

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE SISTEMAS

UNIDAD DE TITULACIÓN

**CONTRASTACIÓN MODELOS DE PREDICCIÓN DE MACHINE
LEARNING PARA LA INFLACIÓN EN ECUADOR PARA EL
PERÍODO 2001-2020.**

**TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL GRADO DE
MAGISTER EN SISTEMAS DE INFORMACIÓN CON MENCIÓN EN
INTELIGENCIA DE NEGOCIOS Y ANALÍTICA DE DATOS MASIVOS**

FRANK GIOVANNY VILLARREAL VELOZ

frankgvillarreal@gmail.com

Director: Henry Patricio Paz arias

henry.paz@epn.edu.ec

2021

APROBACIÓN DEL DIRECTOR

Como director del trabajo de titulación "*Contrastación modelos de predicción de Machine Learning para la inflación en Ecuador para el período 2001-2020*" desarrollado por Frank Giovanni Villarreal Veloz, estudiante de la Maestría en Sistemas de Información con mención en Inteligencia de Negocios y Analítica de Datos Masivos, habiendo supervisado la realización de este trabajo y realizado las correcciones correspondientes, doy por aprobada la redacción final del documento escrito para que prosiga con los trámites correspondientes a la sustentación de la Defensa oral.

Henry Paz

DIRECTOR

DECLARACIÓN DE AUTORÍA

Yo, Frank Giovanni Villarreal Veloz, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentada para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

La Escuela Politécnica Nacional puede hacer uso de los derechos correspondientes a este trabajo, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

Frank Giovanni Villarreal Veloz

ÍNDICE DE CONTENIDO

LISTA DE FIGURAS	i
LISTA DE TABLAS	i
LISTA DE ANEXOS	ii
RESUMEN	iii
<i>ABSTRACT</i>	iv
INTRODUCCIÓN	2
PREGUNTA DE INVESTIGACIÓN	3
OBJETIVO GENERAL	3
OBJETIVOS ESPECÍFICOS	3
CAPÍTULO I: REFERENCIAL TEÓRICO	4
1.1. INFLACIÓN	4
Impacto de la Inflación	6
Inflación en una economía dolarizada	7
Tipo de Cambio	8
1.2. MACHINE LEARNING	9
Redes Neuronales Artificiales	10
Algoritmos de Regresión	12
CAPÍTULO II: ASPECTOS METODOLÓGICOS	16
CAPÍTULO III: RESULTADOS	20
3.1. EVALUACIÓN DE VARIABLES	20
Evaluación de Estacionariedad	21
Selección de Variables	24
3.2. MODELOS DE REGRESIÓN	29
ARMA	32
ARIMA	33
SARIMA	36
VAR	38
3.3. MODELO DE RNA	40
3.4. COMPARACIÓN DE LOS MODELOS	45

CAPÍTULO IV: DISCUSIÓN.....	47
CAPÍTULO V: CONCLUSIONES Y RECOMENDACIONES.....	48
5.1 CONCLUSIONES	48
5.2 RECOMENDACIONES	50
REFERENCIAS BIBLIOGRÁFICAS	51
ANEXOS	55

LISTA DE FIGURAS

Figura 1. Neurona Artificial	11
Figura 2 - Series histórica IPC (base100 = 2010).....	20
Figura 3 - DLOG(IPC)	23
Figura 4 - Pronóstico del IPC modelo MCO	30
Figura 5 - Pronóstico del IPC modelo MCO CE	32
Figura 6 – Pronósticos del IPC modelo ARIMA.....	35
Figura 7 - Pronósticos del IPC modelo SARIMA	38
Figura 8 - Pronósticos IPC del modelo VAR	40
Figura 9 - RMSE vs Epochs	42
Figura 10 - Pronósticos IPC modelo RNA	44

LISTA DE TABLAS

Tabla 1 - Variables del estudio	16
Tabla 2 - Matriz de correlaciones de los logaritmos de las variables.....	21
Tabla 3 - Prueba de raíz unitaria LOG(IPC).....	22
Tabla 4 - Prueba de raíz unitaria DLOG(IPC).....	22
Tabla 5 - Prueba de raíz unitaria variables explicativas	23
Tabla 6 - Matriz de correlaciones de las diferencias de los logaritmos de las variables ...	24
Tabla 7 - MCO para la DLOG(IPC) con 11 variables regresoras	25
Tabla 8 - MCO para la DLOG(IPC) con 9 variables regresoras	25
Tabla 9 - MCO para la DLOG(IPC) con 8 variables regresoras	26
Tabla 10 - MCO para la DLOG(IPC) con corrección de errores	27
Tabla 11 - Modelo MCO grupo de entrenamiento	29
Tabla 12 - Resultados RMSE de los pronósticos MCE.....	30
Tabla 13 - Modelo MCO CE grupo de entrenamiento	31
Tabla 14 - Resultados RMSE modelo MCO CE	32
Tabla 15 - Correlograma DLOG(IPC).....	33
Tabla 16 - Modelos ARIMA	34
Tabla 17 - Modelo ARIMA(3,3)	34

Tabla 18 - Tabla 9 - Resultados RMSE modelo ARIMA.....	36
Tabla 19 - Modelo SARIMA(3,3,12)	36
Tabla 20 - Modelo SARIMA(3,1,1) ₁₂	37
Tabla 21 - Resultados RMSE modelo SARIMA.....	38
Tabla 22 - Prueba de rezagos.....	39
Tabla 23 - Resultados RMSE del modelo VAR.....	40
Tabla 24 - Pruebas de modelos de RNA	42
Tabla 25 - Pruebas de funciones de activación.....	43
Tabla 26 - Resultados RMSE del modelo RNA.....	44
Tabla 27 - Resumen RMSE por modelo.....	45
Tabla 28 - Comparación con resultados de Mora [32]	47

LISTA DE ANEXOS

Anexo 1 - Salario Nominal Ecuador (USD).....	55
Anexo 2 - Diagramas de dispersión de la primera diferencia del logaritmo de las variables explicativas	55
Anexo 3 - Test de Walt para modelo MCO.....	56
Anexo 4 - Prueba de raíz unitaria en residuos	56
Anexo 5 - Valores de los parámetros del modelo VAR	57
Anexo 6 – Prueba de estabilidad modelo VAR.....	59
Anexo 7 - Diagrama modelo RNA.....	60
Anexo 8 - Pesos de conexiones modelo de RNA	61
Anexo 9 - Sesgos neuronas modelo RNA	63
Anexo 10 - Destinos de exportaciones Ecuador 2019.....	64
Anexo 11 - Coeficiente de determinación ajustado modelo RNA	64

RESUMEN

La predicción de la inflación es de relevancia para la toma de decisiones de política económica y en las perspectivas que están directamente relacionadas con el valor esperado del nivel de actividad agregado en la economía. Los análisis de regresión han sido las técnicas utilizadas con mayor frecuencia; sin embargo, el uso de Redes Neuronales se ha popularizado para el pronóstico de series temporales y el presente trabajo pretende comparar su capacidad predictiva frente a los análisis de regresión. En este sentido se realizaron modelos de regresión (MCO, VAR, ARMA, ARIMA, SARIMA) y se compararon los resultados con el modelo de RNA mediante el uso de la raíz del error cuadrático medio (RMSE).

El período de estudio fue de enero-2001 a diciembre 2020, del cual el 80% se tomó como grupo de entrenamiento y el restante 20% como grupo de prueba de los modelos. En primera instancia, se evaluó la influencia de las variables sobre el comportamiento de la inflación, las cuales resultaron ser los tipos de cambio, precios internacionales de materias primas y el salario nominal. En cuanto al pronóstico del IPC, el modelo de mínimos cuadrados ordinarios con corrección de error de corto plazo (MCO CE) mostró los mejores resultados, seguido del modelo VAR y el modelo de RNA. Sin embargo, se debe mencionar que el modelo de RNA si mostró mejores resultados para el pronóstico de las primeras diferencias del logaritmo del IPC.

Palabras clave: IPC, Pronóstico, Regresión, Redes Neuronales, ARIMA

ABSTRACT

Inflation forecast is relevant for economic policy makers and for perspectives that are directly related to the expected value of the aggregate activity in the economy. Regression analyzes have been the most frequently used techniques; nevertheless, the use of Neural Networks has become popular for time series forecasting, so in that way, this project aims to compare their predictive capability against regression analysis. In this sense, regression models (OLS, VAR, ARIMA, SARIMA) and their results were compared with the ANN model using the root-mean-square error (RMSE).

The period of study started in January-2001 to December 2020, of which 80% was taken as the training group and the remaining 20% as the test group. In the first instance, the influence of the variables over Ecuadorian inflation was evaluated, which turned out to be exchange rates, international commodities prices and nominal wage. In the CPI forecast, the ordinary least squares model with short-term error correction (OLS CE) showed the best results, followed by the VAR model and the ANN model. However, it should be mentioned that the RNA model showed better results for predicting the first differences of log CPI.

Keywords: CPI, Forecast, Regression, Neural Networks, ARIMA.

**CONTRASTACIÓN MODELOS DE PREDICCIÓN DE MACHINE
LEARNING PARA LA INFLACIÓN EN ECUADOR PARA EL
PERÍODO 2001-2020.**

INTRODUCCIÓN

La inflación, en economía, es el “cambio porcentual que experimenta el nivel de precios en un período determinado” [1]. Anticipar las variaciones de la inflación es relevante para las decisiones de política económica y en las perspectivas que están directamente relacionadas con el valor esperado del nivel de actividad agregado en la economía. La medida de la precisión de su pronóstico permite la implementación de medidas antinflacionarias para controlar choques externos y mantener el poder adquisitivo del consumidor. Por ello, la predicción de la inflación se ha vuelto de interés tanto para el sector productivo, sector público e incluso los hogares para su toma de decisiones en función de la capacidad adquisitiva futura.

Las técnicas tradicionalmente empleadas para el pronóstico de la inflación han sido los análisis de regresión, entre estos especialmente: los modelos autorregresivos (AR) en alguna de sus variantes y los modelos de vectores autorregresivos (VAR); sin embargo, el uso de Redes Neuronales, una rama del Machine Learning, se ha popularizado sobre las tareas de regresión y predicción de series temporales, especialmente debido a que recogen mejor las relaciones no lineales subyacentes en las series de tiempo económicas [2].

De acuerdo con Granger [3], en las series de tiempo financieras y económicas es más probable que aparezcan relaciones no lineales, ya que la información que subyace a procesos económicos es de naturaleza no lineal. En tal sentido, los modelos lineales y las técnicas de regresión podrían no ser óptimas para el estudio de estas relaciones. El rápido crecimiento de la capacidad computacional de los últimos años ha permitido la aplicación de métodos de machine learning como las redes neuronales artificiales (RNA) para el modelado y pronóstico de series económicas y financieras. El modelado por RNA permite el estudio de problemas económicos sobre modelos poco definidos, ruido y resistencia de efectos no lineales.

Bajo el antecedente expuesto, el presente trabajo busca comparar los resultados de los pronósticos realizados mediante técnicas de machine learning para la inflación en Ecuador desde el 2001 hasta el 2020. Para ello, se plantea la comparación entre las predicciones de modelos de regresión lineal MCO, ARMA, ARIMA, SARIMA y VAR con

los resultados del modelo de RNA y determinar cuál arroja los mejores resultados y muestra un mejor ajuste a los datos observados.

Pregunta de investigación

¿El modelo de predicción de la inflación basado en RNA proporciona un mejor pronóstico que los algoritmos de regresión lineal?

Objetivo general

Contrastar modelos de predicción de Machine Learning para la inflación en Ecuador para el período 2001-2020.

Objetivos específicos

- Analizar la serie de tiempo de valores mensuales de la inflación, y sus relaciones con variables exógenas.
- Elaborar modelos de regresión (MCO, VAR, ARMA, ARIMA, SARIMA) para determinar su capacidad predictiva y determinar cuál modelo responde mejor al comportamiento de la inflación.
- Elaborar un modelo basado en redes neuronales artificiales (RNA) para comparar su capacidad predictiva con los modelos de regresión.

