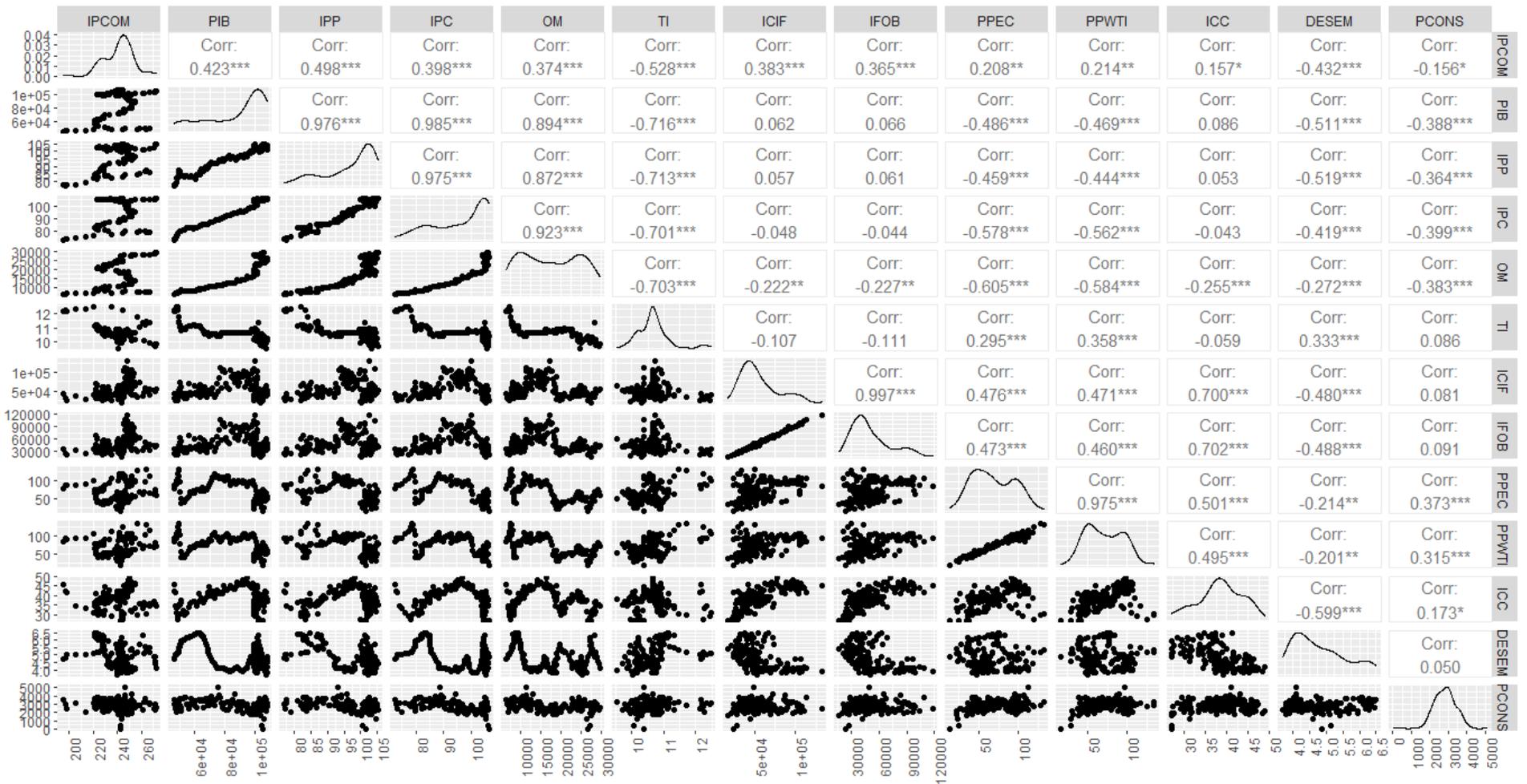


Figure 14: Matriz de diagramas de dispersión y correlaciones
Fuente: Mayra Viscaíno (2022)

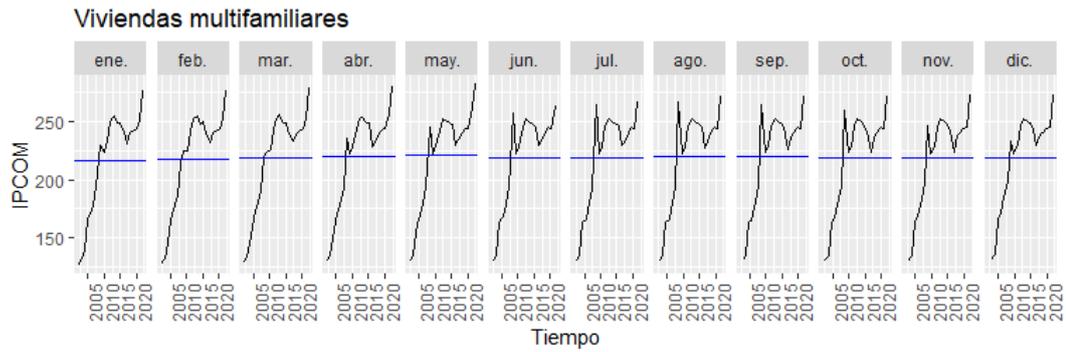


Figure 15: Gráfico de subserie estacional: Serie IPCOM
Fuente: Mayra Viscaíno (2022)

El correlograma del IPCOM, que se muestra en la Figura 16, indica que existen valores autocorrelacionados en la serie más allá del retraso 25, con lo cual esta serie tampoco parece ser estacionaria.

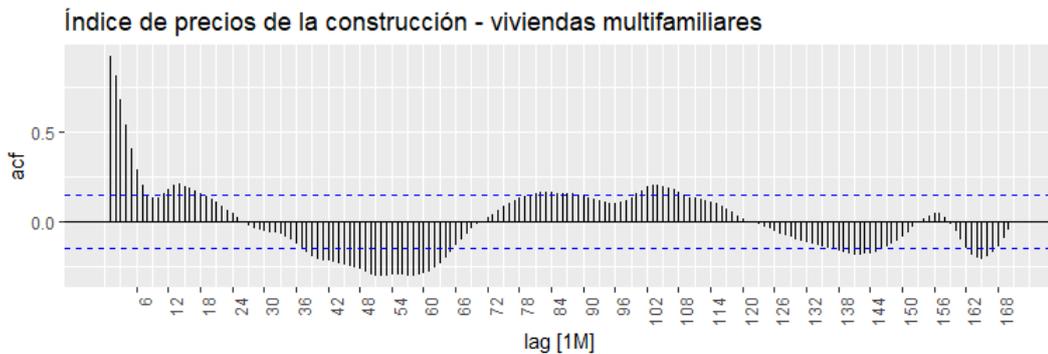


Figure 16: Función de autocorrelación del IPCOM
Fuente: Mayra Viscaíno (2022)

4.3. Selección de modelos candidatos

Luego de conocer algunas características de las series de tiempo de interés y en el proceso del diseño de modelos para predicción de los indicadores: IPCOU e IPCOM, es importante identificar el conjunto de modelos potenciales; éstos han sido clasificados en dos grupos: modelos univariados y multivariados.

4.3.1. Modelos univariados

Catorce modelos univariantes conforman el conjunto de modelos predictores candidatos (ver la Figura 17). Las series cronológicas de todas las variables (respuesta y predictoras), fueron modeladas y luego evaluado su desempeño, para finalmente seleccionar el modelo con menor valor de RMSE, como modelo predictivo para cada variable.



Figure 17: Conjunto de modelos predictores potenciales.
Fuente: Mayra Viscaíno (2022)

4.3.2. Modelos multivariados

Por otra parte, se formularon modelos de regresión lineal múltiple (MRLM) para las variables respuesta IPCOU e IPCOM.

Partiendo de un MRLM inicial que considera todas las variables regresoras candidatas (12 variables) para estimar los índices: IPCOU (ver ecuación 30) e IPCOM (ver ecuación 31), luego ajustándolos a un conjunto de 147 datos (octubre/2007 - diciembre/2019); se revisó qué variables son estadísticamente significativas con el 90% de confianza, de acuerdo a los resultados de la tabla de análisis de varianzas de coeficientes del modelo, tanto para el IPCOU (ver Tabla 11), como para el IPCOM (ver Tabla 12).

$$\begin{aligned}
 IPC\hat{O}U_i = & \hat{\beta}_0 + \hat{\beta}_1 PIB_i + \hat{\beta}_2 IPP_i + \hat{\beta}_3 IPC_i + \hat{\beta}_4 OM_i + \\
 & \hat{\beta}_5 TI_i + \hat{\beta}_6 ICIF_i + \hat{\beta}_7 IFOB_i + \hat{\beta}_8 PPEC_i + \\
 & \hat{\beta}_9 PPWTI_i + \hat{\beta}_{10} ICC_i + \hat{\beta}_{11} DESEM_i + \hat{\beta}_{12} PCONS_i
 \end{aligned} \quad (30)$$

$$\begin{aligned}
 IPC\hat{O}M_i = & \hat{\beta}_0 + \hat{\beta}_1 PIB_i + \hat{\beta}_2 IPP_i + \hat{\beta}_3 IPC_i + \hat{\beta}_4 OM_i + \\
 & \hat{\beta}_5 TI_i + \hat{\beta}_6 ICIF_i + \hat{\beta}_7 IFOB_i + \hat{\beta}_8 PPEC_i + \\
 & \hat{\beta}_9 PPWTI_i + \hat{\beta}_{10} ICC_i + \hat{\beta}_{11} DESEM_i + \hat{\beta}_{12} PCONS_i
 \end{aligned} \quad (31)$$

- **Modelo 1:** Para la estimación del IPCOU
Para el caso del IPCOU, como se observa en la Tabla 11. resultó que: *OM*, *PPEC* y *IFOB* no son estadísticamente significativas, definiéndose así los predictores que intervendrán en el modelo de regresión múltiple para la estimación del IPCOU, que se indica en la ecuación 32.

Table 11: Análisis de varianzas de coeficientes del modelo del IPCOU.

Series: IPCOU
Model: TSLM

Residuals:

	Min	1Q	Median	3Q	Max
	-12.59185	-3.02339	0.01005	3.74792	14.82595

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	2.415e+02	4.576e+01	5.276	5.22e-07	***
PIB	-1.366e-03	3.292e-04	-4.148	5.94e-05	***
IPP	8.402e-01	4.307e-01	1.951	0.053173	.
IPC	2.239e+00	5.859e-01	3.821	0.000203	***
OM	-2.783e-04	3.448e-04	-0.807	0.420997	
TI	-1.457e+01	2.067e+00	-7.048	8.90e-11	***
ICIF	6.794e-04	3.399e-04	1.999	0.047666	*
IFOB	-5.966e-04	3.701e-04	-1.612	0.109384	
PPEC	-1.411e-01	1.166e-01	-1.210	0.228381	
PPWTI	4.190e-01	1.267e-01	3.308	0.001211	**
ICC	-6.596e-01	2.153e-01	-3.064	0.002647	**
DESEM	-4.932e+00	1.275e+00	-3.868	0.000171	***
PCONS	-2.744e-03	1.064e-03	-2.580	0.010978	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.641 on 133 degrees of freedom
Multiple R-squared: 0.8137, Adjusted R-squared: 0.7955
F-statistic: 44.7 on 13 and 133 DF, p-value: < 2.22e-16

Fuente: Mayra Viscaíno (2022)

$$\begin{aligned}
\hat{IPCOU}_i = & \hat{\beta}_0 + \hat{\beta}_1 PIB_i + \hat{\beta}_2 IPP_i + \hat{\beta}_3 IPC_i + \\
& \hat{\beta}_4 TI_i + \hat{\beta}_5 ICIF_i + \hat{\beta}_6 PPWTI_i + \\
& \hat{\beta}_7 ICC_i + \hat{\beta}_8 DESEM_i + \hat{\beta}_9 PCONS_i
\end{aligned} \tag{32}$$

• **Modelo 2:** Para la estimación del IPCOM

De la misma manera para la variable de interés IPCOM, se verificó qué variables predictoras son estadísticamente significativas con el 90% de confianza (ver la Tabla 12), resultó que *IFOB* y *PPEC* no son estadísticamente significativas, con el nivel de confianza indicado. Para este caso se identificó que el modelo más parsimonioso es el que se indica en la ecuación 33.

Table 12: Análisis de varianzas de coeficientes del modelo del IPCOM.

```

Series: IPCOM
Model: TSLM

Residuals:
      Min       1Q   Median       3Q      Max
-14.9660  -3.8275  -0.1887   4.6803  18.3549

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.488e+02  5.347e+01   4.653 7.82e-06 ***
PIB          -1.388e-03  3.847e-04  -3.607 0.000437 ***
IPP           1.224e+00  5.032e-01   2.433 0.016308 *
IPC           1.962e+00  6.845e-01   2.866 0.004836 **
OM           -1.100e-03  4.028e-04  -2.731 0.007161 **
TI           -1.517e+01  2.415e+00  -6.281 4.44e-09 ***
ICIF         6.635e-04  3.971e-04   1.671 0.097140 .
IFOB        -5.772e-04  4.324e-04  -1.335 0.184225
PPEC        -1.761e-01  1.362e-01  -1.293 0.198316
PPWTI       4.580e-01  1.480e-01   3.095 0.002402 **
ICC         -6.669e-01  2.515e-01  -2.651 0.008989 **
DESEM       -5.650e+00  1.490e+00  -3.792 0.000226 ***
PCONS       -3.086e-03  1.243e-03  -2.483 0.014275 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.59 on 133 degrees of freedom
Multiple R-squared: 0.7464, Adjusted R-squared: 0.7216
F-statistic: 30.11 on 13 and 133 DF, p-value: < 2.22e-16

```

Fuente: Mayra Viscaíno (2022)

$$\begin{aligned}
 \hat{IPCOM}_i = & \hat{\beta}_0 + \hat{\beta}_1 PIB_i + \hat{\beta}_2 IPP_i + \\
 & \hat{\beta}_3 IPC_i + \hat{\beta}_4 TI_i + \hat{\beta}_5 ICIF_i + \\
 & \hat{\beta}_6 PPWTI_i + \hat{\beta}_8 DESEM_i + \hat{\beta}_9 PCONS_i
 \end{aligned} \tag{33}$$

4.4. Evaluación del rendimiento de predicción de los modelos

En esta sección se presentan los modelos predictivos seleccionados; así como las predicciones de las variables de interés y predictoras.

4.4.1. Selección de modelos con mejor rendimiento

En Tabla 13, se listan los modelos predictivos univariantes que presentaron menor valor de RMSE y que fueron seleccionados para generar predicciones.

Estos modelos fueron entrenados y validados con un conjunto de 147 datos (octubre/2007 - diciembre/2019), se empleó la técnica de validación cruzada, adaptada para series de tiempo, en el software RStudio. El código empleado se adjunta en los anexos, en el Anexo 3, puede revisarse el código usado para probar los catorce modelos estocásticos, y en el Anexo 4, se indica el código utilizado para probar los modelos de regresión lineal múltiple.

Table 13: Modelos univariantes seleccionados.

No	Variable	Modelo	RMSE
1	IPCOU	ARIMA(1,1,2)	1.46
2	IPCOM	ARIMA(2,1,2)	14.4
3	PIB	M1. Combinado[Naïve, Damped-Holt, ARIMA(0,2,0)(0,0,1)[12]]	6.49
4	IPP	Holt-Winter	1.69
5	IPC	Holt-Winter	0.527
6	TI	Drift	0.209
7	ICIF	Media	8897
8	PPWTI	ARIMA(0,1,2)	6.8
9	ICC	Naïve	1.31
10	DESEM	Naïve	0.453
11	PCONS	ARIMA(0,1,3)(1,0,0)[12]	375

Fuente: Mayra Viscaíno (2022)

4.4.2. Generación de pronósticos para las pruebas

Los valores previstos para las variables exógenas, se emplearán para las pruebas del MRLM propuesto y proceder a comparar su capacidad predictiva con los modelos univariantes para las variables de interés IPCOU e IPCOM.

En la Tabla 23, que se presenta en el Anexo 5, se indica la predicción de los valores de las variables respuesta y las variables exógenas que intervendrán en los MRLM, generados con los modelos indicados en la Tabla 13. A su vez, estos valores serán empleados para generar las predicciones del IPCOU e IPCOM con el modelo de regresión múltiple.

4.5. Evaluación de la capacidad predictiva de los modelos

Para las pruebas de los modelos: univariante y multivariante, se reservó el 16% de los datos, que corresponde a 24 meses (enero/2020 - diciembre/2021). Éstos son datos que los modelos no han visto y permitirán seleccionar el modelo con mejor capacidad predictiva.

4.5.1. Modelo para el IPCOU

En la Tabla 14 se indican los 24 valores reservados para la etapa de evaluación de los modelos para la predicción del IPCOU.

- **Modelo univariante**

De acuerdo a la Tabla 13. el modelo univariante con menor valor de *RMSE*, reportado por RStudio, para predecir la variable respuesta IPCOU que fue seleccionado, fue el modelo $ARIMA(p, d, q) \approx ARIMA(1, 1, 2)$, cuya modelo predictor está dado por la ecuación 34.

$$IPC\hat{O}U_{t+1} = 0.68866IPC\hat{O}U_t - IPC\hat{O}U_{t-1} - 0.0102\epsilon_{t-1} + 0.223\epsilon_{t-2} + \epsilon_t \quad (34)$$

Donde:

- $IPC\hat{O}U_{t+1}$: predicción del valor de IPCOU.
- $IPC\hat{O}U_t$: “*t* – ésimo” valor de la serie IPCOU diferenciada.
- $IPC\hat{O}U_{t-1}$: “*t* – 1” valor de la serie IPCOU diferenciada.
- $\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_t$: son términos de error de ruido blanco.

- **Modelo multivariante**

$$IPC\hat{O}U_i = 199.20 - 0.001294PIB_i + 1.08IPP_i + 2.175IPC_i - 12.46TI_i + 0.0001381ICIF_i + 0.2778PPWTI_i - 0.5517ICC_i - 4.157 * DESEM_i - 0.003389PCONS_i \quad (35)$$

Donde:

- *i*: representa el índice de las observaciones, donde $i = 1, \dots, N$
- $IPC\hat{O}U_i$: predicción del *i* – ésimo valor de IPCOU.
- PIB_i : *i* – ésimo valor del Producto Interno Bruto.
- IPP_i : *i* – ésimo valor del Índice de Precios del Productor.
- IPC_i : *i* – ésimo valor del Índice de Precios del Consumidor.
- TI_i : *i* – ésimo valor de la Tasa de Interés.
- $ICIF_i$: *i* – ésimo valor de Importaciones CIF.
- $PPWTI_i$: *i* – ésimo valor del Precio Promedio del Crudo de West Texas Intermediate.
- ICC_i : *i* – ésimo valor del Índice de Confianza al Consumidor.
- $DESEM_i$: *i* – ésimo valor de la Tasa de Desempleo.
- $PCONS_i$: *i* – ésimo valor del Número de Permisos de Construcción otorgados.

- **Métricas de evaluación**

En la Figura 18, se puede apreciar gráficamente el comportamiento de los dos modelos: $ARIMA(1,1,2)$ y MRLM para el IPCOU. Se observa que el modelo $ARIMA$, tiene mejor comportamiento que el MRLM en el corto plazo, aunque

Table 14: Evaluación de la capacidad predictiva de modelos para el IPCOU.

Fecha	$IPCOU_i$ Real	Modelo de regresión lineal			Modelo ARIMA(1,1,2)		
		$IP\hat{C}OU_i$	$(Y_i - \hat{Y}_i)^2$	$\left \frac{Y_i - \hat{Y}_i}{Y_i} \right $	$IP\hat{C}OU_i$	$(Y_i - \hat{Y}_i)^2$	$\left \frac{Y_i - \hat{Y}_i}{Y_i} \right $
2020 ene.	251.14	253.153	4.051	0.008	250.507	0.401	0.003
2020 feb.	251.10	255.771	21.821	0.019	250.116	0.969	0.004
2020 mar.	251.16	257.210	36.605	0.024	249.847	1.724	0.005
2020 abr.	250.80	256.402	31.382	0.022	249.663	1.293	0.005
2020 may.	250.80	255.952	26.538	0.021	249.536	1.597	0.005
2020 jun.	250.71	255.410	22.09	0.019	249.449	1.589	0.005
2020 jul.	250.64	254.914	18.267	0.017	249.390	1.563	0.005
2020 ago.	250.68	254.929	18.054	0.017	249.349	1.772	0.005
2020 sep.	251.17	256.035	23.666	0.019	249.321	3.420	0.007
2020 oct.	251.54	254.882	11.168	0.013	249.301	5.011	0.009
2020 nov.	252.68	255.134	6.021	0.010	249.288	11.504	0.013
2020 dic.	252.61	255.100	6.202	0.010	249.279	11.095	0.013
2021 ene.	256.69	256.889	0.040	0.001	249.273	55.014	0.029
2021 feb.	257.21	258.621	1.992	0.005	249.269	63.067	0.031
2021 mar.	257.40	260.112	7.355	0.011	249.266	66.168	0.032
2021 abr.	259.17	259.742	0.327	0.002	249.264	98.137	0.038
2021 may.	263.28	259.037	18.002	0.016	249.262	196.499	0.053
2021 jun.	264.18	258.523	32.000	0.021	249.261	222.569	0.056
2021 jul.	267.25	258.264	80.749	0.034	249.261	323.619	0.067
2021 ago.	270.44	258.086	152.629	0.046	249.26	448.586	0.078
2021 sep.	270.74	259.081	135.922	0.043	249.26	461.398	0.079
2021 oct.	270.71	257.889	164.374	0.047	249.26	460.119	0.079
2021 nov.	270.45	258.110	152.28	0.046	249.259	449.038	0.078
2021 dic.	270.77	258.141	159.484	0.047	249.259	462.707	0.079

NOTA: $IPCOU_i = Y_i$ y $IP\hat{C}OU_i = \hat{Y}_i$

Fuente: Mayra Viscaíno (2022)

luego es superado por el MRLM.

Se emplearon dos métricas el RMSE y el MAPE, considerando tres horizontes de pronóstico, como se indican en la Tabla 15.

Table 15: Resultados de RMSE y MAPE para modelos del IPCOU.

Horizonte	Modelo de regresión lineal		Modelo ARIMA(1,1,2)	
	RMSE	MAPE	RMSE	MAPE
Año 2020	4.34	0.83%	1.87	0.66%
Año 2021	8.69	2.65%	16.60	5.84%
2020-2021	6.86	2.15%	11.81	3.25%

Fuente: Mayra Viscaíno (2022)

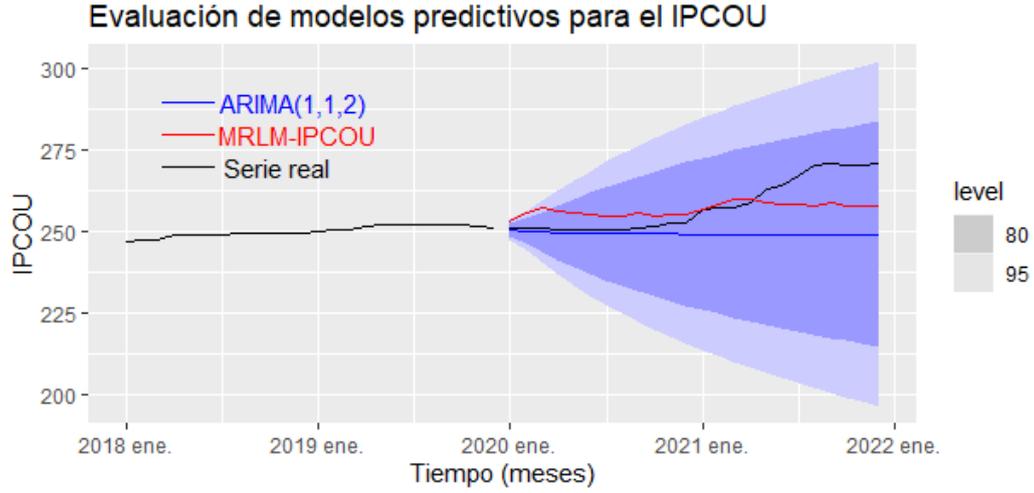


Figure 18: Comparación de las predicciones del $IP\hat{C}OU$ y la serie real.
Fuente: Mayra Viscaíno (2022)

4.5.2. Modelo para el IPCOM

En la Tabla 16 se indican los 24 valores reservados para la etapa de evaluación de los modelos para la predicción del IPCOM.

- **Modelo univariante**

De acuerdo a la Tabla 13. el modelo univariante con menor valor de $RMSE$, seleccionado para predecir valores del IPCOM, fue el modelo $ARIMA(p, d, q) \approx ARIMA(2, 1, 2)$, cuya modelo predictor está dado por la ecuación 36.

$$IP\hat{C}OM_{t+1} = 0.0855 + 2.5771IPCOM_t - 0.7351IPCOM_{t-1} - 1.0407\epsilon_{t-1} + 0.3757\epsilon_{t-2} + \epsilon_t \quad (36)$$

Donde:

- $IP\hat{C}OM_{t+1}$: predicción del valor de IPCOM.
- $IPCOM_t$: “ t – ésimo” valor de la serie IPCOM diferenciada.
- $IPCOM_{t-1}$: “ $t - 1$ ” valor de la serie IPCOM diferenciada.
- $\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_t$: son términos de error de ruido blanco.

- **Modelo multivariante**

$$IP\hat{C}OM_i = 128.2 - 0.001721PIB_i + 1.867IPP_i + 2.094IPC_i - 11.29TI_i + 0.0001526ICIF_i + 0.2782PPWTI_i - 3.575DESEM_i - 0.00365PCONS_i \quad (37)$$

Donde:

- i : representa el índice de las observaciones, donde $i = 1, \dots, N$
- \hat{IPCOM}_i : predicción del i –ésimo valor de IPCOM.
- PIB_i : i –ésimo valor del Producto Interno Bruto.
- IPP_i : i –ésimo valor del Índice de Precios del Productor.
- IPC_i : i –ésimo valor del Índice de Precios del Consumidor.
- TI_i : i –ésimo valor de la Tasa de Interés.
- $ICIF_i$: i –ésimo valor de Importaciones CIF.
- $PPWTI_i$: i –ésimo valor del Precio Promedio del Crudo de West Texas Intermediate.
- $DESEM_i$: i –ésimo valor de la Tasa de Desempleo.
- $PCONS_i$: i –ésimo valor del Número de Permisos de Construcción otorgados.

• Métricas de evaluación

En la Figura 19, se puede apreciar gráficamente el comportamiento de los dos modelos: ARIMA(2,1,2) y MRLM-IPCOM. Se observa que el modelo ARIMA, tiene mejor comportamiento que el MRLM en el horizonte de pronóstico de evaluación (24 meses).

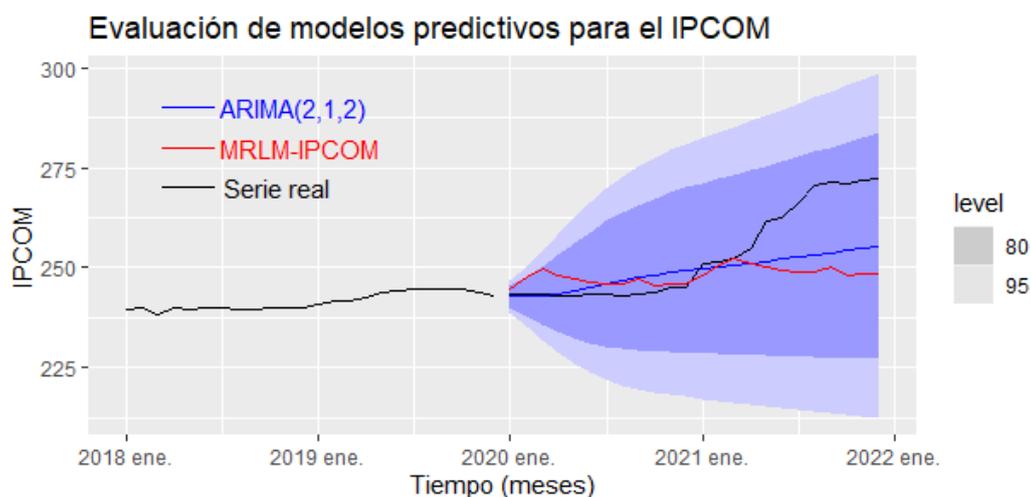


Figure 19: Comparación de las predicciones del \hat{IPCOM} y la serie real.
Fuente: Mayra Viscaíno (2022)

Finalmente, se evalúan los modelos con datos fuera de la muestra, para ello se usaron las métricas de RMSE y MAPE. Al igual que para la evaluación de los modelos para el IPCOU, se consideraron tres horizontes de pronóstico, como se indican en la Tabla 17.