CAPÍTULO I: REFERENCIAL TEÓRICO

1.1. Inflación

Mantener la estabilidad de precios es uno de los objetivos más importantes de una nación. A tal efecto, el artículo 302 de la Constitución de la República de Ecuador, alusivo a la política monetaria señala el propósito de mantener la estabilidad de precios de acuerdo con el objetivo de estabilidad económica, entendido como “el máximo nivel de producción y empleo sostenibles en el tiempo” [4]. La inflación se contrapone a la estabilidad de precios puesto que conlleva efectos negativos sobre el valor del dinero de los consumidores. De acuerdo con Pablo Zambrano [1] la inflación consiste en “el incremento porcentual en los precios promedio de los bienes y servicios disponibles para la producción y consumo”. Las causas de la inflación pueden ser variadas, “a largo plazo, la inflación resulta del fenómeno monetario que ocurre cuando la cantidad de dinero crece más rápido que el PIB potencial” [5]. No obstante, en el corto plazo son diversos los factores que pueden afectar el nivel general de precios, y para su análisis puede estudiarse principalmente desde dos perspectivas: la inflación de la demanda y la inflación de costos.

“La inflación de la demanda ocurre cuando la demanda agregada se eleva con mayor rapidez que el potencial productivo de la economía.” [6] Es decir, la inflación de la demanda asume que, dada la plena ocupación de los factores de producción, la presión alcista sobre los precios se produce porque el poder adquisitivo de la sociedad es mayor a la capacidad productiva de la economía. Este incremento del poder adquisitivo puede corresponder a la reducción de las tasas de interés, ampliación del gasto público, descenso de impuestos, aumento de la inversión privada, y otros. En macroeconomía, la inflación de la demanda y el papel del dinero como su causa ha sido controversial. Por un lado, para los monetaristas la cantidad de dinero es el factor excluyente del aumento de la inflación, pero, por otro lado, para los keynesianos la inflación no es sólo un fenómeno monetario, sino que depende de los aumentos de la demanda agregada efectiva, particularmente en recesiones económicas[7]. De acuerdo a los monetaristas, en general, la inflación se produce por el déficit fiscal del estado, en dicho escenario los estados incrementan el circulante pero la producción no aumenta, así que “una forma muy dañina de inflación de demanda ocurre cuando los gobiernos incurren en gastos deficitarios y se apoyan en la máquina de hacer billetes para pagar sus déficits” [5]. En resumen, el déficit

fiscal y la inflación se relacionan, puesto que un gobierno con desequilibrios presupuestarios sostenidos enfrenta el impedimento de financiamiento con ninguna otra opción más que la creación de dinero.

La inflación de costos reside en el incremento general de los costos de producción. Este tipo de inflación puede ser causado por el incremento en los salarios, los precios de las materias primas, las tasas de interés, etc. Dicho de otra manera, la inflación de costos se forma al elevarse los costos de producción, ya que existe un número menor de proveedores que pueden soportar márgenes de utilidad cada vez más pequeños, lo cual presiona al alza sobre los precios y reduce la oferta agregada. En definitiva, los costos de producción experimentarían aumentos que luego son transferidos a los precios finales.

La literatura económica menciona también otras teorías sobre la inflación que vale la pena mencionar: inflación estructural, inflación inercial y estanflación. En el caso de la inflación estructural su causa “no se encuentra en el desequilibrio entre la demanda y la oferta global, sino en desajustes sectoriales que afecten a bienes determinados” [8]. Es decir, la inflación estructural es causada por desajustes en cómo se regula y funciona la producción en una economía que afectan el normal funcionamiento del mercado. La inflación inercial se produce porque los diferentes agentes económicos tratan de no perder poder adquisitivo, para lo cual actualizan o indexan los valores nominales de sus productos o servicios a la inflación esperada. Por ende, la inflación del período actual dependerá de la inercia inflacionaria procedente del período anterior. La inflación con estancamiento o estanflación, “supone una subida notable y sostenida de los precios, junto con el aumento progresivo de los niveles de paro” [8]. Este tipo de inflación generalmente es observada en períodos de recesión y depresión.

La medida más divulgada para medir la inflación es a través del Índice de Precios al Consumidor (IPC), el cual calcula los cambios promedio de los precios al consumidor de un conjunto de bienes y servicios. En Ecuador, el Instituto Nacional de Estadística y Censos (INEC) es la institución que calcula la inflación mensualmente utilizando el IPC, el cual registra los movimientos en los precios finales de consumo, en un período de tiempo determinado, para una canasta fija de 359 productos distintivos del consumo nacional:

$$Tasa\ de\ inflación = \frac{IPC_t - IPC_{t-1}}{IPC_{t-1}} \quad (1)$$

La metodología para el cálculo IPC parte del Manual de Índices de Precio al Consumidor: Teoría y Práctica¹. El IPC es un indicador calculado en 9 ciudades del país que mide el cambio en los precios al consumo final en hogares de diferentes ingresos del área urbana, para lo cual se toman 25,350 muestras de precios [9]. Entre los artículos que se incluyen se puede encontrar alimentos y bebidas, prendas de vestir, salud, transporte, comunicaciones, entre otros [10]. El IPC se constituye como una variable relevante para la economía, política y la sociedad ya que es un insumo directo de las cuentas nacionales y la política económica.

Impacto de la Inflación

La inflación tiende a reducir la calidad de vida de la sociedad porque merma su capacidad adquisitiva. Mientras mayor es la inflación, menos bienes se podrían adquirir si el salario nominal no se incrementa a un ritmo similar. La inflación puede tener diversos efectos sobre una economía, por un lado, una inflación moderada puede disminuir el valor de las deudas, aumentar la competitividad de los productos de un país y beneficiar a los poseedores de capital porque también incrementa el valor de sus activos, pero, por otro lado, impacta principalmente sobre las personas con menos recursos afectando directamente su poder adquisitivo, desincentivando el ahorro y tiene efectos negativos sobre la redistribución del ingreso y la riqueza. La inflación castiga a la gente que carece de activos haciéndolos más difíciles de adquirir, en este sentido, “la inflación constituye un impuesto regresivo que no ha sido aprobado mediante ley y que grava en mayor proporción a los individuos y familias más pobres” [1].

En cuanto a la producción, la inflación elevada frena las inversiones al generar incertidumbre sobre las decisiones a futuro. Frente a tasas de inflación elevadas es evidente la relación inversa entre la producción e inflación, sin embargo, la relación es

¹ Manual del Índice de Precios al Consumidor: Teoría y Práctica, producido por la Organización Internacional del Trabajo (OIT), la Organización de Cooperación y Desarrollo Económicos (OCDE); la Oficina Estadística de las Comunidades Europeas (EUROSTAT); la Comisión Económica de las Naciones Unidas para Europa (UNECE); el Banco Mundial (BM); y el Fondo Monetario Internacional (FMI)

discutible cuando la inflación es moderada. Una tasa de inflación moderada puede conllevar consecuencias beneficiosas en el corto plazo, ya que al provocar el incremento de la demanda agregada puede elevar la producción y el empleo (si estos se encontraban por debajo del nivel de equilibrio).

Una inflación moderada manejada adecuadamente puede ser positiva para el crecimiento económico. Las autoridades y bancos centrales pueden utilizar diversos mecanismos para controlar la inflación, por ejemplo, pueden modificar el monto de gasto público, los impuestos o las tasas de interés, e incluso diferentes combinaciones de políticas fiscal y monetaria para incentivar a los distintos actores a deshacerse o adquirir liquidez. La inflación es una de las principales áreas de estudio de los bancos centrales, porque al controlar la inflación se crea un clima de estabilidad financiera y monetaria necesaria para el crecimiento económico. Dentro de su política monetaria, varios países tratan de controlar la inflación mediante el establecimiento de metas de inflación, la razón principal es que “permite que la política monetaria responda mejor a eventos adversos” [11]. Por ello, los bancos centrales generalmente fijan las metas de manera flexible, es decir, apuntan a lograrla a mediano plazo.

Inflación en una economía dolarizada

Luego de sufrir una grave crisis económica e inflacionaria el año 1999, Ecuador adoptó el Dólar de Estados Unidos como moneda de curso legal. De acuerdo con registros del Banco Mundial, en solo una semana la moneda había pasado de 7.000 sucres por dólar a 19.000 y Ecuador alcanzó una inflación del 96% en 1999 [12]. La dolarización logró controlar la inflación, que para el 2003 era menor al 8%, estabilizando la economía y provocando un crecimiento moderado. En el caso particular ecuatoriano, la dolarización conllevó también la pérdida de control sobre la oferta monetaria por el Banco Central del Ecuador (BCE), lo cual impide el manejo de su política monetaria y cambiaria.

Ecuador es un país de renta media con un PIB per cápita de USD 6.184 [13], cuyas principales exportaciones son: Petróleo Crudo (37,1%), Camarón (14,8%), Banano (14,4%), Enlatados de Pescado (5,8%) y Flores (4,0%) [14], que al ser en su mayoría bienes primarios hacen que el país sea especialmente sensible a las variaciones de los precios internacionales de las materias primas o *commodities*. De acuerdo con el trabajo de Gachet *et al.* [15], los niveles de inflación en Ecuador en 2008 “son principalmente

causados por los precios internacionales de commodities, tipos de cambio y política pública”. Sin embargo, hacen la precisión de que los precios internacionales de los commodities deberían ser vistos como temporales.

En un trabajo posterior realizado en la Universidad del Azuay, se determinó que, para el período enero-2000 a abril-2018, los principales determinantes de la inflación en Ecuador fueron el tipo de cambio real y los salarios [16]. En relación al tipo de cambio real, el gobierno no tiene la posibilidad de organizar política monetaria dirigida a aumentar el circulante y la competitividad de los productos locales, que, al combinarla con la dependencia de la exportación de materias primas, provoca que los procesos productivos sean afectados por variaciones cambiarias que empujan una inflación de costos. Por tal motivo, la política del gobierno debería ser “sustituir poco a poco la importación de bienes intermedios por bienes nacionales” [16].

La dolarización ha impuesto la renuncia a la posibilidad de hacer política monetaria en Ecuador optando por el manejo consistente de la Reserva Federal de los Estados Unidos. Su principal ventaja es que ha controlado la inflación al eliminar la posibilidad de emisión de moneda lo cual ha generado estabilidad económica, pero por otro lado hacen a Ecuador susceptible a crisis relacionadas con la inestabilidad de los precios de materias primas, particularmente el petróleo [17], y por la pérdida de competitividad asociada a la devaluación de la moneda de los países vecinos.

Tipo de Cambio

De acuerdo con Francisco Mochón “el tipo de cambio, es el precio relativo de una moneda por otra distinta: es el número de unidades de la moneda extranjera que hay que entregar para obtener una unidad de moneda nacional” [18], en otras palabras, el tipo de cambio hace referencia al precio relativo entre dos monedas: es decir, cuántas unidades se necesitan de una moneda para adquirir una unidad de otra. Si el tipo de cambio de la moneda local con respecto a otra aumenta, esto causa que la moneda se deprecie, es decir, que pierda su valor frente a la otra. La devaluación o revaluación de la moneda tiene efecto directo sobre las exportaciones de un país, puesto que si la moneda de curso local se revalúa encarece nuestros productos en el extranjero, y en caso de devaluación pueden incrementarse las exportaciones, por lo cual se desprende una relación inversa entre la inflación y el tipo de cambio.

1.2. Machine Learning

El Machine Learning, generalmente traducido como Aprendizaje Automático, es un área de estudio de la Inteligencia Artificial que explora la construcción de algoritmos computacionales que aprendan de la experiencia. De acuerdo con Chris Mattmann, machine learning “está caracterizado por software que aprende de experiencia previa. Tal programa mejora su desempeño mientras más ejemplos están disponibles” [19]. En este sentido, machine learning es similar a compartir recetas usadas por otros programas. Típicamente, se construye el algoritmo en 2 etapas: aprendizaje e inferencia. La receta es generada en la etapa de aprendizaje cuando la data es analizada y resumida en un modelo (receta). En la etapa de inferencia, el modelo es usado para hacer observaciones sobre data aún no observada, es como usar la receta con otros ingredientes, en esta etapa se prueba el desempeño del modelo [20].