4.6. Prueba de hipótesis y respuesta a pregunta de investigación

En la sección 3.4.1 se realizó el planteamiento de la hipótesis, como se indica a continuación:

Table 16: Evaluación de la capacidad predictiva de modelos para el IPCOM.

Fecha	$IPCOM_i$ Real	Modelo de regresión lineal			Modelo ARIMA(2,1,2)		
		$IPC\hat{O}M_i$	$(Y_i - \hat{Y}_i)^2$	$\left \frac{Y_i - \hat{Y}_i}{Y_i}\right $	$IPC\hat{O}M_i$	$(Y_i - \hat{Y}_i)^2$	$\left \frac{Y_i - \hat{Y}_i}{Y_i}\right $
2020 ene.	243.29	244.614	1.752	0.005	242.86	0.185	0.002
2020 feb.	243.27	247.778	20.326	0.019	242.813	0.209	0.002
2020 mar.	243.22	249.659	41.464	0.026	243.046	0.030	0.001
2020 abr.	243.05	247.946	23.975	0.020	243.534	0.234	0.002
2020 may.	243.07	247.119	16.392	0.017	244.217	1.315	0.005
2020 jun.	243.21	246.400	10.176	0.013	245.021	3.278	0.007
2020 jul.	243.23	245.937	7.328	0.011	245.872	6.981	0.011
2020 ago.	243.02	245.956	8.622	0.012	246.71	13.614	0.015
2020 sep.	243.30	247.345	16.36	0.017	247.49	17.556	0.017
2020 oct.	243.73	245.378	2.715	0.007	248.191	19.897	0.018
2020 nov.	244.91	245.840	0.865	0.004	248.807	15.189	0.016
2020 dic.	245.29	245.883	0.351	0.002	249.35	16.486	0.017
2021 ene.	251.26	248.220	9.242	0.012	249.839	2.019	0.006
2021 feb.	251.64	250.439	1.444	0.005	250.296	1.806	0.005
2021 mar.	252.17	252.373	0.041	0.001	250.743	2.036	0.006
2021 abr.	255.05	251.124	15.410	0.015	251.197	14.842	0.015
2021 may.	261.52	250.018	132.303	0.044	251.671	96.999	0.038
2021 jun.	262.63	249.324	177.042	0.051	252.17	109.419	0.040
2021 jul.	266.49	249.113	301.969	0.065	252.693	190.354	0.052
2021 ago.	270.68	248.920	473.496	0.08	253.238	304.233	0.064
2021 sep.	271.25	250.187	443.659	0.078	253.797	304.596	0.064
2021 oct.	271.11	248.171	526.21	0.085	254.365	280.395	0.062
2021 nov.	271.71	248.596	534.249	0.085	254.934	281.418	0.062
2021 dic.	272.28	248.705	555.786	0.087	255.501	281.543	0.062

NOTA: $IPCOU_i = Y_i$ y $IPC\hat{O}U_i = \hat{Y}_i$

Fuente: Mayra Viscaíno (2022)

Table 17: Resultados de RMSE y MAPE para modelos del IPCOM.

Horizonte	Modelo de regresión lineal		Modelo ARIMA(2,1,2)	
	RMSE	MAPE	RMSE	MAPE
Año 2020	3.54	1.28%	2.81	0.94%
Año 2021	16.26	5.06%	12.48	3.96%
2020-2021	11.76	3.17%	9.05	2.45%

Fuente: Mayra Viscaíno (2022)

- H0: No hay factores que explican la variación del índice de precios de construcción de viviendas unifamiliares y multifamiliares en Ecuador.
- H1: Existen factores que explican la variación del índice precios de construcción de viviendas unifamiliares y multifamiliares en Ecuador.

Las tablas 18 y 19, muestran los resultados de la prueba general del modelo de regresión lineal múltiple, propuesto en la ecuación 35 para predecir valores

Table 18: Análisis de varianzas de coeficientes del modelo del IPCOU.

```

Series: IPCOU
Model: TSLM

Residuals:
      Min       1Q   Median       3Q      Max
-11.3834  -3.6037   0.3102   3.3129  15.4485

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.992e+02  4.126e+01   4.829 3.62e-06 ***
PIBA         -1.294e-03  2.470e-04  -5.237 6.02e-07 ***
IPP          1.080e+00  3.715e-01   2.908 0.004241 **
IPC          2.175e+00  5.826e-01   3.733 0.000276 ***
TI           -1.246e+01  1.708e+00  -7.292 2.22e-11 ***
ICIF         1.381e-04  2.905e-05   4.753 5.01e-06 ***
PPWTI        2.778e-01  3.647e-02   7.617 3.85e-12 ***
ICC          -5.517e-01  1.940e-01  -2.844 0.005138 **
DESEM        -4.157e+00  1.079e+00  -3.854 0.000178 ***
PCONS        -3.389e-03  1.051e-03  -3.225 0.001576 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.714 on 137 degrees of freedom
Multiple R-squared:  0.8031, Adjusted R-squared:  0.7902
F-statistic: 62.09 on 9 and 137 DF, p-value: < 2.22e-16

```

Fuente: Mayra Viscaíno (2022)

del IPCOU y en la ecuación 37 para predecir valores del IPCOM.

Los resultados obtenidos son conducentes a rechazar la hipótesis de nulidad de parámetros en los dos casos. Para el caso del IPCOU, resultó que los parámetros: PIB, IPP, IPC, TI, ICIF, PPWTI, ICC, DESEM, PCONS, son estadísticamente significativos, con el 95% de confianza. Mientras que, para el IPCOM, los parámetros: PIB, IPP, IPC, TI, ICIF, PPWTI, DESEM, PCONS, son estadísticamente significativos, con el 95% de confianza.

Para los dos casos, el hecho de que los predictores resultaran significativos, indica que las variables propuestas pueden ser explicadas, en el caso del IPCOU por nueve variables exógenas, y para el IPCOM por ocho de ellas; de manera que la H0 planteada en este estudio, se rechaza.

Para dar respuesta a la pregunta de investigación *¿Cuál es el modelo con mejor capacidad para predecir valores de IPCO de viviendas unifamiliares y multifamiliares en Ecuador, para el año 2023?*. Se recurre a los resultados obtenidos en la etapa de pruebas presentados en las tablas 15 y 17.

Table 19: Análisis de varianzas de coeficientes del modelo del IPCOM.

```

Series: IPCOM
Model: TSLM

Residuals:
      Min       1Q   Median       3Q      Max
-15.5396  -4.2630  -0.2804   4.4557  17.1344

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.282e+02  3.912e+01   3.276  0.00133 **
PIB          -1.721e-03  2.583e-04  -6.665  5.85e-10 ***
IPP           1.867e+00  4.400e-01   4.243  4.02e-05 ***
IPC           2.094e+00  6.456e-01   3.243  0.00149 **
TI           -1.129e+01  1.995e+00  -5.660  8.41e-08 ***
ICIF         1.526e-04  3.278e-05   4.657  7.48e-06 ***
PPWTI        2.782e-01  4.324e-02   6.435  1.89e-09 ***
DESEM       -3.575e+00  1.248e+00  -2.864  0.00483 **
PCONS       -3.650e-03  1.243e-03  -2.936  0.00389 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.793 on 138 degrees of freedom
Multiple R-squared:  0.7204, Adjusted R-squared:  0.7042
F-statistic: 44.45 on 8 and 138 DF, p-value: < 2.22e-16

```

Fuente: Mayra Viscaíno (2022)

En el caso de la serie del IPCOU, se observa que el modelo ARIMA(1,1,2) presentó un valor menor de RMSE para el primer año, mientras que para el segundo año el MRLM tiene mejor capacidad predictiva. En base a estos resultados, se propone usar los dos modelos para la generación de pronósticos y usarlos como modelos complementarios.

Para la serie del IPCOM, el modelo ARIMA(2,1,2) demostró mejor capacidad predictiva al presentar el menor valor de RMSE, tanto para el primer año, como para el segundo. Por lo cual, se recomienda usar este modelo para generar las predicciones de IPCOM para el año 2023.

4.7. Generación de predicciones para el año 2023

4.7.1. Predicciones de valores del IPCOU

Para la generación de predicciones de IPCOU, que se presentan en la Tabla 20 y pueden observarse en la Figura 20, se empleó el modelo ARIMA(1,1,2)

para un horizonte de 12 meses (julio/2022 - jun/2023); y para las predicciones de julio/2023 a diciembre/2023, se utilizó el modelo multivariante propuesto en la Ecuación 35, con el propósito de reducir la incertidumbre en horizontes mayores a 12 meses.

Table 20: Estimaciones del IPCOU hasta diciembre-2023.

Fecha	Límite Inferior	Límite Superior	Media
2022 jul.	278.28	281.41	279.85
2022 ago.	276.78	282.72	279.75
2022 sep.	275.31	284.46	279.89
2022 oct.	274.01	286.34	280.18
2022 nov.	272.91	288.23	280.57
2022 dic.	271.97	290.10	281.03
2023 ene.	271.17	291.91	281.54
2023 feb.	270.50	293.66	282.08
2023 mar.	269.92	295.36	282.64
2023 abr.	269.44	296.99	283.21
2023 may.	269.02	298.57	283.80
2023 jun.	268.66	300.11	284.38
2023 jul.	273.75	300.67	287.21
2023 ago.	273.92	300.87	287.4
2023 sep.	274.42	301.46	287.94
2023 oct.	274.49	301.53	288.01
2023 nov.	274.79	301.88	288.33
2023 dic.	275.05	302.19	288.62

Fuente: Mayra Viscaíno (2022)

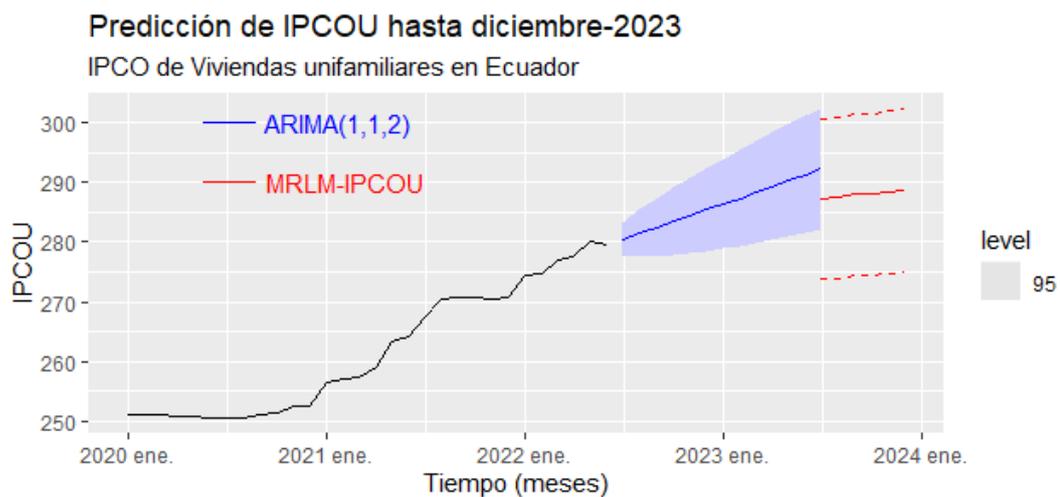


Figure 20: Gráfico de predicciones del IPCOU, h= 18 meses.

4.7.2. Predicciones de valores del IPCOM

En el caso del IPCOM, según las razones expuestas se seleccionó el modelo ARIMA(2,1,2) para obtener las predicciones para un horizonte de 18 meses; los valores pueden observarse en la Tabla 21, y su visualización en la Figura 21.

Table 21: Estimaciones del IPCOM hasta diciembre-2023.

Fecha	Límite Inferior	Límite Superior	Media
2022 jul.	276.16	280.29	278.22
2022 ago.	273.52	280.96	277.24
2022 sep.	270.93	281.93	276.43
2022 oct.	268.67	283.26	275.97
2022 nov.	266.94	284.93	275.94
2022 dic.	265.81	286.84	276.33
2023 ene.	265.23	288.88	277.05
2023 feb.	265.08	290.93	278.00
2023 mar.	265.24	292.89	279.06
2023 abr.	265.55	294.70	280.12
2023 may.	265.89	296.32	281.11
2023 jun.	266.2	297.75	281.98
2023 jul.	266.43	299.01	282.72
2023 ago.	266.56	300.13	283.34
2023 sep.	266.61	301.15	283.88
2023 oct.	266.6	302.13	284.37
2023 nov.	266.57	303.10	284.84
2023 dic.	266.55	304.08	285.31

Fuente: Mayra Viscaíno (2022)

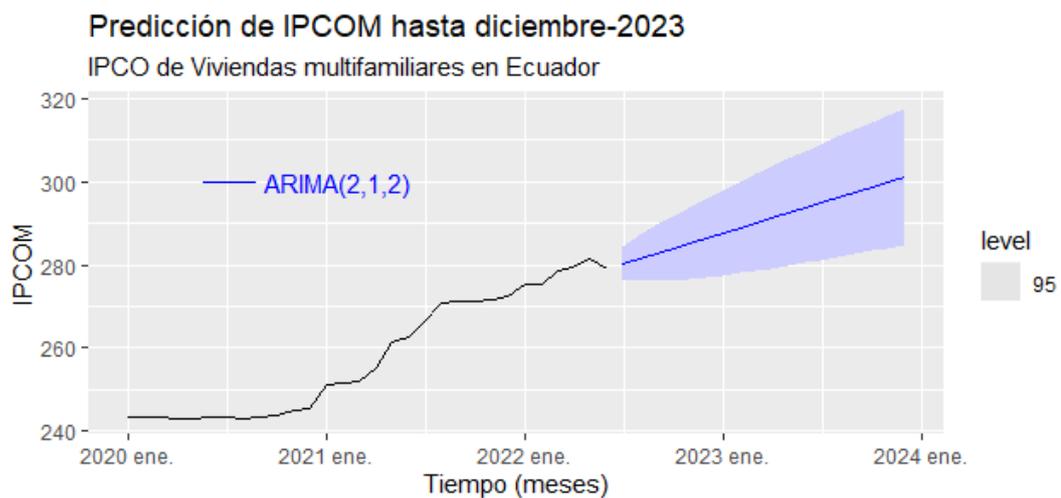


Figure 21: Gráfico de predicciones del IPCOM, h= 18 meses.