Machine learning, en términos de Peter Harrington, “yace en la intersección de la ciencia de la computación, ingeniería, estadística y frecuentemente aparecen otras disciplinas” [21], por ello puede ser aplicada a muchos campos de estudio que necesiten analizar e interpretar data como la economía. Susan Athey, economista estadounidense profesora de Economía de la Tecnología en Stanford afirmó que “el machine learning tendrá un impacto dramático en el campo de la economía dentro de un corto período de tiempo” [22].

Machine learning generalmente, se divide en tres ámbitos: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje por refuerzo.

En el aprendizaje supervisado la máquina aprende de la data en donde se ha identificado y especificado la variable objetivo, en este punto la máquina reduce la tarea a buscar un patrón que llegue desde la data de entrada hasta la variable objetivo. “La esencia del algoritmo de machine learning supervisado es descubrir los parámetros de un modelo que resulten en el menor costo” [19]. Las principales herramientas del aprendizaje supervisado incluyen regresión y clasificación. Si la variable objetivo tiene muchos valores, y éstos tienen un orden natural es un modelo de regresión, mientras si la variable objetivo es una variable discreta se trata de un problema de clasificación. Algunos de los algoritmos usados para regresión son Mínimos Cuadrados Ordinarios (MCO), Vectores Autorregresivos (VAR), Autorregresivo de Media Móvil (ARMA), Autorregresivo Integrado

de Media Móvil (ARIMA), y su extensión con modelado del componente estacional (SARIMA). Para clasificación, algunos de los algoritmos usados son: Regresión Logística, K-Nearest Neighbors y Árbol de Decisión.

En el aprendizaje no supervisado se trata de “entender y encontrar una descripción concisa de la data a través de pasivamente mapear y agrupar la data de acuerdo con algunos principios de ordenamiento” [23]. En otras palabras, en el aprendizaje no supervisado la data solo constituye un conjunto de objetos sin una relación subyacente asociado a ellos, y su meta principal es crear grupos de acuerdo a un criterio de similitud. Las dos herramientas principales del aprendizaje no supervisado son el clustering (agrupamiento) y la reducción dimensional. El clustering “es el proceso de dividir data en grupos individuales de objetos similares” [19], entre los algoritmos más populares podemos encontrar K-means, DBSCAN y Gaussian Mixture Model. Por otro lado, la reducción dimensional “se trata sobre manipular la data para verla desde una perspectiva más simple” [19]. La simplificación de la data ayuda tanto a la visualización como al desempeño del procesamiento de la información, algunos de los algoritmos usados son el Análisis de Componentes Principales (PCA, por sus siglas en inglés) y Autoencoders.

El aprendizaje por refuerzo está relacionado al aprendizaje de agentes mediante prueba y error que procuran recibir una recompensa. “El aprendizaje por refuerzo es un tipo de machine learning que interactúa con el medioambiente para aprender cuál combinación de acciones rinden los resultados más favorables” [19]. Bajo este esquema la máquina intenta aprender basada en un sistema de premios y castigos, que a diferencia de los modelos de aprendizaje supervisado y no supervisado intenta maximizar la recompensa, sin minimizar el coste. Este tipo de aprendizaje automático es muy usado en el campo de la robótica ya que puede acelerar los procesos de automatización.

Redes Neuronales Artificiales

Las Redes Neuronales Artificiales (RNAs) son algoritmos del machine learning que desempeñan una conducta análoga a las neuronas humanas y su aptitud para conectarse, interrelacionándose en varias capas. Las redes neuronales artificiales “consisten de varias unidades de procesamiento no lineales y parametrizadas conectadas por uniones dadas por los pesos de la variable numérica” [24]. En otras palabras, los

modelos de redes neuronales artificiales integran diferentes entradas que son multiplicadas por los pesos de las conexiones antes de alcanzar la principal unidad de procesamiento.

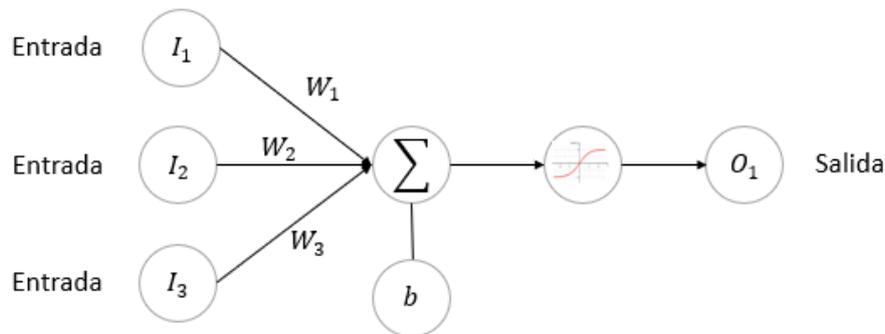


Figura 1. Modelo Básico de RNA

Previo a la salida de una neurona puede incluirse una función que filtra o altera el resultado antes de proseguir a la siguiente neurona, esta es la función de activación y permite a la red resolver problemas más complejos y provoca que los modelos sean generalmente no lineales. El proceso de aprendizaje (entrenamiento) de la red neuronal se da por el ajuste de los pesos (w) y sesgos (b) en la red para generar la respuesta particular deseada. Las tres funciones de activación más usadas son *sigmoid*, *hiperbolic tangent* y un tipo de función de rampa llamada *Unidad de Rectificación Lineal* (ReLU por sus siglas en inglés). Matemáticamente, si la salida para la Figura 1 se generaliza para cualquier número de entradas puede expresarse como:

$$O_1 = f\left(\sum_{i=1}^n w_i \cdot I_i + b\right) \quad (2)$$

Las principales redes neuronales artificiales que utilizan aprendizaje supervisado son las Redes Neuronales Profundas, Redes Neuronales Recurrentes y Redes Neuronales Convolucionales. Las Redes Neuronales Profundas son “una clase de técnicas del machine learning que explotan varias capas de procesamiento de información no lineal para extracción y transformación supervisada y no supervisada” [25], la idea subyacente es que con un mayor número de capas neuronales se puede optimizar las predicciones de datos complejos. Las Redes Neuronales Recurrentes “permiten el funcionamiento de las conexiones de retroalimentación entre las neuronas para que cada neurona esté conectada a otras, incluso consigo misma, para ejecutarse como un sistema dinámico

que depende de un momento pasado en el tiempo” [24], esto permite que la información se guarde durante algunas etapas de entrenamiento y el modelo pueda utilizar rezagos e incluso ser autorregresivo. Este tipo de red neuronal es generalmente utilizada para analizar información de series de tiempo financieras y económicas. Las Redes Neuronales Convolucionales, usadas principalmente en el procesamiento de imágenes, son un tipo de red “con aprendizaje supervisado que procesa sus capas imitando al córtex visual del ojo humano para identificar distintas características en las entradas que en definitiva hacen que pueda identificar objetos” [26], para luego conectarla con una red neuronal tradicional que no evaluará cada pixel sino el conjunto de características en la imagen.

Las redes neuronales artificiales son consideradas como un método bastante efectivo para analizar y predecir diferente tipo de información. Particularmente para las series de tiempo, “la eficiencia del proceso de aprendizaje es una buena herramienta para predecir problemas complejos en series de tiempo no lineales” [2]. Las redes neuronales tienen ciertas ventajas sobre los métodos convencionales ya que son más flexibles y tienen mayor precisión para analizar patrones complejos. Sin embargo, de acuerdo con Vochozka, “la desventaja de las redes neuronales es su necesidad de una gran cantidad de información de entrada” [27], no siempre disponible en series económicas; además, las redes neuronales se convierten en “cajas negras porque es imposible descifrar cómo son estimadas las relaciones de las capas ocultas” [28].

Algoritmos de Regresión

“La regresión es el estudio de cómo encajar mejor una curva para resumir la información y es uno de los más poderosos y mejor estudiados algoritmos del aprendizaje supervisado” [19]. Los análisis de regresión son una herramienta fundamental en la econometría, que es una rama de la economía la cual “se define como la ciencia social en la cual las herramientas de la teoría económica, las matemáticas y la inferencia estadística se aplican al análisis de los fenómenos económicos” [14].

Dentro de los métodos de regresión, el método de Mínimos Cuadrados Ordinarios (MCO), también conocido como modelo de Gauss, modelo clásico o estándar de regresión lineal es el principio de la teoría econométrica, este modelo se puede expresar como:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \quad (3)$$

Donde Y_i es la variable dependiente o regresada, X_i es la variable independiente o regresora, β_1 es el parámetro de intercepto, β_2 es el parámetro de pendiente y u_i es el término de perturbación estocástico (error estocástico o ruido blanco). Los estimadores calculados mediante este método “adquieren ciertas propiedades resumidas en el teorema Gauss-Markov, el cual plantea que [...] los estimadores de mínimos cuadrados tienen una varianza mínima. En resumen, son los mejores estimados lineales insesgados” [29]. La precisión de los estimadores en MCO se mide por su error estándar y la bondad de ajuste del modelo se mide con el coeficiente de correlación.

En economía gran parte de la información está dada como series de temporales, las cuales son “un conjunto de observaciones sobre los valores de una variable en diferentes momentos” [29]. Su análisis econométrico parte de la suposición de que los valores son resultado del componente tendencial, estacional y un término de perturbación estocástico. Algunos de los métodos que han adquirido popularidad en el análisis de series temporales son el Autorregresivo Integrado de Promedios Móviles (ARIMA por sus siglas en inglés) y el de Vectores Autorregresivos (VAR) [30].

Los modelos ARIMA parten de la metodología conocida también como Box-Jenkins, de acuerdo con Gujarati y Porter [29]: “el interés de estos métodos de pronóstico no está en la construcción de modelos uniecuacionales o de ecuaciones simultáneas, sino en el análisis de las propiedades probabilísticas, o estocásticas, de las series de tiempo económicas por sí mismas según la filosofía de que los datos hablen por sí mismos”

En contraste de los modelos de regresión clásicos en los que la variable regresada Y_t puede ser explicada por n variables regresoras X_i , en los modelos autorregresivos la variable Y_t se explica por sus valores rezagados. Partiendo del modelo autorregresivo (AR) se obtiene la siguiente ecuación:

$$(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + u_t \quad (4)$$

Donde δ es la media de Y . Este modelo corresponde a un proceso autorregresivo de primer orden AR(1) los valores de Y se expresan como su desviación respecto a su valor

medio. Dicho de otra manera, este modelo pronostica que el valor de Y_t alrededor de su media es una proporción α_1 de su valor en el período anterior más un término de perturbación aleatoria u_t . Si en el modelo se incluyera un rezago adicional se tendría un proceso autorregresivo de segundo orden AR(2), y así en general se tendrían procesos autorregresivos del orden p , o AR(p) donde p es el número de rezagos incluidos en el modelo.

Otra forma para generar Y en el tiempo t puede ser a través de un proceso de promedios móviles (MA por sus siglas en inglés), esta ecuación considera que Y_t es igual a una constante μ más un promedio móvil de los términos de error actual y pasado, siguiendo esta expresión:

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} \quad (5)$$

Al igual que en un modelo AR, su orden dependerá del nivel de rezagos que se incluyan donde q es el número de rezagos MA(q). Si la serie de tiempo en análisis tiene características de AR y MA se tiene un proceso de regresión ARMA, que no es sino la combinación de un proceso autorregresivo y de promedios móviles ARMA (p, q):

$$Y_t = \theta + \alpha_1 Y_{t-1} + \beta_0 u_t + \beta_1 u_{t-1} \quad (6)$$

En este caso θ es constante y actúa como intercepto al eje vertical. Los modelos ARMA se ajustan mediante MCO para minimizar el término de error en los parámetros. El modelo ARMA “se basa en el supuesto de que las series de tiempo consideradas son (débilmente) estacionarias” [29], es decir su media y varianza son constantes y su covarianza es invariante. “La no estacionariedad, una propiedad común de muchas series de tiempo económicas y financieras, significa que una variable no tiene una tendencia clara a retornar a un valor constante” [30], por lo cual su análisis se vuelve poco práctico. Una serie de tiempo no es estacionaria suele serlo en sus primeras diferencias, así al convertir una serie en estacionaria mediante diferenciarla d veces y aplicar el modelo ARMA se obtiene un modelo ARIMA(p, d, q), es decir, un modelo Autorregresivo Integrado de Promedios Móviles donde, p es el número de rezagos de AR, d el número de veces que la serie se ha diferenciado para hacerla estacionaria y q el número de rezagos de MA. Si se considera un modelo ARIMA(1,1,1) se tendría la siguiente ecuación:

$$(Y_t - Y_{t-1}) = \theta + \alpha_1(Y_t - Y_{t-1})_{t-1} + \beta_0 u_t + \beta_1 u_{t-1} \quad (7)$$

Por la naturaleza misma de las series de tiempo económicas, en las que influyen la época del año en el que se toma la medición, como pueden ser períodos de incremento en el gasto a finales del año o en períodos de vacaciones es útil eliminar el componente estacional para concentrar el análisis en la tendencia subyacente. Para ello, se realiza la corrección estacional, lo cual nos da como resultado un modelo Autorregresivo Integrado de Promedios Móviles con Ajuste de Estacionalidad (SARIMA, por sus siglas en inglés).