4.8. Discusión

El enfoque de este estudio es diseñar modelos para explicar y predecir valores de IPCO de viviendas en el Ecuador, basado en este objetivo se planteó una hipótesis y una pregunta de investigación, que cubre el doble propósito del estudio. La hipótesis nula, cubre el enfoque explicativo del modelo e indica que: “*No hay factores que explican la variación del índice de precios de construcción de viviendas unifamiliares y multifamiliares en Ecuador*”.

Mediante la prueba de hipótesis lineal general, conocida como “*Overall regression*”, para el caso del IPCOU y IPCOM, se obtuvo un $p\text{-valor} = 2.22e - 16$; para los dos casos, con lo cual se rechaza categóricamente con el 99% de confianza la hipótesis de nulidad simultánea de los coeficientes de las variables explicativas de los modelos de regresión lineal múltiple, tanto para IPCOU, como para IPCOM

Esto indica en términos generales, que al menos uno de los coeficientes de las variables es distinto de cero; es importante destacar que la prueba de hipótesis lineal general no considera la nulidad del intercepto β_0 . Con lo cual se puede garantizar que la variable respuesta es explicada al menos por una de las variables exógenas, y de manera más específica por las variables que resultan estadísticamente significativas con el 95% de confianza, como se detalló en las tablas 18 y 19.

Se ha verificado que existen variables exógenas que explican el comportamiento del índice de precios de la construcción (IPCO) de viviendas en el Ecuador. Comparando con un estudio similar [16], se concuerda en las siguientes variables exógenas: número de permisos de construcción otorgados (PCONS), el índice de precios al consumidor (IPC), la tasa de desempleo (DESM), el precio promedio del crudo West Texas Intermediate (PPWTI), el índice de precios al productor (IPP), el producto interno bruto (PIB), las importaciones de materiales de construcción (ICIF).

Tanto en este estudio como en [16], las variables que presentan valores de correlación superiores a 0.5 son: PIB, IPP, IPC. En los dos casos la variable DESEM, tiene una correlación negativa, lo que indica que cuando la variable DESEM crece el IPCO decrece.

En el caso de los modelos de regresión lineal múltiple, los valores de los coeficientes de determinación son considerados altos (≥ 0.5), en el caso del MRLM para el IPCOU el valor fue de 0.8031 y para el IPCOM, el valor fue de 0.7204. Los valores de este estadístico determinan que la proporción de variación de los resultados de la variable respuesta puede ser explicada por los regresores asociados a los modelos.

Es conocido que mientras mayor sea la cantidad de regresores que intervienen en un modelo, el valor del coeficiente de determinación tiende a ser más alto, esta puede ser la razón por la cual para el caso de IPCOM, se obtenga un

menor valor de R^2 , puesto que en el caso del MRLM para el IPCOU el número de predictores es de nueve, mientras que para el IPCOM es de ocho.

En el caso particular del Ecuador, las variables TI y DESEM, tienen una relación inversamente proporcional con el IPCOU y el IPCOM. Y en promedio, por cada unidad que TI decrece, el índice de precios de la construcción de viviendas se incrementa en 11.88 unidades. Mientras que por cada unidad que DESEM crece, el índice de precios de la construcción de viviendas se reduce en 2.13 unidades.

Por otra parte, la pregunta de investigación: *¿Cuál es el modelo con mejor capacidad para predecir valores del IPCO de viviendas unifamiliares y multifamiliares en Ecuador, para el año 2023?*, aborda la búsqueda del modelo con la mejor capacidad predictiva para un horizonte de pronóstico de 18 meses. En cuanto a modelos predictivos para el IPCO, se encontró en la revisión de la literatura que los modelos autorregresivos integrados de media móvil, son modelos que demuestran buena eficacia predictiva a corto plazo [16], esta afirmación es apoyada por este estudio, en base a los resultados obtenidos.

En el caso de los modelos para el IPCOU y el IPCOM, el promedio de los valores de MAPE obtenidos fueron de 3.25% y 2.45% respectivamente. Los modelos planteados para predecir el IPCOU y el IPCOM presentaron valores de MAPE inferiores al 10%, lo cual según [16], [19] y [27], demuestran una buena capacidad predictiva.

Este trabajo, empleó técnicas de aprendizaje de máquina, como: validación cruzada para series de tiempo y *Hold-out* para entrenar y validar a los modelos estadísticos, abordando así una brecha que según [16], no la abordó. Para investigadores interesados en este tema, se recomienda usar técnicas más sofisticadas para entrenar a los modelos como, la técnica de *bootstrapping*.

Se considera como una limitación de este estudio, el horizonte de pronóstico de 18 meses que cubren los modelos estocásticos aquí diseñados; sin embargo, esto puede ser un asunto cuestionable y discutible, ya que se debe valorar el contexto en donde se emplazan este tipo de obras. En el contexto local, proyectos constructivos de viviendas unifamiliares y multifamiliares no superan los 18 meses en construirse. En este sentido, conviene valorar si es necesario y útil generar pronósticos para horizontes mayores.

Se encontró en la revisión de la literatura, estudios que proponen modelos predictivos con horizontes a largo plazo como: 60 [24], 84 [8] y 96 [21] meses. Sin embargo, para nuevos investigadores se recomienda estudiar el contexto local, en el cual se pretenden emplear los valores pronosticados de IPCO y en función de ello evaluar cuál es el horizonte más apropiado, para el diseño de sus modelos. Recordando que la incertidumbre de las predicciones es mayor para horizontes de pronóstico más largos.

CAPÍTULO V. CONCLUSIONES Y REFERENCIAS BIBLIOGRÁFICAS

5.1. Conclusiones

El IPCO para viviendas unifamiliares y multifamiliares en Ecuador, puede ser pronosticado usando modelos univariados y multivariados, existe la información disponible en fuentes de información como el INEC, BCE y US-EIA, que proporcionaron información de 12 variables predictoras candidatas y dos variables respuesta. En el caso de las variables PIB y DESEM, fue necesario realizar interpolación de los datos, ya que la periodicidad con la que se generan estos indicadores es anual y semestral, respectivamente. El proceso de recopilación de información tomó un tiempo considerable, lo cual es un aspecto negativo para el uso de MRLM.

Las variables que resultaron estadísticamente significativas para realizar predicciones del IPCOU e IPCOM, fueron estadísticas reportadas por las diferentes fuentes (BCE, INEC, US-EIA), en las siguientes áreas: monetarias y financieras (TI), del comercio exterior (ICIF), del sector real (PPWTI, PIB, IPP, IPC, ICC), sociodemográficas y sociales (DESEM) y de edificaciones (PCONS).

Del análisis descriptivo de las series de tiempo IPOCU e IPCOM, se observó que las series tienen una tendencia creciente, son series no estacionarias. Existen variables exógenas que presentan altos valores de correlación positiva como: PIB(0.652), IPC(0.639), OM(0.614) y IPP(0.710), lo que indica que la relación es directamente proporcional, mientras estas variables crezcan, las variables respuesta (IPCOU e IPCOM), también crecerán.

Apesar de ello OM, que tiene un valor de correlación “alto”, no resultó estadísticamente significativo para el modelo. Y variables con coeficiente de correlación nulo como el PPWTI, resultó estadísticamente significativa. Se debe recordar que valores bajos de correlación, indican que las variables no se relacionan linealmente.

Para la generación de las predicciones de las variables respuesta (2 variables) y las variables regresoras candidatas (9 variables), se formó un conjunto de modelos potenciales conformado por modelos univariantes y modelos de regresión múltiple (sólo para variables respuesta). Resultó que los modelos simples, como: Drift, Media, Naive, tuvieron valores bajos de RMSE para pronosticar TI, ICIF, ICC y DESEM. Para las variables: IPCOU, IPCOM, PPWTI y PCNOS, los modelos ARIMA presentaron buena capacidad predictiva. Para

dos variables (IPP, IPC) el modelo de SES, Holt-Winter fue el mejor. Sólo para un caso (PIB), resultó eficiente un modelo combinado entre: un modelo simple, ARIMA y SES.

De los 14 modelos estocásticos que fueron probados, los que demostraron efectividad predictiva para la variable respuesta IPCOU fueron los modelos ARIMA(1,1,2) y modelo de regresión múltiple con nueve predictores (PIB, IPP, IPC, TI, ICIF, PPWTI, ICC, DESEM, PCONS). Para el IPCOM, el modelo que mejor desempeño tuvo fue el modelo ARIMA(2,1,2).

Para la etapa de diseño y evaluación de la capacidad predictiva de los modelos, se dividió el conjunto de datos en dos partes, tomando en cuenta el 84% de los datos (123 meses), para entrenar los modelos utilizando validación cruzada para series de tiempo, que resultó una técnica apropiada para la cantidad de datos disponibles y el 16% (24 meses) para realizar evaluaciones de predicción fuera de la muestra y comparar el desempeño de la predicción.

Con los modelos indicados, se generaron las predicciones de los valores de IPCOU e IPCOM, hasta el año 2023, lo que implica un horizonte de predicción de 18 meses. En el caso del IPCOU, se emplearon dos modelos. El modelo ARIMA(1,1,2) cubrió los 12 primeros meses, ya que tiene efectividad predictiva a corto plazo, y se usó el modelo de regresión lineal múltiple para generar las predicciones desde el mes 13 hasta el mes 18. Los valores generados, representan la media de los valores pronosticados con un intervalo de confianza del 95%.

Los modelos predictivos para el IPCOU e IPCOM propuestos, consideran la incertidumbre en los resultados de predicción, pues se proponen predicciones por intervalos. En los resultados de los pronósticos del IPCOU e IPCOM, para el año 2023, se indicó el límite superior e inferior del rango en el que se espera que la variable de salida se encuentre con un nivel de confianza del 95%. El ancho del intervalo crece a medida que se incrementa el horizonte de pronóstico.

5.2. Referencias bibliográficas

- [1] P. T. Nguyen and Q. Le Hoang Thuy To Nguyen, “Critical factors affecting construction price index: An integrated fuzzy logic and analytical hierarchy process,” *Journal of Asian Finance, Economics and Business*, vol. 7, no. 8, pp. 197–204, 2020.
- [2] Banco Central del Ecuador, “Boletín del índice de actividad económica coyuntural no. 37 1.-,” no. 37, 2021.
- [3] L. Zhao, J. Mbachu, and H. Zhang, “Forecasting residential building costs in New Zealand using a univariate approach,” *International Journal of Engineering Business Management*, vol. 11, pp. 1–13, 2019.
- [4] S. Moradi, R. Ansari, and R. Taherkhani, “A Systematic Analysis of Construction Performance Management: Key Performance Indicators from 2000 to 2020,” *Iranian Journal of Science and Technology - Transactions of Civil Engineering*, vol. 46, no. 1, pp. 15–31, 2022.
- [5] K. Lucero, “La construcción, un pilar de la economía debilitado por la pandemia,” 2020.
- [6] Organización Internacional del Trabajo, “El impacto de la COVID-19 en el sector de la construcción,” *Organización Internacional del Trabajo*, p. 14, 2021.
- [7] A. M. Sánchez, T. Vayas, F. Mayorga, and C. Freire, “Sector Construcción Ecuador,” vol. 2018, pp. 1–4, 2020.
- [8] Y. Elfahham, “Estimation and prediction of construction cost index using neural networks, time series, and regression,” *Alexandria Engineering Journal*, vol. 58, pp. 499–506, 2019.
- [9] J. Dong, Y. Chen, and G. Guan, “Cost Index Predictions for Construction Engineering Based on LSTM Neural Networks,” *Advances in Civil Engineering*, vol. 2020, 2020.
- [10] S. Moon, S. Chi, and D. Y. Kim, “Predicting Construction Cost Index Using the Autoregressive Fractionally Integrated Moving Average Model,” *Journal of Management in Engineering*, vol. 34, no. 2, p. 04017063, 2018.
- [11] ENR(Engineering News-Record), “Construction Cost Index History - As of June 2022,” 2022.

- [12] T. Moon and D. H. Shin, “Forecasting Construction Cost Index Using Interrupted Time-Series,” *KSCE Journal of Civil Engineering*, vol. 22, no. 5, pp. 1626–1633, 2018.
- [13] INEC, “Manual Del Investigador y Supervisor IPCO Índice de Precios de la Construcción,” 2021.
- [14] A. Kamal, M. Abas, D. Khan, and R. W. Azfar, “Risk factors influencing the building projects in Pakistan: from perspective of contractors, clients and consultants,” *International Journal of Construction Management*, vol. 22, no. 6, pp. 1141–1157, 2022.
- [15] S. H. Ji, J. Ahn, H. S. Lee, and K. Han, “Cost Estimation Model Using Modified Parameters for Construction Projects,” *Advances in Civil Engineering*, vol. 2019, 2019.
- [16] F. Jiang, J. Awaitey, and H. Xie, “Analysis of construction cost and investment planning using time series data,” *Sustainability (Switzerland)*, vol. 14, 2022.
- [17] Barbara Kitchenham and S. Charters, “Guidlines for performing Systematic Literature Riviws in Software Engineering,” 2007.
- [18] M. Mir, H. M. Kabir, F. Nasirzadeh, and A. Khosravi, “Neural network-based interval forecasting of construction material prices,” *Journal of Building Engineering*, vol. 39, no. February, p. 102288, 2021.
- [19] S. A. M. Faghih, Y. Gholipour, and H. Kashani, “Time Series Analysis Framework for Forecasting the Construction Labor Costs,” *KSCE Journal of Civil Engineering*, vol. 25, no. 8, pp. 2809–2823, 2021.
- [20] C.-Y. Choi, K. R. Ryu, and M. Shahandashti, “Predicting City-Level Construction Cost Index Using Linear Forecasting Models,” *Journal of Construction Engineering and Management*, vol. 147, no. 2, pp. 1–12, 2021.
- [21] Y. Cao and B. Ashuri, “Predicting the Volatility of Highway Construction Cost Index Using Long Short-Term Memory,” *Journal of Management in Engineering*, vol. 36, no. 4, p. 04020020, 2020.
- [22] O. Swei, “Forecasting Infidelity: Why Current Methods for Predicting Costs Miss the Mark,” *Journal of Construction Engineering and Management*, vol. 146, no. 2, p. 04019100, 2020.
- [23] J. Ahn, S. H. Ji, S. J. Ahn, M. Park, H. S. Lee, N. Kwon, E. B. Lee, and Y. Kim, “Performance evaluation of normalization-based CBR models for improving construction cost estimation,” *Automation in Construction*, vol. 119, no. January 2017, p. 103329, 2020.
- [24] K. Pakalapati, M. Khalafalla, and J. Rueda-Benavides, “Using moving-window cross-validation algorithm to optimize bid-based cost estimating data usage,” *International Journal of Construction Management*, vol. 0, no. 0, pp. 1–9, 2020.

- [25] D. Chakraborty, H. Elhegazy, H. Elzarka, and L. Gutierrez, “A novel construction cost prediction model using hybrid natural and light gradient boosting,” *Advanced Engineering Informatics*, vol. 46, no. November, p. 101201, 2020.
- [26] S. Mao and F. Xiao, “A novel method for forecasting Construction Cost Index based on complex network,” *Physica A: Statistical Mechanics and its Applications*, vol. 527, p. 121306, 2019.
- [27] S. A. M. Faghih and H. Kashani, “Forecasting Construction Material Prices Using Vector Error Correction Model,” *Journal of Construction Engineering and Management*, vol. 144, no. 8, pp. 1–12, 2018.
- [28] M. H. Rafiei and H. Adeli, “Novel Machine-Learning Model for Estimating Construction Costs Considering Economic Variables and Indexes,” *Journal of Construction Engineering and Management*, vol. 144, no. 12, p. 04018106, 2018.
- [29] R. Zhang, B. Ashuri, Y. Shyr, and Y. Deng, “Forecasting Construction Cost Index based on visibility graph: A network approach,” *Physica A: Statistical Mechanics and its Applications*, vol. 493, pp. 239–252, 2018.
- [30] T. Moon and D. H. Shin, “Forecasting Model of Construction Cost Index Based on VECM with Search Query,” *KSCE Journal of Civil Engineering*, vol. 22, no. 8, pp. 2726–2734, 2018.
- [31] O. S. Oshodi, O. A. Ejohwomu, I. O. Famakin, and P. Cortez, “Comparing univariate techniques for tender price index forecasting: Box-Jenkins and neural network model,” *Construction Economics and Building*, vol. 17, no. 3, pp. 109–123, 2017.
- [32] R. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice, 3rd edition*. 2021.
- [33] N. Ramón and J. López, *Econometría: Series Temporales y Modelos de Ecuaciones Simultáneas*. 2016.
- [34] R. Krispin, *Hands-On Time Series Analysis with R : Perform Time Series Analysis and Forecasting Using R*. 2019.
- [35] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2022.
- [36] RStudio Team, *RStudio: Integrated Development Environment for R*. RStudio, PBC, Boston, MA, 2022.

ANEXOS

Anexo 1: Extracción de información de estudios revisados.