En el caso de causalidad multivariada se utiliza la técnica de autorregresión vectorial (VAR, por sus siglas en inglés), “un VAR es un modelo de ecuaciones [...] que es muy útil cuando existe evidencia de simultaneidad entre un grupo de variables y que sus relaciones se transmiten a lo largo de un determinado número de períodos” [30]. Los modelos VAR partes de un modelo estructural dinámico, que en el caso de dos variables se expresaría como:

$$X_t = \gamma_1 + \sum_{i=1}^n \alpha_i Y_{t-i} + \sum_{j=1}^n \beta_j X_{t-j} + u_{1t} \quad (8)$$

$$Y_t = \gamma_2 + \sum_{i=1}^n \gamma_i Y_{t-i} + \sum_{j=1}^n \delta_j X_{t-j} + u_{2t} \quad (9)$$

Entre algunas de las ventajas de un modelo VAR se cita que “las predicciones obtenidas mediante este método son en muchos casos mejores que las obtenidas con modelos de ecuaciones simultáneas más complejos. [29]” y que es relativamente fácil de especificar y de estimar, y las variables pueden ser no-estacionarias. Sin embargo, también se mencionan que “con frecuencia es difícil interpretar los coeficientes individuales estimados en los modelos VAR, los practicantes de esta técnica a menudo estiman la llamada función de impulso- respuesta (FIR) [29]”

CAPÍTULO II: ASPECTOS METODOLÓGICOS

En este análisis se utilizará información histórica con periodicidad mensual, que va desde enero del 2001 hasta diciembre del 2020 (240 observaciones mensuales), para ello se dispuso de diversas fuentes como el INEC, el BCE, Ministerio de Economía y Finanzas (MEF) y el Banco Mundial (BM). El IPC con base al 2010 será tomado como proxy para la inflación y como variable dependiente en adelante.

Las variables independientes para su prueba inicial han sido tomadas en referencia a trabajos previos (Gachet *et al.* [15] de 2008, Diana Morán [32] de 2014, Santiago Guerrero [33] de 2017, Anthony Mora [34] del 2019 y Adrián Calderón [16] del 2019) y de acuerdo a la teoría económica [18] [35] sobre la naturaleza y determinantes de la inflación:

Tabla 1 - Variables del estudio

N°	Variable	Descripción	Tipo	Fuente
1	IPC	Índice de Precios al Consumidor con base = 2010	Dependiente	INEC
2	TCR	Índice Tipo de Cambio Efectivo Real con base = 2014	Independiente	BCE
3	USDCLP	Tipo de Cambio de US Dólar a Peso Chileno	Independiente	MEF
4	USDCNY	Tipo de Cambio de US Dólar a Yuan Chino	Independiente	MEF
5	USDCOP	Tipo de Cambio de US Dólar a Peso Colombiano	Independiente	MEF
6	USDPEN	Tipo de Cambio de US Dólar a Sol Peruano	Independiente	MEF
7	USDEUR	Tipo de Cambio US Dólar a Euro	Independiente	MEF
8	WTI	Precio mensual del West Texas Intermediate	Independiente	BM
9	AgIndex	Índice Mensual de Precios Agrícolas con base = 2010	Independiente	BM
10	SalarioNominal	Salario Nominal anual ecuatoriano en USD	Independiente	Varias Fuentes

11	GastoPublico	Gasto Total mensual en operaciones del sector público no financiero (USD Millones)	Independiente	BCE
12	OfertaMonetaria	Total de las Especies Monetarias en Circulación, más Moneda Fraccionaria, más Dinero Electrónico, más Depósitos a la Vista (USD Millones)	Independiente	BCE

En cuanto a los tipos de cambio, la información fue tomada de los registros del BCE. Se consideró como variables, a más del TCR, el tipo de cambio de los principales destinos de las exportaciones ecuatorianas cuya moneda no sea el dólar estadounidense: Europa (19.38%), China (12,5%), Chile (6,56%), Perú (3,34%) y Colombia (3,73%) [36].

Para los precios internacionales de materias primas se tomó la información disponible en los registros del Banco Mundial. Se incluyó el West Texas Intermediate (CL1) que es el crudo de referencia para Ecuador y el índice de precios de productos agrícolas (AgIndex). El índice de precios agrícolas del Banco Mundial (base = 2010) está compuesto por los precios nominales, en dólares estadounidenses, de 56 productos agrícolas agrupados en bebidas, comida y materia prima, lo cual se constituye como una buena aproximación al comportamiento general del precio internacional de los productos agrícolas ecuatorianos ya que entre los productos del indicador se incluyen: banano, café, cacao y camarones, los cuales son los principales bienes de exportación no petrolera de Ecuador de acuerdo con la información citada por la revista Ekos [14]

La selección de variables independientes más influyentes sobre el IPC se realizará mediante un modelo preliminar de MCO. En el modelo preliminar de MCO se analizarán y seleccionarán las variables con mayor significancia. Las variables seleccionadas servirán como variables explicativas para el modelo final de MCO y de variables de entrada para el modelo de RNA. A priori se espera, de acuerdo a la teoría económica, que la variable de mayor influencia en el modelo sea la Oferta Monetaria; no obstante, el trabajo de Calderón *et al.* muestra que el salario nominal y el tipo de cambio real también tienen impacto en el comportamiento de la inflación [16].

El manejo de series de tiempo requiere un enfoque hacia series no estacionarias. Por ello, se evaluará la estacionariedad de las variables aplicando la prueba Dickey-Fuller Aumentada. De acuerdo con Gujarati y Porter [29], si esta evaluación indica la presencia de raíz unitaria en la serie, se presentarían problemas en la inferencia estadística por lo cual deberá ser diferenciada y evaluada nuevamente hasta que la serie sea estacionaria.

Una vez corregidas las series en cuanto a su estacionariedad se dividirá el *dataset* de forma consecutiva y ordenada en grupo de entrenamiento y prueba, tomando un 80% y 20% respectivamente. Los modelos autorregresivos serán construidos siguiendo la metodología Box-Jenkins [29]. Esta metodología sigue cuatro pasos:

1. Identificación: Encontrar los valores apropiados de p, d y q
2. Estimación: Estimar los parámetros de los términos autorregresivos.
3. Diagnóstico: Evaluar la distribución de los residuos
4. Pronóstico

Los modelos autorregresivos suelen tener mejores pronósticos que los modelos econométricos tradicionales en particular en el corto plazo [29], por ello se espera que su curva de regresión esté mejor ajustada en los primeros datos del grupo de prueba.

El algoritmo de entrenamiento para la RNA será el de retropropagación. No existe una regla generalmente aceptada para la selección del número de nodos; sin embargo, se han propuesto varias guías [37], en este proyecto el número de nodos ocultos será 2^k , con k incrementando de 1 a 5 para evaluar mediante ensayo y error el modelo que arroje el menor RMSE. El número de ciclos y error permitido también serán fijados por ensayo y error. Una vez definidos los parámetros del modelo de RNA, se generarán los valores pronosticados por la red utilizando el grupo de entrenamiento y grupo de prueba para determinar el RMSE en ambos grupos similar a lo realizado por Egüez [38] en sus modelos de predicción.

Hecho esto, se ejecutarán los modelos de regresión en los dos grupos para evaluar su coeficiente de determinación (r^2) y la raíz del error cuadrático medio (RMSE, por sus siglas en inglés). El RMSE mide la diferencia entre las observaciones reales y los pronósticos mediante la siguiente ecuación:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \quad (10)$$

Finalmente, se compararán los resultados de los modelos de regresión con el modelo de RNA a través de la raíz del error cuadrático medio de cada uno. En modelos de RNA, donde el objetivo es reproducir el comportamiento de los datos minimizando la función de coste o error, entendiendo que la función de coste mide la diferencia entre el valor de salida y el valor real de la red neuronal, se utilizará el estimador RMSE como función de coste [38].

CAPÍTULO III: RESULTADOS

3.1. Evaluación de Variables

La inflación mide cambio porcentual en el nivel de precios de los bienes y servicios de una economía, y para su cálculo se miden los cambios en el IPC. El IPC es un índice que se calcula mensualmente con referencia a una base igual a 100. La interpretación de sus valores es sencilla, un valor por encima del 100 implica que los precios han subido en ese porcentaje, y un valor por debajo del 100 implica que los precios han bajado en ese porcentaje. En particular, si un IPC (base=2010) tiene un valor de 120 en julio de 2018 significa que los precios han subido 20% para julio de 2018 respecto a su base. Por otro lado, si el valor del IPC a diciembre de 2005 fue de 80, significa que los precios eran 20% menores a esa fecha. Para el período enero-2001 a diciembre-2020, el IPC (base=2010) de Ecuador publicado por el BCE se muestra en la figura 2

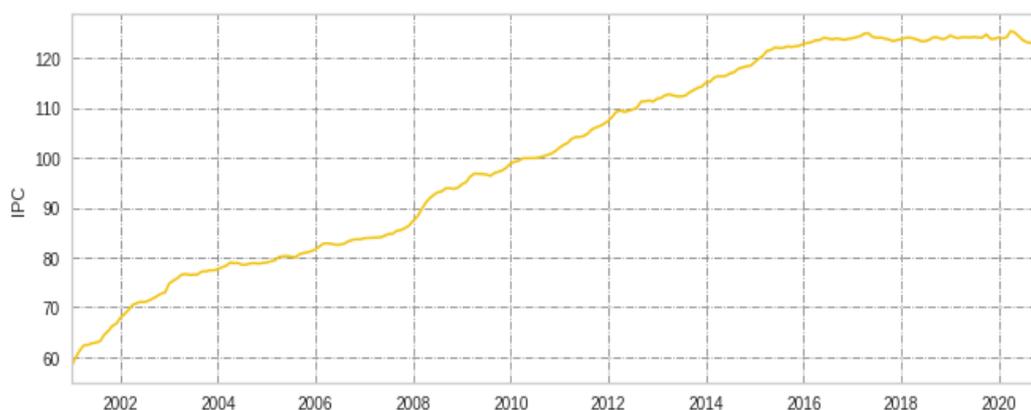


Figura 2 - Serie histórica IPC (base100 = 2010)

La serie del IPC cuenta con 240 observaciones, el valor más bajo registrado fue 58,29 en enero de 2001 y el más alto fue 125,40 registrado en abril de 2020. La tendencia ha sido principalmente creciente desde el 2001; sin embargo, desde el 2016 se observa que el IPC empieza a estabilizarse alrededor de los 123 puntos.