Nombre del revisor: Mayra Velasco										Fecha de la revisión:		Inicio: 20-Quime-2022		Fin: 20-Agosto-2022			
CODIGO	TITULO	AUTORES	FECHA DE PUBLICACION	REVISTA	TIPO DE REVISOR	LINK	OBJETIVO	ÁMBITO DE APLICACIÓN	MODELOS DE PROGNÓSTICO		VARIABLES INDEPENDIENTES VARIABLES EXPLICATIVAS / PREDICTORAS	MÉTRICA DE EVALUACIÓN	MEJOR MODELO PARA PREDICCIÓN DE ICC	HORIZONTE DE PREDICCIÓN	OBSERVACIONES		
									MODELOS ESTADÍSTICOS	ML							
P1	Analysis of Construction Cost and Investment Planning Using Time Series Data	Jiang, F., Jia, Y., Xie, H.	1-Feb-22	Sustainability	ML	https://doi.org/10.3390/su13020205	Analizar los costos de construcción desde el punto de vista de los modelos de predicción de costos multivariados en la predicción del índice de costos de construcción (ICC)	ICC	1. Modelo ARIMA para ICC, 2. Modelo de regresión, 3. Pronóstico de variables independientes		1. Índice de costos de la construcción (ICC) 2. Índice de precios al consumidor (IPC00 976) 3. Tasa de desempleo en los EE. UU. (U3) 4. Tasa de cambio del dólar - EE. UU. (CONUSDOL) 5. Índice de precios al productor (PP03 823) 6. Precio del petróleo crudo (OIL00 093) 7. Producto interno bruto (PIB) (U033) 8. Precio del consumo de electricidad (ELEC00 023) 9. Índice de precios al consumidor (CPI) (U033) 10. Índice de precios al productor (PPI) (U033) 11. Índice de precios al consumidor (CPI) (U033) 12. Índice de precios al productor (PPI) (U033) 13. Índice de precios al consumidor (CPI) (U033) 14. Índice de precios al productor (PPI) (U033) 15. Índice de precios al consumidor (CPI) (U033) 16. Índice de precios al productor (PPI) (U033) 17. Índice de precios al consumidor (CPI) (U033) 18. Índice de precios al productor (PPI) (U033) 19. Índice de precios al consumidor (CPI) (U033) 20. Índice de precios al productor (PPI) (U033)	RMS, MAPE, MAE, R-squared	ARIMA	18 MESES	1. No indica el valor de desviación estándar		
P2	Neural network-based interval forecasting of construction material prices	Nofreza Nair, M. Dipo Kabir, Farhad Akbari, Amirhossein Ghobadipour	19-Feb-21	Journal of Building Engineering	Other	https://doi.org/10.1016/j.jobe.2021.103448	Proposición de un método novedoso para la predicción de precios de materiales.	ICC-ÍNDICE DE COSTOS DE CONSTRUCCIÓN	Modelos de predicción VEC (vectores de corrección de error vectoriales)		RMS, MAPE, MAE, U-Stat	LUBE-ANN	5 MESES	RMS y MAPE no indicados			
P3	Time Series Analysis Framework for Forecasting the Construction Labor Costs	Mohamed Elmaghrabi, Mohamed Elmaghrabi, Mohamed Elmaghrabi, Mohamed Elmaghrabi	5-mar-2021	KSCJ Journal of Civil Engineering	Other	https://doi.org/10.1016/j.kscj.2021.03.001	Desarrollar modelos que pronostican los movimientos a corto y largo plazo de los ingresos generados por hora de mano de obra de la construcción	Costos laborales en la construcción (Ingresos generados por hora de la mano de obra de la construcción)	Modelos de predicción VEC (vectores de corrección de error vectoriales)		RMS, MAPE, MAE, U-Stat	VEC	CORTO Y LARGO PLAZO	1. No especifica qué es el valor de desviación estándar 2. No indica el valor de desviación estándar			
P4	Predicting City-Level Construction Cost Index Using Linear Forecasting Models	Chi Young Choi, Young Eun Park, Jaehwan Kim, Jaehwan Kim	19-mar-21	Journal of Construction Engineering and Management	ML	https://doi.org/10.1061/(ASCE)CEM.19E.6983.00011973	El objetivo de este estudio es proporcionar pronósticos para el índice de precios al consumidor (IPC) a nivel de ciudad después de examinar el estado y las correlaciones de la heterogeneidad entre ciudades en los ICC de la ciudad	ICC	Modelos de pronóstico: modelo de regresión lineal, modelo de regresión logística, modelo de regresión de mínimos cuadrados, modelo de regresión de mínimos cuadrados ponderados		RMS, MAPE, MAE	VEC y ARIMA	12, 24, 36 MESES	NO PRESENTA VALORES DE DESVIACIÓN ESTÁNDAR			
P5	Predicting the Volatility of Highway Construction Cost Index Using Life	Yang Cao, Baozuo An, Jun Wang	28-mar-20	Journal of Management in Engineering	ML	https://doi.org/10.1061/(ASCE)JME.19E.6983.00011973	Modelar y predecir la variación del índice de costos de construcción de carreteras (ICC)	Índice de costos de construcción de carreteras	Modelos de pronóstico: modelo de regresión lineal, modelo de regresión logística, modelo de regresión de mínimos cuadrados, modelo de regresión de mínimos cuadrados ponderados		RMS, MAPE, MAE	LISTA	CORTO PLAZO (18 años)	No proporciona los valores de desviación estándar No indica			
P6	Forecasting Inflation Why Current Methods for Forecasting Costs May Be Inadequate	Omair Saeed	4-may-20	Journal of Construction Engineering and Management	ML	https://doi.org/10.1061/(ASCE)CEM.19E.6983.00011973	Proponer un marco ARIMA para pronosticar los precios de los materiales que capturen una compleja relación con los precios de los materiales energéticos, transacción y otros factores como una variable	A través de un estudio de caso con basalt, un material de construcción de principal producción para las agencias de transporte	Modelos ARIMA/VEC (Vectores de Corrección de Error Vectoriales)		MAE, MAPE	ARIMA/VEC	CORTO PLAZO (6 MESES)	NO ESPECIFICA EL VALOR DE DESVIACIÓN ESTÁNDAR NO PRESENTA RESULTADOS DE PREDICCIÓN			
P7	Performance evaluation of conventional-based CBI models for forecasting construction cost estimation	Yongjun Jang, Jungho Kim, Minhyun Lee, Namsik Lee, Namsik Lee	13-jun-20	Automation in Construction	ML	https://doi.org/10.1016/j.autcon.2020.103280	Examinar los efectos de los métodos de estimación a través de simulaciones de estimación de modelos CBI basados en normalización para mejorar la precisión de los índices de construcción en los proyectos de construcción	Índice de costos de construcción de viviendas (ICC)	CBM (Modelamiento basado en el tiempo)		RMS, MAPE, MAE	CBM	NO INDICA	NO ESPECIFICA EL VALOR DE DESVIACIÓN ESTÁNDAR NO PRESENTA RESULTADOS DE PREDICCIÓN			
P8	Cost Index Prediction for Construction Engineering Based on LSTM Neural Networks	Deng, Feihuang	14-sep-20	Advances in Civil Engineering	ML	https://doi.org/10.1061/(ASCE)ASCE.19E.6983.00011973	Analizar la aplicabilidad y el mecanismo de aplicación del algoritmo LSTM en el campo de la predicción de índices de construcción		Red neuronal LSTM		RMS, MAPE, MAE	LISTA	CORTO PLAZO	1. No especifica el tiempo (2017-2019)			
P9	Using moving window cross-correlation algorithm to optimize the base data estimating data usage	Parvati Patel, Parvati Patel, Parvati Patel	9-ene-20	International Journal of Advanced Management	Other	https://doi.org/10.1061/(ASCE)CEM.19E.6983.00011973	Desarrollar un enfoque que ayude con la definición de los datos de entrada para la estimación de los índices de construcción basados en la estimación de datos en un proyecto de construcción de carreteras	proyectos de construcción de transporte	Regresión no lineal		MAPE	Regresión no lineal	11, 30 MESES	Estimación de precios basados en precios referidos			
P10	A novel construction cost prediction model using hybrid neural and soft computing techniques	Adnan Qureshi, Hossain Ghaffar, Hossain Ghaffar, Hossain Ghaffar	28-oct-20	Advanced Engineering Research	Other	https://doi.org/10.1061/(ASCE)CEM.19E.6983.00011973	Evaluar seis modelos de ML para la predicción de costos de construcción de elementos estructurales horizontales	Costos de construcción de elementos estructurales	Integración de: 1. Red neuronal 2. Máquina de soporte vectorial 3. Máquina de aprendizaje 4. Máquina de aprendizaje profundo		RMS, MAPE, MAE	LOGBOOST, XGBOOST	NO INDICAR	No especifica qué es el valor de desviación estándar No especifica qué es el valor de desviación estándar			
P11	Estimation and prediction of construction cost index using neural networks, time series, and regression	Yasser Elshahhat	23-may-19	Advanced Engineering Journal	Other	https://doi.org/10.1061/(ASCE)CEM.19E.6983.00011973	El objetivo de este trabajo es analizar el valor del Índice de Costos de la Construcción con base en datos históricos de construcción de un caso de construcción de un apartamento en el futuro.	ÍNDICE DE COSTOS DE CONSTRUCCIÓN	1. Series de tiempo autorregresiva (AR) 2. Regresión lineal		RMS, MAPE, MAE	AR	7 AÑOS - LARGO PLAZO	1. No especifica qué es el valor de desviación estándar 2. No especifica qué es el valor de desviación estándar			
P12	Forecasting residential building costs in New Zealand using a concrete approach	Lixin Zhao, Jianhua Wang, Jianhua Wang, Jianhua Wang	2-mar-19	International Journal of Engineering, Business and Technology Management	ML	https://doi.org/10.1061/(ASCE)CEM.19E.6983.00011973	Pronosticar el costo de construcción de cinco categorías de edificios residenciales: apartamento, vivienda de pueblo para adultos, casa de campo, apartamento y vivienda de pueblo para niños	Índice de costos de la construcción de viviendas (Índice de viviendas)	1. Bases de datos históricas de precios de materiales 2. Bases de datos históricas de precios de mano de obra 3. Bases de datos históricas de precios de alquiler 4. Bases de datos históricas de precios de impuestos 5. Bases de datos históricas de precios de servicios públicos 6. Bases de datos históricas de precios de energía 7. Bases de datos históricas de precios de agua 8. Bases de datos históricas de precios de gas 9. Bases de datos históricas de precios de electricidad 10. Bases de datos históricas de precios de calefacción 11. Bases de datos históricas de precios de refrigeración 12. Bases de datos históricas de precios de ventilación 13. Bases de datos históricas de precios de aislamiento 14. Bases de datos históricas de precios de pintura 15. Bases de datos históricas de precios de carpintería 16. Bases de datos históricas de precios de electricidad 17. Bases de datos históricas de precios de gas 18. Bases de datos históricas de precios de agua 19. Bases de datos históricas de precios de calefacción 20. Bases de datos históricas de precios de refrigeración 21. Bases de datos históricas de precios de ventilación 22. Bases de datos históricas de precios de aislamiento 23. Bases de datos históricas de precios de pintura 24. Bases de datos históricas de precios de carpintería 25. Bases de datos históricas de precios de electricidad 26. Bases de datos históricas de precios de gas 27. Bases de datos históricas de precios de agua 28. Bases de datos históricas de precios de calefacción 29. Bases de datos históricas de precios de refrigeración 30. Bases de datos históricas de precios de ventilación 31. Bases de datos históricas de precios de aislamiento 32. Bases de datos históricas de precios de pintura 33. Bases de datos históricas de precios de carpintería 34. Bases de datos históricas de precios de electricidad 35. Bases de datos históricas de precios de gas 36. Bases de datos históricas de precios de agua 37. Bases de datos históricas de precios de calefacción 38. Bases de datos históricas de precios de refrigeración 39. Bases de datos históricas de precios de ventilación 40. Bases de datos históricas de precios de aislamiento 41. Bases de datos históricas de precios de pintura 42. Bases de datos históricas de precios de carpintería 43. Bases de datos históricas de precios de electricidad 44. Bases de datos históricas de precios de gas 45. Bases de datos históricas de precios de agua 46. Bases de datos históricas de precios de calefacción 47. Bases de datos históricas de precios de refrigeración 48. Bases de datos históricas de precios de ventilación 49. Bases de datos históricas de precios de aislamiento 50. Bases de datos históricas de precios de pintura 51. Bases de datos históricas de precios de carpintería 52. Bases de datos históricas de precios de electricidad 53. Bases de datos históricas de precios de gas 54. Bases de datos históricas de precios de agua 55. Bases de datos históricas de precios de calefacción 56. Bases de datos históricas de precios de refrigeración 57. Bases de datos históricas de precios de ventilación 58. Bases de datos históricas de precios de aislamiento 59. Bases de datos históricas de precios de pintura 60. Bases de datos históricas de precios de carpintería 61. Bases de datos históricas de precios de electricidad 62. Bases de datos históricas de precios de gas 63. Bases de datos históricas de precios de agua 64. Bases de datos históricas de precios de calefacción 65. Bases de datos históricas de precios de refrigeración 66. Bases de datos históricas de precios de ventilación 67. Bases de datos históricas de precios de aislamiento 68. Bases de datos históricas de precios de pintura 69. Bases de datos históricas de precios de carpintería 70. Bases de datos históricas de precios de electricidad 71. Bases de datos históricas de precios de gas 72. Bases de datos históricas de precios de agua 73. Bases de datos históricas de precios de calefacción 74. Bases de datos históricas de precios de refrigeración 75. Bases de datos históricas de precios de ventilación 76. Bases de datos históricas de precios de aislamiento 77. Bases de datos históricas de precios de pintura 78. Bases de datos históricas de precios de carpintería 79. Bases de datos históricas de precios de electricidad 80. Bases de datos históricas de precios de gas 81. Bases de datos históricas de precios de agua 82. Bases de datos históricas de precios de calefacción 83. Bases de datos históricas de precios de refrigeración 84. Bases de datos históricas de precios de ventilación 85. Bases de datos históricas de precios de aislamiento 86. Bases de datos históricas de precios de pintura 87. Bases de datos históricas de precios de carpintería 88. Bases de datos históricas de precios de electricidad 89. Bases de datos históricas de precios de gas 90. Bases de datos históricas de precios de agua 91. Bases de datos históricas de precios de calefacción 92. Bases de datos históricas de precios de refrigeración 93. Bases de datos históricas de precios de ventilación 94. Bases de datos históricas de precios de aislamiento 95. Bases de datos históricas de precios de pintura 96. Bases de datos históricas de precios de carpintería 97. Bases de datos históricas de precios de electricidad 98. Bases de datos históricas de precios de gas 99. Bases de datos históricas de precios de agua 100. Bases de datos históricas de precios de calefacción 101. Bases de datos históricas de precios de refrigeración 102. Bases de datos históricas de precios de ventilación 103. Bases de datos históricas de precios de aislamiento 104. Bases de datos históricas de precios de pintura 105. Bases de datos históricas de precios de carpintería 106. Bases de datos históricas de precios de electricidad 107. Bases de datos históricas de precios de gas 108. Bases de datos históricas de precios de agua 109. Bases de datos históricas de precios de calefacción 110. Bases de datos históricas de precios de refrigeración 111. Bases de datos históricas de precios de ventilación 112. Bases de datos históricas de precios de aislamiento 113. Bases de datos históricas de precios de pintura 114. Bases de datos históricas de precios de carpintería 115. Bases de datos históricas de precios de electricidad 116. Bases de datos históricas de precios de gas 117. Bases de datos históricas de precios de agua 118. Bases de datos históricas de precios de calefacción 119. Bases de datos históricas de precios de refrigeración 120. Bases de datos históricas de precios de ventilación 121. Bases de datos históricas de precios de aislamiento 122. Bases de datos históricas de precios de pintura 123. Bases de datos históricas de precios de carpintería 124. Bases de datos históricas de precios de electricidad 125. Bases de datos históricas de precios de gas 126. Bases de datos históricas de precios de agua 127. Bases de datos históricas de precios de calefacción 128. Bases de datos históricas de precios de refrigeración 129. Bases de datos históricas de precios de ventilación 130. Bases de datos históricas de precios de aislamiento 131. Bases de datos históricas de precios de pintura 132. Bases de datos históricas de precios de carpintería 133. Bases de datos históricas de precios de electricidad 134. Bases de datos históricas de precios de gas 135. Bases de datos históricas de precios de agua 136. Bases de datos históricas de precios de calefacción 137. Bases de datos históricas de precios de refrigeración 138. Bases de datos históricas de precios de ventilación 139. Bases de datos históricas de precios de aislamiento 140. Bases de datos históricas de precios de pintura 141. Bases de datos históricas de precios de carpintería 142. Bases de datos históricas de precios de electricidad 143. Bases de datos históricas de precios de gas 144. Bases de datos históricas de precios de agua 145. Bases de datos históricas de precios de calefacción 146. Bases de datos históricas de precios de refrigeración 147. Bases de datos históricas de precios de ventilación 148. Bases de datos históricas de precios de aislamiento 149. Bases de datos históricas de precios de pintura 150. Bases de datos históricas de precios de carpintería 151. Bases de datos históricas de precios de electricidad 152. Bases de datos históricas de precios de gas 153. Bases de datos históricas de precios de agua 154. Bases de datos históricas de precios de calefacción 155. Bases de datos históricas de precios de refrigeración 156. Bases de datos históricas de precios de ventilación 157. Bases de datos históricas de precios de aislamiento 158. Bases de datos históricas de precios de pintura 159. Bases de datos históricas de precios de carpintería 160. Bases de datos históricas de precios de electricidad 161. Bases de datos históricas de precios de gas 162. Bases de datos históricas de precios de agua 163. Bases de datos históricas de precios de calefacción 164. Bases de datos históricas de precios de refrigeración 165. Bases de datos históricas de precios de ventilación 166. Bases de datos históricas de precios de aislamiento 167. Bases de datos históricas de precios de pintura 168. Bases de datos históricas de precios de carpintería 169. Bases de datos históricas de precios de electricidad 170. Bases de datos históricas de precios de gas 171. Bases de datos históricas de precios de agua 172. Bases de datos históricas de precios de calefacción 173. Bases de datos históricas de precios de refrigeración 174. Bases de datos históricas de precios de ventilación 175. Bases de datos históricas de precios de aislamiento 176. Bases de datos históricas de precios de pintura 177. Bases de datos históricas de precios de carpintería 178. Bases de datos históricas de precios de electricidad 179. Bases de datos históricas de precios de gas 180. Bases de datos históricas de precios de agua 181. Bases de datos históricas de precios de calefacción 182. Bases de datos históricas de precios de refrigeración 183. Bases de datos históricas de precios de ventilación 184. Bases de datos históricas de precios de aislamiento 185. Bases de datos históricas de precios de pintura 186. Bases de datos históricas de precios de carpintería 187. Bases de datos históricas de precios de electricidad 188. Bases de datos históricas de precios de gas 189. Bases de datos históricas de precios de agua 190. Bases de datos históricas de precios de calefacción 191. Bases de datos históricas de precios de refrigeración 192. Bases de datos históricas de precios de ventilación 193. Bases de datos históricas de precios de aislamiento 194. Bases de datos históricas de precios de pintura 195. Bases de datos históricas de precios de carpintería 196. Bases de datos históricas de precios de electricidad 197. Bases de datos históricas de precios de gas 198. Bases de datos históricas de precios de agua 199. Bases de datos históricas de precios de calefacción 200. Bases de datos históricas de precios de refrigeración 201. Bases de datos históricas de precios de ventilación 202. Bases de datos históricas de precios de aislamiento 203. Bases de datos históricas de precios de pintura 204. Bases de datos históricas de precios de carpintería 205. Bases de datos históricas de precios de electricidad 206. Bases de datos históricas de precios de gas 207. Bases de datos históricas de precios de agua 208. Bases de datos históricas de precios de calefacción 209. Bases de datos históricas de precios de refrigeración 210. Bases de datos históricas de precios de ventilación 211. Bases de datos históricas de precios de aislamiento 212. Bases de datos históricas de precios de pintura 213. Bases de datos históricas de precios de carpintería 214. Bases de datos históricas de precios de electricidad 215. Bases de datos históricas de precios de gas 216. Bases de datos históricas de precios de agua 217. Bases de datos históricas de precios de calefacción 218. Bases de datos históricas de precios de refrigeración 219. Bases de datos históricas de precios de ventilación 220. Bases de datos históricas de precios de aislamiento 221. Bases de datos históricas de precios de pintura 222. Bases de datos históricas de precios de carpintería 223. Bases de datos históricas de precios de electricidad 224. Bases de datos históricas de precios de gas 225. Bases de datos históricas de precios de agua 226. Bases de datos históricas de precios de calefacción 227. Bases de datos históricas de precios de refrigeración 228. Bases de datos históricas de precios de ventilación 229. Bases de datos históricas de precios de aislamiento 230. Bases de datos históricas de precios de pintura 231. Bases de datos históricas de precios de carpintería 232. Bases de datos históricas de precios de electricidad 233. Bases de datos históricas de precios de gas 234. Bases de datos históricas de precios de agua 235. Bases de datos históricas de precios de calefacción 236. Bases de datos históricas de precios de refrigeración 237. Bases de datos históricas de precios de ventilación 238. Bases de datos históricas de precios de aislamiento 239. Bases de datos históricas de precios de pintura 240. Bases de datos históricas de precios de carpintería 241. Bases de datos históricas de precios de electricidad 242. Bases de datos históricas de precios de gas 243. Bases de datos históricas de precios de agua 244. Bases de datos históricas de precios de calefacción 245. Bases de datos históricas de precios de refrigeración 246. Bases de datos históricas de precios de ventilación 247. Bases de datos históricas de precios de aislamiento 248. Bases de datos históricas de precios de pintura 249. Bases de datos históricas de precios de carpintería 250. Bases de datos históricas de precios de electricidad 251. Bases de datos históricas de precios de gas 252. Bases de datos históricas de precios de agua 253. Bases de datos históricas de precios de calefacción 254. Bases de datos históricas de precios de refrigeración 255. Bases de datos históricas de precios de ventilación 256. Bases de datos históricas de precios de aislamiento 257. Bases de datos históricas de precios de pintura 258. Bases de datos históricas de precios de carpintería 259. Bases de datos históricas de precios de electricidad 260. Bases de datos históricas de precios de gas 261. Bases de datos históricas de precios de agua 262. Bases de datos históricas de precios de calefacción 263. Bases de datos históricas de precios de refrigeración 264. Bases de datos históricas de precios de ventilación 265. Bases de datos históricas de precios de aislamiento 266. Bases de datos históricas de precios de pintura 267. Bases de datos históricas de precios de carpintería 268. Bases de datos históricas de precios de electricidad 269. Bases de datos históricas de precios de gas 270. Bases de datos históricas de precios de agua 271. Bases de datos históricas de precios de calefacción 272. Bases de datos históricas de precios de refrigeración 273. Bases de datos históricas de precios de ventilación 274. Bases de datos históricas de precios de aislamiento 275. Bases de datos históricas de precios de pintura 276. Bases de datos históricas de precios de carpintería 277. Bases de datos históricas de precios de electricidad 278. Bases de datos históricas de precios de gas 279. Bases de datos históricas de precios de agua 280. Bases de datos históricas de precios de calefacción 281. Bases de datos históricas de precios de refrigeración 282. Bases de datos históricas de precios de ventilación 283. Bases de datos históricas de precios de aislamiento 284. Bases de datos históricas de precios de pintura 285. Bases de datos históricas de precios de carpintería 286. Bases de datos históricas de precios de electricidad 287. Bases de datos históricas de precios de gas 288. Bases de datos históricas de precios de agua 289. Bases de datos históricas de precios de calefacción 290. Bases de datos históricas de precios de refrigeración 291. Bases de datos históricas de precios de ventilación 292. Bases de datos históricas de precios de aislamiento 293. Bases de datos históricas de precios de pintura 294. Bases de datos históricas de precios de carpintería 295. Bases de datos históricas de precios de electricidad 296. Bases de datos históricas de precios de gas 297. Bases de datos históricas de precios de agua 298. Bases de datos históricas de precios de calefacción 299. Bases de datos históricas de precios de refrigeración 300. Bases de datos históricas de precios de ventilación 301. Bases de datos históricas de precios de aislamiento 302. Bases de datos históricas de precios de pintura 303. Bases de datos históricas de precios de carpintería 304. Bases de datos históricas de precios de electricidad 305. Bases de datos históricas de precios de gas 306. Bases de datos históricas de precios de agua 307. Bases de datos históricas de precios de calefacción 308. Bases de datos históricas de precios de refrigeración 309. Bases de datos históricas de precios de ventilación 310. Bases de datos históricas de precios de aislamiento 311. Bases de datos históricas de precios de pintura 312. Bases de datos históricas de precios de carpintería 313. Bases de datos históricas de precios de electricidad 314. Bases de datos históricas de precios de gas 315. Bases de datos históricas de precios de agua 316. Bases de datos históricas de precios de calefacción 317. Bases de datos históricas de precios de refrigeración 318. Bases de datos históricas de precios de ventilación 319. Bases de datos históricas de precios de aislamiento 320. Bases de datos históricas de precios de pintura 321. Bases de datos históricas de precios de carpintería 322. Bases de datos históricas de precios de electricidad 323. Bases de datos históricas de precios de gas 324. Bases de datos históricas de precios de agua 325. Bases de datos históricas de precios de calefacción 326. Bases de datos históricas de precios de refrigeración 327. Bases de datos históricas de precios de ventilación 328. Bases de datos históricas de precios de aislamiento 329. Bases de datos históricas de precios de pintura 330. Bases de datos históricas de precios de carpintería 331. Bases de datos históricas de precios de electricidad 332. Bases de datos históricas de precios de gas 333. Bases de datos históricas de precios de agua 334. Bases de datos históricas de precios de calefacción 335. Bases de datos históricas de precios de refrigeración 336. Bases de datos históricas de precios de ventilación 337. Bases de datos históricas de precios de aislamiento 338. Bases de datos históricas de precios de pintura 339. Bases de datos históricas de precios de carpintería 340. Bases de datos históricas de precios de electricidad 341. Bases de datos históricas de precios de gas 342. Bases de datos históricas de precios de agua 343. Bases de datos históricas de precios de calefacción 344. Bases de datos históricas de precios de refrigeración 345. Bases de datos históricas de precios de ventilación 346. Bases de datos históricas de precios de aislamiento 347. Bases de datos históricas de precios de pintura 348. Bases de datos históricas de precios de carpintería 349. Bases de datos históricas de precios de electricidad 350. Bases de datos históricas de precios de gas 351. Bases de datos históricas de precios de agua 352. Bases de datos históricas de precios de calefacción 353. Bases de datos históricas de precios de refrigeración 354. Bases de datos históricas de precios de ventilación 355. Bases de datos históricas de precios de aislamiento 356. Bases de datos históricas de precios de pintura 357. Bases de datos históricas de precios de carpintería 358. Bases de datos históricas de precios de electricidad 359. Bases de datos históricas de precios de gas 360. Bases de datos históricas de precios de agua 361. Bases de datos históricas de precios de calefacción 362. Bases de datos históricas de precios de refrigeración 363. Bases de datos históricas de precios de ventilación 364. Bases de datos históricas de precios de aislamiento 365. Bases de datos históricas de precios de pintura 366. Bases de datos históricas de precios de carpintería 367. Bases de datos históricas de precios de electricidad 368. Bases de datos históricas de precios de gas 369. Bases de datos históricas de precios de agua 370. Bases de datos históricas de precios de calefacción 371. Bases de datos históricas de precios de refrigeración 372. Bases de datos históricas de precios de ventilación 373. Bases de datos históricas de precios de aislamiento 374. Bases de datos históricas de precios de pintura 375. Bases de datos históricas de precios de carpintería 376. Bases de datos históricas de precios de electricidad 377. Bases de datos históricas de precios de gas 378. Bases de datos históricas de precios de agua 379. Bases de datos históricas de precios de calefacción 380. Bases de datos históricas de precios de refrigeración 381. Bases de datos históricas de precios de ventilación 382. Bases de datos históricas de precios de aislamiento 383. Bases de datos históricas de precios de pintura 384. Bases de datos históricas de precios de carpintería 385. Bases de datos históricas de precios de electricidad 386. Bases de datos históricas de precios de gas 387. Bases de datos históricas de precios de agua 388. Bases de datos históricas de precios de calefacción 389. Bases de datos históricas de precios de refrigeración 390. Bases de datos históricas de precios de ventilación 391. Bases de datos históricas de precios de aislamiento 392. Bases de datos históricas de precios de pintura 393. Bases de datos históricas de precios de carpintería 394. Bases de datos históricas de precios de electricidad 395. Bases de datos históricas de precios de gas 396. Bases de datos históricas de precios de agua 397. Bases de datos históricas de precios de calefacción 398. Bases de datos históricas de precios de refrigeración 399. Bases de datos históricas de precios de ventilación 400. Bases de datos históricas de precios de aislamiento 401. Bases de datos históricas de precios de pintura 402. Bases de datos históricas de precios de carpintería 403. Bases de datos históricas de precios de electricidad 404. Bases de datos históricas de precios de gas 405. Bases de datos históricas de precios de agua 406. Bases de datos históricas de precios de calefacción 407. Bases de datos históricas de precios de refrigeración 408. Bases de datos históricas de precios de ventilación 409. Bases de datos históricas de precios de aislamiento 410. Bases de datos históricas de precios de pintura 411. Bases de datos históricas de precios de carpintería 412. Bases de datos históricas de precios de electricidad 413. Bases de datos históricas de precios de gas 414. Bases de datos históricas de precios de agua 415. Bases de datos históricas de precios de calefacción 416. Bases de datos históricas de precios de refrigeración 417. Bases de datos históricas de precios de ventilación 418. Bases de datos históricas de precios de aislamiento 419. Bases de datos históricas de precios de pintura 420. Bases de datos históricas de precios de carpintería 421. Bases de datos históricas de precios de electricidad 422. Bases de datos históricas de precios de gas 423. Bases de datos históricas de precios de agua 424. Bases de datos históricas de precios de calefacción 425. Bases de datos históricas de precios de refrigeración 426. Bases de datos históricas de precios de ventilación 427. Bases de datos históricas de precios de aislamiento 428. Bases de datos históricas de precios de pintura 429. Bases de datos históricas de precios de carpintería 430. Bases de datos históricas de precios de electricidad 431. Bases de datos históricas de precios de gas 432. Bases de datos históricas de precios de agua 433. Bases de datos históricas de precios de calefacción 434. Bases de datos históricas de precios de refrigeración 435. Bases de datos históricas de precios de ventilación 436. Bases de datos históricas de precios de aislamiento 437. Bases de datos históricas de precios de pintura 438. Bases de datos históricas de precios de carpintería 439. Bases de datos históricas de precios de electricidad 440. Bases de datos históricas de precios de gas 441. Bases de datos históricas de precios de agua 442. Bases de datos históricas de precios de calefacción 443. Bases de datos históricas de precios de refrigeración 444. Bases de datos históricas de precios de ventilación 445. Bases de datos históricas de precios de aislamiento 446. Bases de datos históricas de precios de pintura 447. Bases de datos históricas de precios de carpintería 448. Bases de datos históricas de precios de electricidad 449. Bases de datos históricas de precios de gas 450. Bases de datos históricas de precios de agua 451. Bases de datos históricas de precios de calefacción 452. Bases de datos históricas de precios de refrigeración 453. Bases de datos históricas de precios de ventilación 454. Bases de datos históricas de precios de aislamiento 455. Bases de datos históricas de precios de pintura 456. Bases de datos históricas de precios de carpintería 457. Bases de datos históricas de precios de electricidad 458. Bases de datos históricas de precios de gas 459. Bases de datos históricas de precios de agua 460. Bases de datos históricas de precios de calefacción 461. Bases de datos históricas de precios de refrigeración 462. Bases de datos históricas de precios de ventilación 463. Bases de datos históricas de precios de aislamiento 464. Bases de datos históricas de precios de pintura 465. Bases de datos históricas de precios de carpintería 466. Bases de datos históricas de precios de electricidad 467. Bases de datos históricas de precios de gas 468. Bases de datos históricas de precios de agua 469. Bases de datos históricas de precios de calefacción 470. Bases de datos históricas de precios de refrigeración 471. Bases de datos históricas de precios de ventilación 472. Bases de datos históricas de precios de aislamiento 473. Bases de datos históricas de precios de pintura 474. Bases de datos históricas de precios de carpintería 475. Bases de datos históricas de precios de electricidad 476. Bases de datos históricas de precios de gas 477. Bases de datos históricas de precios de agua 478. Bases de datos históricas de precios de calefacción 479. Bases de datos históricas de precios de refrigeración 480. Bases de datos históricas de precios de ventilación 481. Bases de datos históricas de precios de aislamiento 482. Bases de datos históricas de precios de pintura 483. Bases de datos históricas de precios de carpintería 484. Bases de datos históricas de precios de electricidad 485. Bases de datos históricas de precios de gas 486. Bases de datos históricas de precios de agua 487. Bases de datos históricas de precios de calefacción 488. Bases de datos históricas de precios de refrigeración 489. Bases de datos								

Anexo 2: Cantidad de datos de las variables que intervienen en el modelado.

Table 22: Cantidad de datos de las variables que intervienen en el modelado.

No	Variable	Descripción	Número de datos
1	IPCOU	Índice de precios de la construcción de viviendas unifamiliares.	245
2	IPCOM	Índice de precios de la construcción de viviendas multifamiliares.	245
3	PIB	Producto interno bruto.	229
4	IPP	Índice de Precios al Productor de Disponibilidad Nacional.	247
5	IPC	Indice de precios al consumidor.	247
6	OM	Oferta monetaria.	245
7	TI	Tasas de interés vigentes para el segmento inmobiliario.	245
8	ICIF	Importaciones CIF por uso o destino económico (Materias primas para materiales de construcción).	245
9	IFOB	Importaciones FOB por uso o destino económico (Materias primas para materiales de construcción).	245
10	PPEC	Precios promedio del crudo ecuatoriano cesta Oriente y Napo.	246
11	PPWTI	Precios promedio del crudo internacional - petróleo West Texas Intermediate.	246
12	ICC	Índice de confianza del consumidor nacional (Quito, Guayaquil, Cuenca, Ambato y Machala).	176
13	DESEM	Tasa de desempleo.	177
14	PCONS	Permisos de construcción otorgados a nivel nacional	240

Anexo 3: Código R para prueba de modelos estocásticos.

```
##### PRUEBA DE MODELOS #####
##### Realizado por: Mayra Viscaino
##### date: 15-08-2022
##### DATOS REALES - PREDICTORES INDIVIDUALES

### Cargar librerías
library(tidyverse)

library(readxl)
```

```

library(fabletools)
library(fable)
library(tsibble)
library(tsibbledata)
library(dplyr)
# Librerías para exportar predicciones
library(rio)
library(haven)

##### 1. CARGAR DATOS #####
# Importar datos desde xlsx --> cargar: library(readxl)
DataIPCOU <- read_excel("D:/MAESTRIA UNIVERSIDAD TECNICA DE AMBATO
/MODULO 11- TEMAS DE TESIS/7. LIBRO DE PRONOSTICO
/DatasetPredictoresTrainValid/IPCOU.xlsx")
#View(DataIPCOU)

# Transformar datos a serie de tiempo
DataIPCOU <- DataIPCOU %>%
  mutate(Month = yearmonth(Month)) %>%
  as_tsibble(index = Month)
view(DataIPCOU)
str(DataIPCOU) # Muestra informaciOn de los datos

##### 2. GRAFICAR LOS DATOS #####

# Gráfico de la serie de tiempo de la variable explicativa candidata
DataIPCOU %>%
  autoplot(.vars = IPCOU) +
  labs(title = "Valores de IPCOU",
        subtitle = "IPCOU",
        y = "IPCOU")

##### 3. MODELOS GRUPO 1 #####
##### 3.1 Modelos básicos - Validación cruzada #####

# División del conjunto de datos
DataIPCOU_Train <- DataIPCOU %>%
  stretch_tsibble(.init = 120, .step = 12) %>%
  relocate(Month, .id)
view(DataIPCOU_Train)
# Exactitud de testeo aplicando Validación cruzada
DataIPCOU_Train %>%
  model(
    Mean = MEAN(IPCOU),
    `Naïve` = NAIVE(IPCOU),
    `Seasonal naïve` = SNAIVE(IPCOU),
    `Drift` = NAIVE(IPCOU ~ drift())
  ) %>%
  forecast(h = 24) %>%

```

```

accuracy(DataIPCOU)

# Exactitud de entrenamiento
DataIPCOU %>%
  model(
    Mean = MEAN(IPCOU),
    `Naïve` = NAIVE(IPCOU),
    `Seasonal naïve` = SNAIVE(IPCOU),
    `Drift` = NAIVE(IPCOU ~ drift())
  ) %>%
  accuracy()

# NOTA: Seleccionar el modelo con menor valor de RMSE y MAPE.

##### 4. MODELOS GRUPO 2 #####
##### 4.1 Modelos de Suavizado Exponencial #####

# División del conjunto de datos y exactitud de testeo
DataIPCOU %>%
  stretch_tsibble(.init = 120,.step = 12) %>%
  model(
    SES = ETS(IPCOU ~ error("A") + trend("N") + season("N")),
    Holt = ETS(IPCOU ~ error("A") + trend("A") + season("N")),
    Damped = ETS(IPCOU ~ error("A") + trend("Ad") + season("N")),
    `Damped Holt's method` = ETS(IPCOU ~ error("A") +
      trend("Ad", alpha = 0.5, beta = 0.5, phi = 0.9) + season("N"))
  ) %>%
  forecast(h = 24) %>%
  accuracy(DataIPCOU)

# Exactitud de entrenamiento
DataIPCOU %>%
  model(
    SES = ETS(IPCOU ~ error("A") + trend("N") + season("N")),
    Holt = ETS(IPCOU ~ error("A") + trend("A") + season("N")),
    Damped = ETS(IPCOU ~ error("A") + trend("Ad") + season("N")),
    `Damped Holt's method` = ETS(IPCOU ~ error("A") +
      trend("Ad", alpha = 0.5, beta = 0.5, phi = 0.9) + season("N"))
  ) %>%
  accuracy()

##### 5. MODELOS GRUPO 3 #####
##### 5.1 Modelos ARIMA Y STL #####
# Sección 9.10: La función ARIMA -> devuelve el mejor modelo

DataIPCOU %>%
  stretch_tsibble(.init = 120,.step = 12) %>%
  model(
    ARIMA(IPCOU)
  )

```

```

) %>%
forecast(h = 24) %>%
accuracy(DataIPCOU) #Evaluar el mejor modelo comparando los indicadores

# Exactitud de entrenamiento aplicando Validación-cruzada
DataIPCOU %>%
  model(
    ARIMA(IPCOU)
  ) %>%
  accuracy()

# Reporte del modelo ARIMA
ARIMA_IPCOU <- DataIPCOU %>%
  model(ARIMA(IPCOU))
report(ARIMA_IPCOU) # Reporte del modelo: ARIMA(1,1,2)

####      Modelo STL
STLF <- decomposition_model(
  STL(IPCOU ~ season(window = Inf)),
  ETS(IPCOU ~ season("N"))
)

DataIPCOU %>%
  stretch_tsibble(.init = 100,.step = 12) %>%
  model(
    stlf = STLF
  ) %>%
  forecast(h = 24) %>%
  accuracy(DataIPCOU) #Evaluar el mejor modelo comparando los indicadores

# Exactitud de entrenamiento aplicando Validación-cruzada
DataIPCOU %>%
  model(
    stlf = STLF,
  ) %>%
  accuracy()

#####      6.  MODELOS  GRUPO 4      #####
#####      6.1 Modelos COMBINADOS      #####

# STLF <- decomposition_model(
#   STL(IPCOU ~ season(window = Inf)),
#   ETS(IPCOU ~ season("N"))
# )

DataIPCOU %>%
  stretch_tsibble(.init = 120,.step = 12) %>%
  model(
    Naive = NAIVE(IPCOU),