Las variables explicativas para el IPC que se evaluarán son: tipo de cambio, precio internacional de materias primas, salario nominal, gasto público y oferta monetaria. Como parte del tipo de cambio se incluyen: tipo de cambio real (TCR), dólares de estadounidenses a; pesos chilenos (Exch_Chil), yuanes chinos (Exch_Chin), pesos colombianos (Exch_Col), soles peruanos (Exch_Per) y euros (Exch_Eur). Como materias

primas se incluyen el precio del WTI (Generic 1st 'CL' Future o CL1) y el índice de productos agrícolas del Banco Mundial (AgIndex). Para todas las series se transformará a logaritmos para suavizar los cambios y oscilaciones, además que permite medir los cambios como elasticidades de cada una de las variables. Al evaluar la correlación de Pearson del logaritmo de las variables independientes respecto al logaritmo de la variable dependiente se obtienen los resultados que se muestran en la tabla 2:

Tabla 2 - Matriz de correlaciones de los logaritmos de las variables

Variable	Correlación de Pearson
LOGTCR	-0,678584
LOGUSDCLP	0,109667
LOGUSDCNY	-0,8633697
LOGUSDCOP	0,24418
LOGUSDEUR	-0,242425
LOGUSDPEN	-0,324949
LOGWTI	0,420392
LOGAGIND	0,752918
LOGSALARIONOMINAL	0,996528
LOGGASTOPUBLICO	0,948299
LOGOFERTAMONETARIA	0,989457

Antes de un análisis más profundo, la tabla muestra elevadas correlaciones respecto al LOGTCR, LOGUSDCNY, LOGSALARIONOMINAL, LOGGASTOPUBLICO y LOGOFERTAMONETARIA, lo cual implicaría que la relación lineal entre las dos variables es casi perfecta, sin embargo, al evaluar series de tiempo es preciso evaluar la estacionariedad puesto que esta relación *perfecta* puede deberse a la inercia propia de las series de tiempo, simplemente tienden a incrementar en el tiempo, y no debido a una relación de causalidad subyacente.

Gráficamente se puede evaluar la estacionariedad del IPC en la figura 2, esta valoración a priori, muestra que la serie es no estacionaria en su media porque se visualiza una tendencia creciente muy marcada; sin embargo, si es estacionaria respecto a su varianza dado que las fluctuaciones alrededor de la tendencia central muestran, en términos generales, una amplitud constante. En el siguiente apartado se realizará la prueba de Dickey-Fuller Aumentada para determinar formalmente la estacionariedad de la serie IPC.

Evaluación de Estacionariedad

El primer paso para trabajar con series de tiempo es evaluar su estacionariedad, es decir evaluar que su media y varianza permanezcan constantes en el período de estudio. Para

evaluar la estacionariedad en la serie del LOG(IPC) se aplicó la prueba de raíz unitaria (Dickey-Fuller Aumentada) y como resultado se obtuvo que el IPC no es estacionario, tal como se recoge en la tabla 3:

Tabla 3 - Prueba de raíz unitaria LOG(IPC)

Null Hypothesis: LOGIPC has a unit root Exogenous: Constant Lag Length: 12 (Fixed)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.309668	0.6254
Test critical values:	1% level	-3.459101
	5% level	-2.874086
	10% level	-2.573533

La probabilidad de que IPC tenga raíz unitaria es el 63%, y no es estadísticamente significativo al nivel del 5% por lo tanto la serie es no estacionaria. Con este resultado se confirma lo que se evaluó visualmente en la figura 2, ya que se observa una notoria tendencia creciente en el IPC. Se espera que en sus primeras diferencias la serie sea estacionaria, porque las series de tiempo suelen ser estacionarias en sus variaciones. Al realizar la prueba en su primera diferencia del LOG(IPC) se obtuvieron los siguientes resultados de la prueba de raíz unitaria (Tabla 4):

Tabla 4 - Prueba de raíz unitaria DLOG(IPC)

Null Hypothesis: D(LOGIPC) has a unit root Exogenous: Constant Lag Length: 12 (Fixed)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-2.910178	0.0457
Test critical values:	1% level	-3.459231
	5% level	-2.874143
	10% level	-2.573563

El valor de la probabilidad del estadístico implica el rechazo de la hipótesis nula (IPC no tiene raíz unitaria) al nivel de confianza del 5% y se asume que la serie en su primera diferencia es estacionaria; por lo cual, el orden de integración del LOG(IPC) es $I(1)$ y debe modelarse en su primera diferencia:

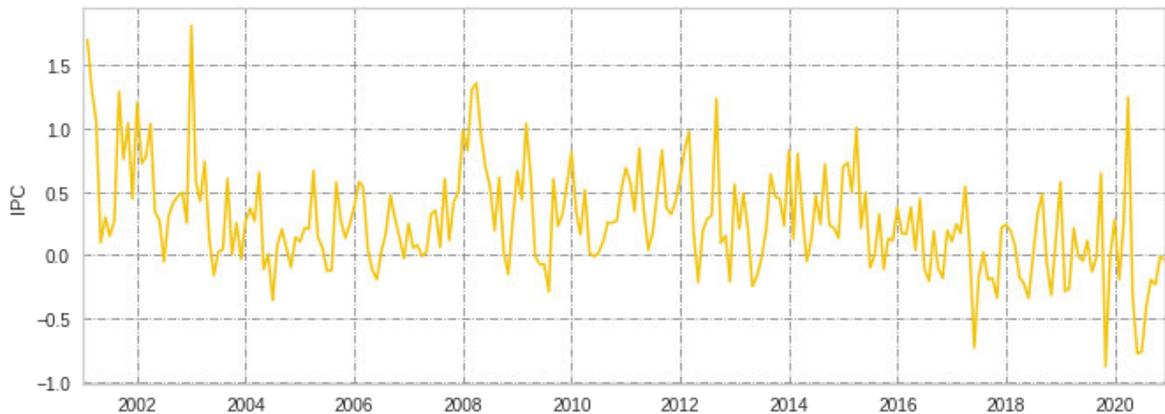


Figura 3 - DLOG(IPC)

El gráfico del DLOG(IPC) muestra que la tendencia creciente desaparece en las primeras diferencias y debido a la transformación logarítmica las variaciones tienen una variación estable. En la evaluación de estacionariedad para las variables explicativas se prescindirá del Salario Nominal puesto que su comportamiento escalonado (anexo 1) con periodicidad anual haría que en sus primeras diferencias se anularan cerca del 92% de las observaciones, lo cual sería de poca utilidad para el desarrollo del modelo.

La evaluación de estacionariedad de las variables independientes mostró que todas las series son no estacionarias al nivel del 5%. Al igual que el LOG(IPC) se aplicó la prueba de raíz unitaria a todas las series, los resultados de la prueba de estacionariedad para las variables independientes se resumen en la tabla 5:

Tabla 5 - Prueba de raíz unitaria variables explicativas

Null Hypothesis: Variable has a unit root					
Exogenous: Constant, Linear Trend					
Lag Length: Automatic - based on SIC, maxlag=14					
Variable	p	Statistic	1%	5%	10%
LOG(TCR)	0.9502	-0,9262	-3,9971	-3,4288	-3,1379
LOG(USDCLP)	0.6876	0,4406	-3,9971	-3,4288	-3,1379
LOG(USDCNY)	0.8947	-1,2603	-3,9971	-3,4288	-3,1379
LOG(USDCOP)	0.7770	-1,6334	-3,9971	-3,4288	-3,1379
LOG(USDPER)	0.7088	-1,1184	-3,9971	-3,4288	-3,1379
LOG(USDEUR)	0.9590	-0,8453	-3,9971	-3,4288	-3,1379
LOG(WTI)	0.2874	-2,5856	-3,9971	-3,4288	-3,1379
LOG(AGIND)	0.7072	-1,7892	-3,9971	-3,4288	-3,1379
LOG(GASTOPUBLICO)	0.9839	-0,4785	-3,9992	-3,4298	-3,1384
LOG(OFERTAMONETARIA)	0.9982	0,2397	-3,9990	-3,4297	-3,1384

Sin embargo, al realizar la misma prueba en sus primeras diferencias se halló que las variables explicativas en sus primeras diferencias si son estacionarias; por tanto, su orden de integración también es $I(1)$ y serán modeladas de esa manera. Ahora, al evaluar la correlación de Pearson de la diferencia del logaritmo de las variables independientes respecto a la primera diferencia del logaritmo de la variable dependiente se obtuvieron resultados mostrados en la tabla 6:

Tabla 6 - Matriz de correlaciones de las diferencias de los logaritmos de las variables

Variable	Correlación
DLOGTCR	-0.385629
DLOGEXCH_CHL	-0.028022
DLOG EXCH_CHN	-0.058784
DLOG EXCH_COL	-0.017934
DLOGEXCH_PER	-0.134809
DLOG EXCH_EUR	0.031226
DLOGWTI	-0.032929
DLOGAGIND	0.059211
DLOGGASTOPUBLICO	-0.140578
DLOGOFERTAMONETARIA	-0.019134

Los gráficos de dispersión de la primera diferencia de los logaritmos de las variables se muestran en el anexo 2. Eliminando el componente tendencial creciente propio de las series de tiempo se puede observar que la correlación disminuye respecto a los resultados de la tabla 2. Previo a un análisis más profundo, las variables explicativas que muestran mayor correlación son el Tipo de Cambio Real y el Gasto Público.

Selección de Variables

Para el cálculo de la primera diferencia del LOG(IPC) o DLOG(IPC), se partirá de una regresión lineal multivariada. De acuerdo a la teoría económica sobre la relación entre las variables se esperarías: una relación inversa entre la inflación y el tipo de cambio y una relación positiva respecto a los precios de las materias primas, salario nominal, gasto público y oferta monetaria. Para determinar su pertinencia en el modelo se ejecutará un modelo de MCO para el período completo (enero 2001 a diciembre 2020). Los resultados del modelo preliminar, con la incorporación de todas las variables explicativas consideradas, se resume en la tabla 7:

Tabla 7 - MCO para la DLOG(IPC) con 11 variables regresoras

Dependent Variable: DLOG(IPC)				
Method: Least Squares				
Date: 06/02/21 Time: 18:49				
Sample (adjusted): 2001M02 2020M12				
Included observations: 239 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
DLOG(TCRB2014)	-0.825635	0.037818	-21.83184	0.0000
DLOG(EXCH_CHL)	-0.046774	0.008891	-5.260526	0.0000
DLOG(EXCH_CHN)	-0.160415	0.028386	-5.651133	0.0000
DLOG(EXCH_COL)	-0.111142	0.009517	-11.67772	0.0000
DLOG(EXCH_PER)	-0.068629	0.018276	-3.755170	0.0002
DLOG(EXCH_EUR)	-0.126264	0.011362	-11.11265	0.0000
DLOG(CL1)	0.004733	0.001861	2.543211	0.0116
DLOG(AGINDEX)	0.026055	0.008649	3.012311	0.0029
DLOG(GASTOPUBLICO)	-0.000361	0.000895	-0.403605	0.6869
LOG(SALARIONOMINAL)	0.000384	3.87E-05	9.925833	0.0000
DLOG(OFERTAMONETARIA)	-0.010004	0.009806	-1.020211	0.3087
R-squared	0.680294	Mean dependent var		0.003113
Adjusted R-squared	0.666272	S.D. dependent var		0.004878
S.E. of regression	0.002818	Akaike info criterion		-8.860751
Sum squared resid	0.001810	Schwarz criterion		-8.700747
Log likelihood	1069.860	Hannan-Quinn criter.		-8.796274
Durbin-Watson stat	1.163710			

Se observa que 2 de las variables no superaron el nivel de significancia del 5%, estas son: el gasto público y la oferta monetaria. La baja interacción de estas variables con la inflación en el caso ecuatoriano se podría explicar por la imposibilidad del estado ecuatoriano de financiarlas mediante la emisión de moneda, por lo que queda supeditado al endeudamiento que, en el mediano y largo plazo, implica también la disminución de circulante por servicio de deuda y por ello su impacto sería equiparado y no tendría una fuerte influencia sobre el IPC.

Al retirar las variables no significativas DLOG(GASTOPUBLICO) y DLOG(OFERTAMONETARIA) se obtuvo el modelo de regresión lineal de la tabla 8:

Tabla 8 - MCO para la DLOG(IPC) con 9 variables regresoras

Dependent Variable: DLOG(IPC)				
Method: Least Squares				
Date: 06/02/21 Time: 18:50				
Sample (adjusted): 2001M02 2020M12				
Included observations: 239 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
DLOG(TCRB2014)	-0.823175	0.037146	-22.16057	0.0000
DLOG(EXCH_CHL)	-0.045885	0.008881	-5.166321	0.0000

DLOG(EXCH_CHN)	-0.162062	0.028415	-5.703465	0.0000
DLOG(EXCH_COL)	-0.112304	0.009439	-11.89849	0.0000
DLOG(EXCH_PER)	-0.066538	0.018226	-3.650724	0.0003
DLOG(EXCH_EUR)	-0.123848	0.011275	-10.98445	0.0000
DLOG(CL1)	0.004613	0.001861	2.478860	0.0139
DLOG(AGINDEX)	0.027656	0.008564	3.229451	0.0014
LOG(SALARIONOMINAL)	0.000364	3.49E-05	10.42523	0.0000
R-squared	0.676443	Mean dependent var	0.003113	
Adjusted R-squared	0.665189	S.D. dependent var	0.004878	
S.E. of regression	0.002822	Akaike info criterion	-8.865512	
Sum squared resid	0.001832	Schwarz criterion	-8.734599	
Log likelihood	1068.429	Hannan-Quinn criter.	-8.812758	
Durbin-Watson stat	1.189226			

En el modelo de la tabla 8 todas las variables son significativas al nivel del 5%, y el coeficiente de determinación ajustado es del 66,52% lo cual indica que las variables independientes seleccionadas en el modelo explican el 66,52% de las variaciones en la inflación. Para evitar el sobreajuste del modelo se retiró la variable DLOG(CL1) correspondiente a la primera diferencia del logaritmo de los precios del WTI dado que es la variable del modelo con la menor significancia, el modelo resultante se muestra en la tabla 9:

Tabla 9 - MCO para la DLOG(IPC) con 8 variables regresoras

Dependent Variable: DLOG(IPC)				
Method: Least Squares				
Date: 06/02/21 Time: 18:52				
Sample (adjusted): 2001M02 2020M12				
Included observations: 239 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
DLOG(TCRB2014)	-0.807398	0.037002	-21.82045	0.0000
DLOG(EXCH_CHL)	-0.046155	0.008979	-5.140232	0.0000
DLOG(EXCH_CHN)	-0.156210	0.028630	-5.456183	0.0000
DLOG(EXCH_COL)	-0.115082	0.009476	-12.14522	0.0000
DLOG(EXCH_PER)	-0.066192	0.018427	-3.592114	0.0004
DLOG(EXCH_EUR)	-0.122532	0.011387	-10.76062	0.0000
DLOG(AGINDEX)	0.031131	0.008542	3.644722	0.0003
LOG(SALARIONOMINAL)	0.000369	3.53E-05	10.47214	0.0000
R-squared	0.667799	Mean dependent var	0.003113	
Adjusted R-squared	0.657732	S.D. dependent var	0.004878	
S.E. of regression	0.002854	Akaike info criterion	-8.847515	
Sum squared resid	0.001881	Schwarz criterion	-8.731148	
Log likelihood	1065.278	Hannan-Quinn criter.	-8.800622	
Durbin-Watson stat	1.100885			

El coeficiente de determinación ajustado es del 65,77%, apenas 0,75% menor al modelo que incluye DLOG(CL1), lo cual implica que la variable no realiza mayor aporte al modelo porque puede estar correlacionada con otra variable ya incluida y con su eliminación se evita el sobre ajuste. El coeficiente de determinación ajustado indica que el modelo pronostica la variación en la inflación con una precisión del 65,77%. Todas las variables regresoras son significativas al nivel del 5%. Al aplicar la prueba de significancia general (Test de Walt) el p del F observado es lo bastante bajo lo cual confirma la capacidad global de los coeficientes de las regresoras para explicar las variaciones del IPC (anexo 3). Además, el valor del estadístico Durbin-Watson (mayor al coeficiente de determinación ajustado), confirma que no se trata de una relación espuria entre las variables.

En conclusión, se ha determinado que las variables con mayor incidencia sobre la inflación en Ecuador para el período 2001-2020 son las de tipo de cambio, los precios internacionales de materias primas y el salario nominal. Se entiende que, Ecuador al ser un país dolarizado, sin capacidad de señoreaje, juega un papel crucial la balanza comercial. De acuerdo al modelo de la tabla 6, la inflación depende principalmente del circulante generado por los tipos de cambio a través de las exportaciones netas; en otras palabras, de existir un superávit comercial, mayor será la cantidad de dinero en circulación presionando al alza sobre los precios; así mismo, la apreciación del dólar perjudicaría las exportaciones haciendo más caros los productos exportación en detrimento de la competitividad del país.

Una complicación que puede presentarse en series de tiempo integradas es la cointegración, lo cual implica que las variables tienen relación estable solo en el largo plazo. Para determinar si existe cointegración se aplica la prueba de raíz unitaria a los residuos de la regresión (anexo 4), esta prueba mostró que los residuos son estacionarios al nivel del 5%, lo cual indica que las variables están cointegradas. Para abordar el problema derivado de la cointegración es necesario incluir el valor rezagado de los residuos y la primera diferencia de las variables no estacionarias para corregir el comportamiento en el corto plazo [29], el modelo que incluye el valor rezagado de los residuos se muestra en la tabla 10:

Tabla 10 - MCO para la DLOG(IPC) con corrección de errores

Dependent Variable: DLOG(IPC) Method: Least Squares Date: 06/27/21 Time: 18:53 Sample (adjusted): 2001M02 2020M11
--

Included observations: 238 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
DLOG(TCR)	-0.776148	0.033280	-23.32197	0.0000
DLOG(EXCH_CHL)	-0.048838	0.008022	-6.087826	0.0000
DLOG(EXCH_CHN)	-0.152052	0.025574	-5.945597	0.0000
DLOG(EXCH_COL)	-0.106793	0.008554	-12.48449	0.0000
DLOG(EXCH_PER)	-0.048940	0.016659	-2.937777	0.0036
DLOG(EXCH_EUR)	-0.117748	0.010185	-11.56092	0.0000
DLOG(AGINDEX)	-0.022148	0.007711	2.872398	0.0045
LOG(SALARIONOMINAL)	0.000382	3.16E-05	12.07307	0.0000
RESIDLOG(1)	0.487029	0.062455	7.798118	0.0000
R-squared	0.737073	Mean dependent var		0.003127
Adjusted R-squared	0.727888	S.D. dependent var		0.004883
S.E. of regression	0.002547	Akaike info criterion		-9.070555
Sum squared resid	0.001486	Schwarz criterion		-8.939251
Log likelihood	1088.396	Hannan-Quinn criter.		-9.017637
Durbin-Watson stat	1.858917			

Para el período de enero-2001 a diciembre-2020 el coeficiente de determinación ajustado de la regresión fue de 72,79%, es decir cerca de un 7% mayor respecto al modelo anterior. Además, el estadístico Durbin-Watson paso de 1,10 a 1,86 lo cual indica que no hay correlación serial en los residuos (d cercano a 2). En este caso y de acuerdo a lo esperado, los tipos de cambio tienen una relación negativa con el IPC y los precios de las materias primas, así como el salario nominal tienen una relación positiva con el IPC. Por otro lado, el parámetro del rezago de los residuos tiene un valor positivo lo cual implica que la $DLOG(IPC)$ tiende a estar por debajo de su nivel de equilibrio de corto plazo. En este contexto y con los resultados de la tabla 10 el modelo de MCO con corrección de errores es el siguiente:

$$\begin{aligned}
 DLOG(IPC) = & -\beta_1 * DLOG(TCR) - \beta_2 * DLOG(EXCH_{CHL}) - \beta_3 * DLOG(EXCH_{CHN}) - \beta_4 \\
 & * DLOG(EXCH_{COL}) - \beta_5 * DLOG(EXCH_{PER}) - \beta_6 * DLOG(EXCH_{EUR}) - \beta_7 \\
 & * DLOG(AGINDEX) + \beta_8 * LOG(SALARIONOMINAL) + \beta_9 * RESID_{t-1}
 \end{aligned} \tag{11}$$

$$\begin{aligned}
 t = & (-23,32) \quad (-6,09) \quad (-5,95) \quad (-12,48) \quad (-2,94) \quad (11,56) \quad (2,87) \quad (12,07) \quad (7,80) \\
 & R^2 = 0,727888 \quad d = 1,86
 \end{aligned}$$

3.2. Modelos de Regresión

MCO

En el apartado anterior se evaluaron las variables que tienen mayor incidencia sobre las variaciones del IPC en el conjunto total de los datos y se encontró que el modelo de regresión de MCO con corrección de errores (MCO CE) es el que logra los mejores resultados. En este apartado se construirá el modelo en el grupo de entrenamiento y se evaluará en el grupo de prueba para también valorar el sobreajuste. Primero se presentará en el modelo de MCO para compararlo con el modelo MCO-CE y determinar si existen variaciones significativas, mismo que se muestra en la tabla 11:

Tabla 11 - Modelo MCO grupo de entrenamiento

Dependent Variable: DLOG(IPC)				
Method: Least Squares				
Date: 07/06/21 Time: 12:40				
Sample (adjusted): 2001M02 2016M12				
Included observations: 191 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
DLOG(TCR)	-0.830688	0.041499	-20.01718	0.0000
DLOG(EXCH_CHL)	-0.047584	0.009574	-4.969913	0.0000
DLOG(EXCH_CHN)	-0.247191	0.043500	-5.682613	0.0000
DLOG(EXCH_COL)	-0.111792	0.010002	-11.17690	0.0000
DLOG(EXCH_PER)	-0.069964	0.020564	-3.402216	0.0008
DLOG(EXCH_EUR)	-0.127814	0.012032	-10.62309	0.0000
DLOG(AGINDEX)	0.030635	0.008965	3.417280	0.0008
LOG(SALARIONOMINA L)	0.000402	4.16E-05	9.665651	0.0000
R-squared	0.683042	Mean dependent var		0.003944
Adjusted R-squared	0.670918	S.D. dependent var		0.004897
S.E. of regression	0.002809	Akaike info criterion		-8.870754
Sum squared resid	0.001444	Schwarz criterion		-8.734533
Log likelihood	855.1570	Hannan-Quinn criter.		-8.815578
Durbin-Watson stat	1.092131			

Para el período de enero-2001 a diciembre-2016(Grupo de entrenamiento) el coeficiente de determinación ajustado de la regresión fue de 67,09% y el RMSE de 2,875303. En el grupo de prueba, correspondiente al 20% restante de los datos (enero-2017 a diciembre-2020) la raíz de la suma del error cuadrático que se obtiene de 3,529178. Finalmente, el ajuste de los datos reales a los pronósticos del modelo MCO se muestran en la siguiente figura:

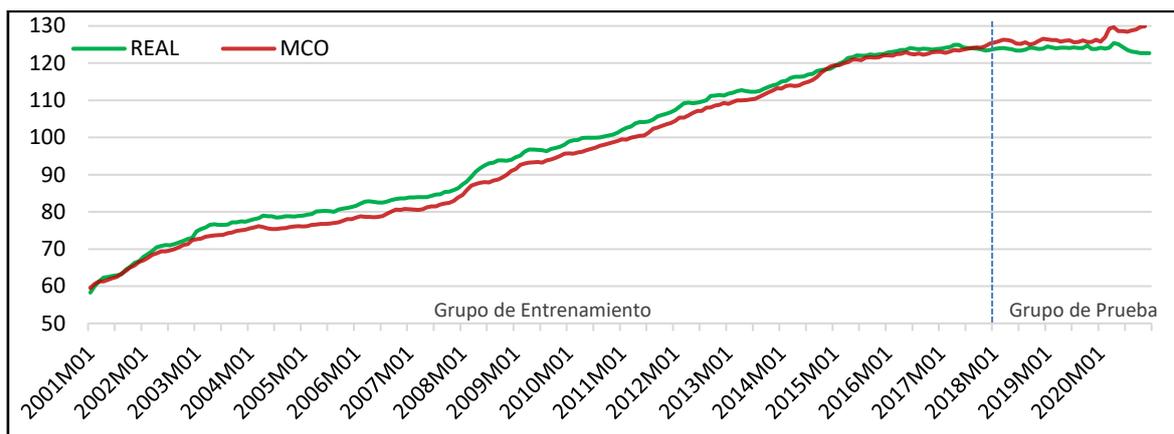


Figura 4 - Pronóstico del IPC modelo MCO

En la figura se observa que, en el grupo de entrenamiento, gran parte de la serie subestima el IPC real hasta que las series se igualan nuevamente en 2015. Por otro lado, en el grupo de prueba los pronósticos del modelo MCO sobrestiman el IPC real especialmente durante 2020. Los RMSE de ambos grupos se resumen en la tabla 12:

Tabla 12 - Resultados RMSE de los pronósticos MCO

Indicador	Modelo	Entrenamiento	Prueba
RMSE	MCO	2,875303	3,529178

El desempeño observado es mejor en el grupo de entrenamiento, pero se debe observar que el año 2020 presenta valores atípicos generados por la grave crisis económica provocada por la pandemia del covid-19 y los consecuentes cierres y cuarentenas alrededor del mundo que frenaron el comercio mundial. Los datos particularmente bajos del IPC en 2020 elevan el RMSE en el grupo de prueba. Este efecto, de un RMSE alto en el grupo de prueba en comparación con el grupo de entrenamiento, se esperaría observar también los modelos de las siguientes secciones de este estudio porque todos incluirán el 2020 en sus observaciones. Si se calculara el RMSE solo para los pronósticos de enero-2017 a diciembre-2019 se obtendría un RMSE de 2,7167, es decir que el desempeño del modelo sería incluso mejor en el grupo de prueba si no se consideraran los datos atípicos del 2020.