```

```

    SES = ETS(IPCOU ~ error("A") + trend("N") + season("N")),
    #ets = ETS(IPCOU),
    #Drift = NAIVE(IPCOU ~ drift()),
    arima = ARIMA(IPCOU)
  ) %>%
mutate(M1_Combinado = (Naive + SES + arima) / 3) %>%
forecast(h = 24) %>%
accuracy(DataIPCOU) #Evaluar el mejor modelo comparando los indicadores

# Exactitud de entrenamiento aplicando Validación-cruzada
DataIPCOU %>%
  model(
    Naive = NAIVE(IPCOU),
    SES = ETS(IPCOU ~ error("A") + trend("N") + season("N")),
    #ets = ETS(IPCOU),
    #Drift = NAIVE(IPCOU ~ drift()),
    arima = ARIMA(IPCOU)
  ) %>%
mutate(M1_Combinado = (Naive + SES + arima) / 3) %>%
accuracy()

##### 6.2 Combinación de modelos básicos - validación cruzada #####
DataIPCOU %>%
  stretch_tsibble(.init = 120,.step = 12) %>%
  model(
    Naive = NAIVE(IPCOU),
    Seasonal_naive = SNAIVE(IPCOU),
    Drift = NAIVE(IPCOU ~ drift()),
  ) %>%
mutate(M2_Combinado = (Naive + Seasonal_naive + Drift) / 3) %>%
forecast(h = 24) %>%
accuracy(DataIPCOU)

# Exactitud de entrenamiento aplicando Validación-cruzada
DataIPCOU %>%
  model(
    Naive = NAIVE(IPCOU),
    Seasonal_naive = SNAIVE(IPCOU),
    Drift = NAIVE(IPCOU ~ drift()),
  ) %>%
mutate(M2_Combinado = (Naive + Seasonal_naive + Drift) / 3) %>%
accuracy()

##### 7. MODELOS GRUPO 5 #####
##### 7.1 Selección del Modelo #####

# Conjunto de entrenamiento
DataTrain <- DataIPCOU %>%
  filter_index("2002-01" ~ "2017-12") # 75% Datos para entrenamiento

```

```

# Conjunto de prueba
DataTest <- DataIPCOU %>%
  filter_index("2018-01" ~ .)          # 15% Datos para validación

# Entrenamiento y testeo de los modelos con la función model()
### Entrenamiento
Model_IPCOU_fit <- DataTrain %>%
  model(
    #Drift = NAIVE(IPCOU ~ drift()),
    SES = ETS(IPCOU ~ error("A") + trend("N") + season("N")),
    `Arima` = ARIMA(IPCOU),
    Naive = NAIVE(IPCOU)#,
    #Seasonal_naive = SNAIVE(IPCOU)
  ) %>%
  mutate(M_Comb=(Naive + SES + Arima) / 3)
### Testeo
IPCOU_predic <- Model_IPCOU_fit %>%
  forecast(DataTest)

# Evaluación del modelo
accuracy(IPCOU_predic, DataIPCOU)

#####      8. GENERACIÓN DE PRONÓSTICOS      #####
#####      8.1 Generación de pronósticos      #####
# Para la generación de los pronósticos, se empleará el conjunto
completo de datos
Model_IPCOU_fc <- DataIPCOU %>% # Modelo entrenado con todos
los valores
  model(
    #Drift = NAIVE(IPCOU ~ drift()),
    #SES = ETS(IPCOU ~ error("A") + trend("N") + season("N")),
    `Arima` = ARIMA(IPCOU)#,
    #Naive = NAIVE(IPCOU),
    #Seasonal_naive = SNAIVE(IPCOU)
  )# %>%
  #mutate(M_Combinado=(Naive + Seasonal_naive + Drift) / 3)

IPCOU_predic_2021 <- Model_IPCOU_fc %>% # Proyección de valores
para modelo de regresión múltiple
  forecast(h=24)

#####      8.2 Exportación de datos      #####
# Directorio para guardar el archivo
setwd("D:/MAESTRIA UNIVERSIDAD TECNICA DE AMBATO/MODULO 11- TEMAS
DE TESIS/12. RESULTADOS")
export(IPCOU_predic_2021, "IPCOU_predic_2021.xlsx")

```

Anexo 4: Código R para prueba de modelos de regresión múltiple.

```
##### MODELOS DE REGRESIÓN DE SERIES DE TIEMPO #####
##### Realizado por: Mayra Viscaíno.
##### Fecha: "17-08-2022"
##### SELECCIÓN DE MODELOS DE REGRESIÓN MÚLTIPLE

### Carga de librerías
library(tidyverse)
library(fable)
library(tsibbledata)
library(tsibble)
library(lubridate)
library(readxl)
library(dplyr)
library(fabletools)

##### Importación de datos #####

# Importar datos desde xls --> cargar: library(readxl)
Dataset_Train_Valid_2019 <- read_excel("D:/MAESTRIA UNIVERSIDAD TECNICA
DE AMBATO/MODULO 11- TEMAS DE TESIS/7. LIBRO DE PRONOSTICO
/DatasetPredictoresTrainValid/Dataset_Train_Valid_2019.xlsx")

# Transformando a formato serie de tiempo
Dataset_FC_2021 <- Dataset_Train_Valid_2019 %>%
  mutate(Month = yearmonth(Month)) %>%
  as_tsibble(index = Month)
str(Dataset_FC_2021) # Muestra información de los datos

##### MODELOS DE REGRESIÓN LINEAL MÚLTIPLE #####
##### Entrenamiento con validación cruzada #####

# División del conjunto de datos
Data_FC_Train <- Dataset_FC_2021 %>%
  stretch_tsibble(.init = 110, .step = 12) %>%
  relocate(Month, .id)
#view(DataIPCOM_Train)

# Excititud de testeo aplicando Validación cruzada
Data_FC_Train %>%
  model(
    `tslm_1` = TSLM(IPCOM ~ PIBA+PIBT+IPP+IPC+OM+TI+ICIF+IFOB+PPEC+
PPWTI+ICC+DESEM+PCONS),
    `tslm_2` = TSLM(IPCOM ~ PIBA+IPP+IPC+TI+ICIF+IFOB+PPWTI+ICC
+DESEM+PCONS),
    `tslm_3` = TSLM(IPCOM ~ PIBA+IPP+IPC+TI+ICIF+PPWTI+ICC+DESEM+PCONS)
    `tslm_4` = TSLM(IPCOM ~ PIBA+IPP+IPC+TI+ICIF+PPWTI+DESEM+PCONS)
```

```

) %>%
forecast(Data_FC_Train) %>%
accuracy(Dataset_FC_2021)

# Exactitud de entrenamiento
Dataset_FC_2021 %>%
  model(
    `tslm_1` = TSLM(IPCOM ~ PIBA+PIBT+IPP+IPC+OM+TI+ICIF+IFOB+PPEC
+PPWTI+ICC+DESEM+PCONS),
    `tslm_2` = TSLM(IPCOM ~ PIBA+IPP+IPC+TI+ICIF+IFOB+PPWTI+ICC
+DESEM+PCONS), # -PIBT, -OM, -PPEC
    `tslm_3` = TSLM(IPCOM ~ PIBA+IPP+IPC+TI+ICIF+PPWTI+ICC
+DESEM+PCONS) #IFOB
    `tslm_4` = TSLM(IPCOM ~ PIBA+IPP+IPC+TI+ICIF+PPWTI+DESEM
+PCONS) # ICC
  ) %>%
  accuracy()

#### NOTA: Hyndman recomienda usar criterios de información
para evaluar los modelos
# estos indicadores tienen como objetivo, evaluar el poder
preductivo de los modelos.

modelo_tslm1 <- Dataset_FC_2021 %>%
  model(`tslm_1`=TSLM(IPCOM ~ PIBA+PIBT+IPP+IPC+OM+TI+ICIF+
IFOB+PPEC+PPWTI+ICC+DESEM+PCONS))
#report(modelo_tslm1)

glance(modelo_tslm1) %>%
  select(adj_r_squared, CV, AIC, AICc, BIC)

modelo_tslm2 <- Dataset_FC_2021 %>%
  model(`tslm_2`=TSLM(IPCOM ~ PIBA+IPP+IPC+TI+ICIF+IFOB+PPWTI
+ICC+DESEM+PCONS))
#report(modelo_tslm1)

glance(modelo_tslm2) %>%
  select(adj_r_squared, CV, AIC, AICc, BIC)

modelo_tslm3 <- Dataset_FC_2021 %>%
  model(`tslm_3`=TSLM(IPCOM ~ PIBA+IPP+IPC+TI+ICIF+PPWTI+ICC
+DESEM+PCONS))
#report(modelo_tslm1)

glance(modelo_tslm3) %>%
  select(adj_r_squared, CV, AIC, AICc, BIC)

modelo_tslm4 <- Dataset_FC_2021 %>%

```

```

  model(`tslm_4` = TSLM(IPCOM ~ PIBA+IPP+IPC+TI+ICIF
+PPWTI+DESEM+PCONS))# -ICC

#report(modelo_tslm1)

glance(modelo_tslm4) %>%
  select(adj_r_squared, CV, AIC, AICc, BIC)

#RESULTADOS: El mejor modelo es el modelo No.4
report(modelo_tslm4) #Tomado para la tesis

# Valores ajustados
# Diagrama de tiempo de valores de IPCOM e IPCOM pronosticado
augment(modelo_tslm3) %>%
  ggplot(aes(x = Month)) +
  geom_line(aes(y = IPCOM, colour = "Datos")) +
  geom_line(aes(y = .fitted, colour = "Ajuste")) +
  labs(y = NULL,
       title = "Modelo de regresión múltiple para predicción
de Índice de precios de construcción - Viviendas unifamiliares"
) +
  scale_colour_manual(values=c(Datos="black",Ajuste="#D55E00")) +
  guides(colour = guide_legend(title = NULL))

```

Anexo 5: Predicciones de variables respuesta y explicativas usadas en las pruebas.

Table 23: Predicciones de variables respuesta y explicativas usadas en las pruebas.

Fecha	IPC $\hat{O}U$	IPC $\hat{O}M$	P $\hat{I}B$	IP \hat{P}	IPC	T \hat{I}	IC $\hat{I}F$	PP $\hat{W}TI$	IC \hat{C}	DE $\hat{S}EM$	PC $\hat{O}NS$
2020 ene.	250.5065319	242.9665791	108164.37	104.99	105.62	10.15	47006.26	61.47	37.24	3.80	2928.05
2020 feb.	250.1155535	242.8038594	108220.87	105.66	105.87	10.13	47006.26	61.95	37.24	3.80	2615.76
2020 mar.	249.8471239	242.6736836	108276.19	106.32	106.17	10.11	47006.26	61.95	37.24	3.80	2642.29
2020 abr.	249.6628311	242.5695428	108330.45	105.33	106.52	10.10	47006.26	61.95	37.24	3.80	2834.62
2020 may.	249.5363034	242.4862301	108383.76	104.85	106.32	10.08	47006.26	61.95	37.24	3.80	2736.36
2020 jun.	249.4494346	242.4195798	108436.20	104.67	106.12	10.06	47006.26	61.95	37.24	3.80	2756.07
2020 jul.	249.3897941	242.3662596	108487.88	104.78	105.93	10.04	47006.26	61.95	37.24	3.80	2863.29
2020 ago.	249.3488474	242.3236033	108538.86	104.80	105.74	10.02	47006.26	61.95	37.24	3.80	2789.82
2020 sep.	249.320735	242.2894783	108589.21	105.21	105.90	10.00	47006.26	61.95	37.24	3.80	2749.80
2020 oct.	249.3014343	242.2621782	108639.00	104.22	105.78	9.99	47006.26	61.95	37.24	3.80	2742.64
2020 nov.	249.2881832	242.2403382	108688.28	104.53	105.65	9.97	47006.26	61.95	37.24	3.80	2736.66
2020 dic.	249.2790855	242.2228661	108737.11	104.68	105.54	9.95	47006.26	61.95	37.24	3.80	2771.31
2021 ene.	249.2728394	242.2088884	108785.52	105.48	105.94	9.93	47006.26	61.95	37.24	3.80	2804.77
2021 feb.	249.2685511	242.1977063	108833.57	106.14	106.18	9.91	47006.26	61.95	37.24	3.80	2711.51
2021 mar.	249.2656069	242.1887605	108881.28	106.79	106.48	9.89	47006.26	61.95	37.24	3.80	2719.43
2021 abr.	249.2635856	242.1816039	108928.70	105.79	106.82	9.87	47006.26	61.95	37.24	3.80	2776.86
2021 may.	249.2621978	242.1758787	108975.84	105.30	106.62	9.86	47006.26	61.95	37.24	3.80	2747.52
2021 jun.	249.261245	242.1712984	109022.74	105.10	106.41	9.84	47006.26	61.95	37.24	3.80	2753.41
2021 jul.	249.2605909	242.1676343	109069.43	105.21	106.21	9.82	47006.26	61.95	37.24	3.80	2785.43
2021 ago.	249.2601418	242.1647029	109115.91	105.21	106.01	9.80	47006.26	61.95	37.24	3.80	2763.49
2021 sep.	249.2598334	242.1623578	109162.22	105.62	106.17	9.78	47006.26	61.95	37.24	3.80	2751.54
2021 oct.	249.2596218	242.1604817	109208.37	104.61	106.04	9.76	47006.26	61.95	37.24	3.80	2749.40
2021 nov.	249.2594764	242.1589809	109254.38	104.91	105.92	9.75	47006.26	61.95	37.24	3.80	2747.61
2021 dic.	249.2593766	242.1577802	109300.26	105.05	105.80	9.73	47006.26	61.95	37.24	3.80	2757.96

Fuente: Mayra Viscaíno (2022)