En la tabla 7 se determinó que un modelo de MCO con corrección de error mejoraría su capacidad predictiva, por lo cual al modelo de MCO construido en el grupo de

entrenamiento se le incluirá el rezago del error para crear el modelo MCO CE. Al ejecutar el modelo se obtienen los resultados de la tabla 13:

Tabla 13 - Modelo MCO CE grupo de entrenamiento

Dependent Variable: DLOG(IPC) Method: Least Squares Date: 06/29/21 Time: 15:06 Sample (adjusted): 2001M02 2016M12 Included observations: 191 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
DLOG(TCR)	-0.798854	0.037358	-21.38364	0.0000
DLOG(EXCH_CHL)	-0.050593	0.008564	-5.907729	0.0000
DLOG(EXCH_CHN)	-0.218123	0.039087	-5.580476	0.0000
DLOG(EXCH_COL)	-0.106262	0.008971	-11.84516	0.0000
DLOG(EXCH_PER)	-0.061680	0.018409	-3.350461	0.0010
DLOG(EXCH_EUR)	-0.121215	0.010791	-11.23344	0.0000
DLOG(AGINDEX)	0.019807	0.008161	2.426928	0.0162
LOG(SALARIONOMINAL)	0.000399	3.71E-05	10.73968	0.0000
RESIDL(1)	0.479859	0.069749	6.879768	0.0000
R-squared	0.748458	Mean dependent var		0.003944
Adjusted R-squared	0.737401	S.D. dependent var		0.004897
S.E. of regression	0.002510	Akaike info criterion		-9.091443
Sum squared resid	0.001146	Schwarz criterion		-8.938195
Log likelihood	877.2329	Hannan-Quinn criter.		-9.029371
Durbin-Watson stat	1.811346			

Para el período de enero-2001 a diciembre-2016(grupo de entrenamiento) el coeficiente de determinación ajustado de la regresión fue de 73,74% y el RMSE de 2,162637. Se observa una considerable mejoría respecto al modelo sin corrección de errores con una reducción del 25% del RMSE. En el grupo de prueba, correspondiente al 20% restante de los datos (enero-2017 a diciembre-2020) la raíz de la suma del error cuadrático que se obtiene es de 2,045096, este RMSE es 42,05% menor que el registrado en el modelo sin corrección de errores. En la figura 5 se muestra el ajuste de los datos reales a los pronósticos del modelo MCO CE:

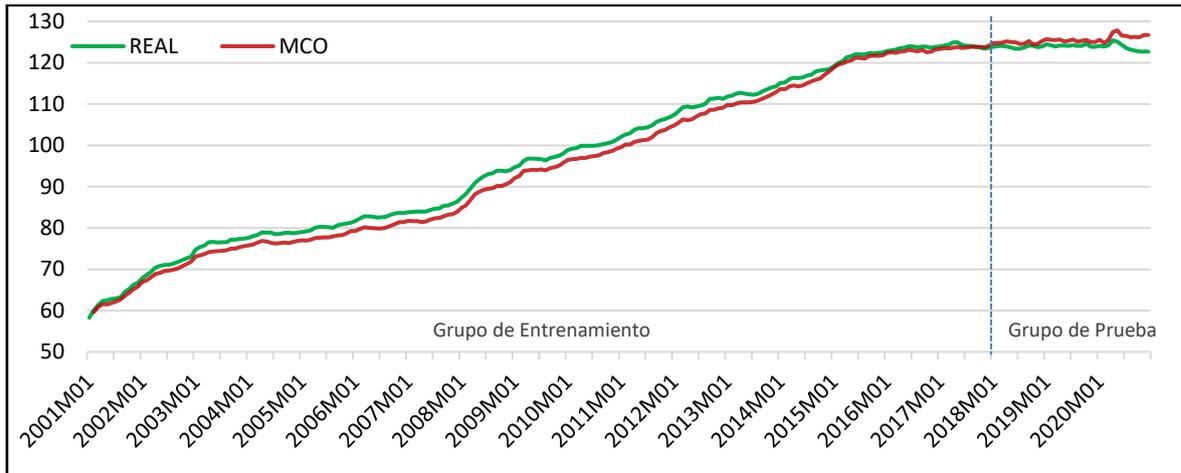


Figura 5 - Pronóstico del IPC modelo MCO CE

En el grupo de entrenamiento los valores pronosticados también tienden a ser consistentemente inferiores dentro del período que va desde el 2003 hasta 2014, pero en menor medida que el modelo de MCO sin corrección de errores. Por otro lado, desde el 2018 el pronóstico tiende a sobrestimarse respecto a los datos reales, hecho que es particularmente notable en el 2020. El RMSE más bajo se dio en el grupo de prueba como se muestra en la tabla 14:

Tabla 14 - Resultados RMSE modelo MCO CE

Indicador	Modelo	Entrenamiento	Prueba
RMSE	MCO CE	2,162637	2,045096

ARMA

Al realizar la prueba de estacionariedad al IPC resultó que es una serie integrada de orden 1, por lo cual no tendría valor realizar un modelo ARMA ya que la serie debe estar integrada, en otras palabras, la construcción de un modelo autorregresivo (AR) y de medias móviles (MA) para el IPC debe realizarse en sus primeras diferencias como parte de un modelo integrado ARIMA. “Si una serie de tiempo es no estacionaria, solo podemos estudiar su comportamiento durante el período en consideración [...] en consecuencia, no es posible generalizar para otros períodos. Así para propósitos de pronóstico, tales series de tiempo tienen poco valor práctico” [29]. Bajo el antecedente expuesto no construirá el modelo ARMA.

ARIMA

La metodología Box-Jenkins, o metodología ARIMA, es aplicada para encontrar un modelo autorregresivo de medias móviles que mejor se ajuste a una serie de tiempo. Esta metodología iterativa se basa en la evaluación de estacionalidad y estacionalidad a través de los correlogramas de las funciones de autocorrelación (FAC) y autocorrelación parcial (FACP) de las series.

La tabla 12 muestra los correlogramas de las funciones FAC y FACP de la serie DLOG(IPC), en estas gráficas se evalúa la presencia de autocorrelación en los rezagos, si las barras se encontraran fuera de los límites verticales sería indicación de correlación serial. Además, la función de autocorrelación permite evaluar comportamientos cíclicos que muestren estacionalidad. En este sentido, los correlogramas nos pueden sugerir qué rezagos incluir en el modelo ARIMA para iniciar con el proceso iterativo. La tabla 15 muestra los correlogramas de la serie DLOG(IPC).

Tabla 15 - Correlograma DLOG(IPC)

Date: 06/02/21 Time: 18:55 Sample: 2001M01 2020M12 Included observations: 239						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.515	0.515	64.104	0.000
		2	0.376	0.152	98.511	0.000
		3	0.228	-0.020	111.16	0.000
		4	0.174	0.034	118.59	0.000
		5	0.190	0.103	127.44	0.000
		6	0.228	0.114	140.29	0.000
		7	0.277	0.123	159.36	0.000
		8	0.180	-0.074	167.44	0.000
		9	0.233	0.125	181.03	0.000
		10	0.240	0.101	195.52	0.000
		11	0.279	0.102	215.14	0.000
		12	0.393	0.224	254.26	0.000
		13	0.217	-0.178	266.28	0.000
		14	0.135	-0.089	270.96	0.000
		15	0.080	0.018	272.59	0.000
		16	0.058	-0.048	273.47	0.000
		17	0.050	-0.046	274.11	0.000
		18	0.048	-0.086	274.72	0.000
		19	0.088	-0.012	276.76	0.000
		20	0.047	0.015	277.35	0.000
		21	0.056	-0.047	278.19	0.000
		22	0.106	0.068	281.18	0.000
		23	0.152	0.079	287.37	0.000
		24	0.195	0.046	297.56	0.000

En el correlograma FACP se aprecian barras fuera de los límites, esto es señal de que no se trata de una serie de tiempo puramente estocástica, en otras palabras, si existe correlación serial en el DLOG(IPC). En el correlograma FACP se identifica autocorrelación AR(1), AR(12) y AR(13). Mientras que en la gráfica de FAC se identifica MA(1), MA(2) y MA(12). Dicho de otra manera, existe un proceso de autocorrelación y de medias móviles con parámetros $p = 3$ y $q = 3$. Una vez identificado un modelo base, se realizaron pruebas con distintos rezagos para evaluar el modelo que mejores resultados obtenga, estas pruebas se resumen en la tabla 16:

Tabla 16 - Modelos ARIMA

Modelo	AR	MA	Adjusted R-squared	Schwarz criterion
1	AR(1) AR(12) AR(13)	MA(1) MA(2) MA(12)	0,489244	-8,765972
2	AR(1) AR(6) AR(12) AR(13)	MA(1) MA(12)	0,491387	-8,770176
3	AR(1) AR(6) AR(12) AR(13)	MA(1) MA(2) MA(11) MA(12)	0,484132	-8,736752
4	AR(1) AR(12) AR(13)	MA(1) MA(12) MA(13)	0,510448	-8,808373
5	AR(1) AR(12) AR(13)	MA(1) MA(2) MA(12) MA(13)	0,510818	-8,789700
6	AR(1) AR(12)	MA(1) MA(12)	0,384147	-8,604616

Para seleccionar el modelo ARIMA se utilizó el criterio de Schwarz, el cual permite seleccionar el modelo más adecuado entre un grupo de modelos similares imponiendo una penalización al sobreajuste. De acuerdo al criterio de Schwarz, el mejor modelo ARIMA (menor valor) es el 4(Ver tabla 16), el cual incluye los rezagos AR(1) AR(12) AR(13) y MA(1) MA (12) MA (13). Al ejecutar este modelo, ARIMA(3,3), se obtuvieron los resultados de la tabla 17:

Tabla 17 - Modelo ARIMA(3,3)

Dependent Variable: DLOG(IPC)				
Method: Least Squares				
Date: 06/29/21 Time: 16:25				
Sample (adjusted): 2002M03 2016M12				
Included observations: 178 after adjustments				
Failure to improve likelihood (non-zero gradients) after 13 iterations				
Coefficient covariance computed using outer product of gradients				
MA Backcast: 2001M02 2002M02				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.798269	0.075815	10.52923	0.0000
AR(12)	0.920786	0.018593	49.52385	0.0000
AR(13)	-0.733798	0.071834	-10.21520	0.0000
MA(1)	-0.449785	0.114804	-3.917863	0.0001

MA(12)	-0.946692	0.015854	-59.71251	0.0000
MA(13)	0.409051	0.111711	3.661691	0.0003
R-squared	0.540768	Mean dependent var	0.003313	
Adjusted R-squared	0.527418	S.D. dependent var	0.003880	
S.E. of regression	0.002667	Akaike info criterion	-8.982544	
Sum squared resid	0.001223	Schwarz criterion	-8.808373	
Log likelihood	805.4464	Hannan-Quinn criter.	-8.939050	
Durbin-Watson stat	2.039085			

De acuerdo con estos resultados, el modelo ARIMA(3,3) tiene una capacidad de predecir el 52,74% de las variaciones en el DLOG(IPC), además se puede notar que los parámetros con mayor valor son AR(12) y MA(12) lo cual evidencia el fuerte componente estacional interanual en la serie.

El modelo ARIMA, en el grupo de prueba, tiene un RMSE de 3,635990 y en el grupo de entrenamiento un RMSE de 4,618656. La predicción mediante el modelo ARIMA y los datos reales del IPC se muestran en la figura 6:

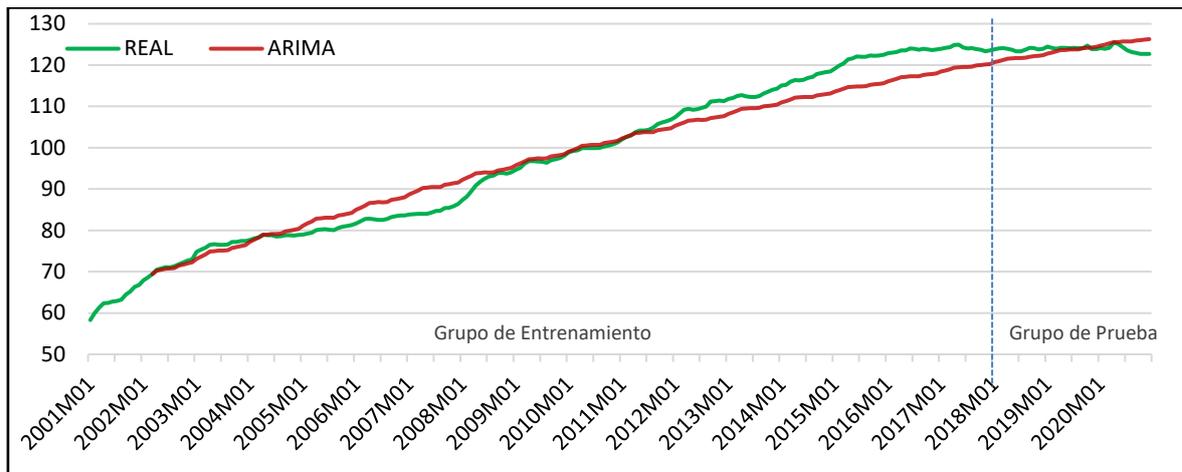


Figura 6 – Pronósticos del IPC modelo ARIMA

En la figura 6 se puede observar que la línea de regresión del modelo ARIMA no responde muy bien a los cambios en la tendencia del IPC real. De 2005 a 2008 el pronóstico ARIMA sobrestima el IPC real, mientras que desde el segundo semestre del 2012 hasta finales de 2019 el pronóstico subestima el IPC real. Este efecto se produce porque se incorporaron los rezagos 12 y 13 tanto en AR como en MA, lo cual ocasiona un retraso de 1 año en incorporar el peso de estos valores al pronóstico. La raíz del error cuadrático medio es menor en el grupo de entrenamiento como se resume en la tabla 18:

Tabla 19 - Resultados RMSE modelo ARIMA

Indicador	Modelo	Entrenamiento	Prueba
RMSE	ARIMA	3,635990	4,618656

SARIMA

La serie del IPC es elaborada por el INEC con periodicidad mensual, lo cual es antecedente de la posible presencia correlación serial interanual. El correlograma de la tabla-12 ya mostró un patrón estacional cada 12 meses, en el correlograma de la FAC se observó un ciclo cada 12 observaciones. Por ello, al proceso MA del modelo se le agregará una corrección estacional en forma de un término de promedio móvil estacional multiplicativo en el rezago 12.

Al modelo ARIMA elaborado en el apartado anterior se le agregará la corrección estacional interanual, de forma que se obtiene un modelo SARIMA(3,3,1)₁₂ el cual se muestra en la tabla 19:

Tabla 20 - Modelo SARIMA(3,3,12)

Dependent Variable: DLOG(B100_2010)				
Method: ARMA Conditional Least Squares (Gauss-Newton / Marquardt steps)				
Date: 06/29/21 Time: 16:29				
Sample (adjusted): 2002M03 2016M12				
Included observations: 178 after adjustments				
Failure to improve likelihood (non-zero gradients) after 12 iterations				
Coefficient covariance computed using outer product of gradients				
MA Backcast: 2000M02 2002M02				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.788709	0.077807	10.13672	0.0000
AR(12)	0.887014	0.021978	40.35889	0.0000
AR(13)	-0.690217	0.073563	-9.382677	0.0000
MA(1)	-0.408075	0.116815	-3.493352	0.0006
MA(12)	0.049122	0.080052	0.613625	0.5403
MA(13)	-0.084715	0.074706	-1.133976	0.2584
SMA(12)	-0.910065	0.029868	-30.46949	0.0000
R-squared	0.526322	Mean dependent var		0.003313
Adjusted R-squared	0.509702	S.D. dependent var		0.003880
S.E. of regression	0.002717	Akaike info criterion		-8.940337
Sum squared resid	0.001262	Schwarz criterion		-8.815211
Log likelihood	802.6900	Hannan-Quinn criter.		-8.889595
Durbin-Watson stat	2.077440			

Los resultados de la regresión muestran que los términos de medias móviles de los rezagos 12 y 13 se vuelven no significativos al nivel del 5% cuando se introduce la corrección estacional, por lo cual son retirados del modelo para evitar el sobreajuste (Ver tabla 20):

Tabla 21 - Modelo SARIMA(3,1,1)₁₂

Dependent Variable: DLOG(B100_2010)				
Method: Least Squares				
Date: 06/29/21 Time: 16:30				
Sample (adjusted): 2002M03 2016M12				
Included observations: 178 after adjustments				
Failure to improve likelihood (non-zero gradients) after 12 iterations				
Coefficient covariance computed using outer product of gradients				
MA Backcast: 2001M02 2002M02				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.793039	0.078151	10.14750	0.0000
AR(12)	0.923510	0.018548	49.79125	0.0000
AR(13)	-0.730778	0.074040	-9.869976	0.0000
MA(1)	-0.438587	0.117224	-3.741455	0.0002
SMA(12)	-0.948222	0.014942	-63.45892	0.0000
R-squared	0.538094	Mean dependent var		0.003313
Adjusted R-squared	0.527415	S.D. dependent var		0.003880
S.E. of regression	0.002667	Akaike info criterion		-8.987976
Sum squared resid	0.001231	Schwarz criterion		-8.898600
Log likelihood	804.9298	Hannan-Quinn criter.		-8.951731
Durbin-Watson stat	2.039838			

Este modelo es un modelo SARIMA(3,1,1)₁₂, es decir es un proceso autorregresivo de orden 3, de medias móviles de orden 1 y corrección estacional con periodicidad mensual. Respecto al modelo anterior (tabla N°16), al retirar los procesos de MA(12) y MA(13) el coeficiente de determinación ajustado se mantiene en 52,74%, pero su RMSE se reduce a 3,111389 en el grupo de entrenamiento. Sin embargo, en el grupo de prueba, correspondiente al 20% de los datos (enero-2017 a diciembre-2020) el RMSE se incrementa a 5,400959. La figura 7 muestra los datos reales del IPC y los valores pronosticados por el modelo SARIMA:

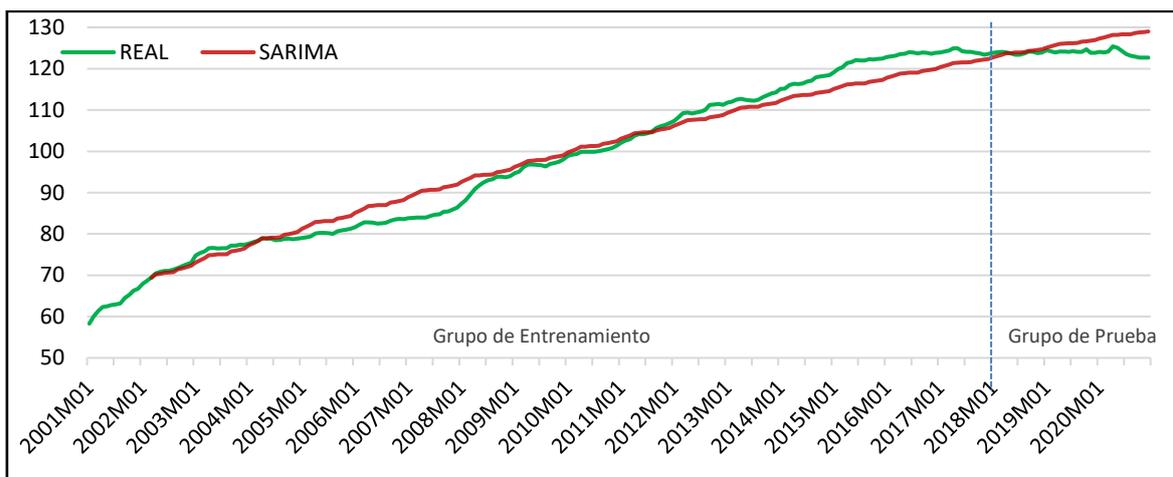


Figura 7 - Pronósticos del IPC modelo SARIMA

La línea de regresión del modelo SARIMA se ajusta mejor los cambios en la tendencia del IPC real que el modelo ARIMA, no obstante, la diferencia entre el valor pronosticado y el real es mayor en 2020. De igual manera que en el modelo ARIMA, existe sobrestimación de 2005 a 2008 y subestimación desde el segundo semestre del 2012 hasta finales de 2018. La raíz del error cuadrático medio 14,4% menor que en el modelo ARIMA en el grupo de entrenamiento, pero es 16,93% mayor en el grupo de prueba como se puede observar en la tabla 21:

Tabla 22 - Resultados RMSE modelo SARIMA

Indicador	Modelo	Entrenamiento	Prueba
RMSE	SARIMA	3,111389	5,400959

VAR

Un modelo de vectores autorregresivos (VAR) es un sistema de ecuaciones simultáneas donde, las variables endógenas son variables cuyo valor está determinado dentro del mismo sistema de ecuaciones y es incluida como sus valores rezagados. Mientras que las variables exógenas, las que no dependen de las otras variables del sistema y son incluidas sin rezagos. En el caso del IPC de Ecuador, las variables de tipo de cambio y del precio internacional de materias primas serán consideradas como variables exógenas puesto que están supeditadas a factores externos al sistema, con lo cual se creará un sistema de dos ecuaciones: para el DLOG(IPC) y el Salario Nominal.

La decisión de la longitud máxima de rezagos utilizará el criterio de Schwarz, o criterio de información Bayesiano, para elegir la ecuación de DLOG(IPC) con el número de rezagos que reduzca el sobreajuste y maximice el coeficiente de determinación ajustado como se observa en la tabla 22:

Tabla 23 - Prueba de rezagos

Rezagos	Schwarz criterion
0	-7.729629
1	-13.37576
2	-13.31141
3	-13.24894
4	-13.33349
5	-13.35002
6	-13.37279
7	-13.33777
8	-13.31785
9	-13.42511*
10	-13.37713
11	-13.36553
12	-13.40770

En el sistema de dos ecuaciones, el menor coeficiente de Schwarz se alcanza a los 9 rezagos por lo cual se modelará de esa manera. El sistema de ecuaciones estaría formado de la siguiente manera (Ver ecuación 12 y 13):

DLOG(IPC)

$$\begin{aligned}
 &= \sum_{i=1}^9 [\beta_i * DLOG(IPC)_{t-i}] + \sum_{j=1}^9 [\beta_j * LOG(SALARIONOMINAL)_{t-i}] \\
 &+ \beta_0 - \beta_1 * DLOG(TCR) - \beta_2 * DLOG(EXCH_{CHL}) - \beta_3 * DLOG(EXCH_{CHN}) \\
 &- \beta_4 * DLOG(EXCH_{COL}) - \beta_5 * DLOG(EXCH_{PER}) - \beta_6 * DLOG(EXCH_{EUR}) \\
 &+ \beta_7 * DLOG(AGINDEX)
 \end{aligned} \tag{12}$$

LOG(SALARIONOMINAL)

$$\begin{aligned}
 &= \sum_{i=1}^9 [\theta_i * DLOG(IPC)_{t-i}] + \sum_{j=1}^9 [\theta_j * LOG(SALARIONOMINAL)_{t-i}] + \theta_0 \\
 &- \theta_1 * DLOG(TCR) - \theta_2 * DLOG(EXCH_{CHL}) - \theta_3 * DLOG(EXCH_{CHN}) - \theta_4 \\
 &* DLOG(EXCH_{COL}) - \theta_5 * DLOG(EXCH_{PER}) - \theta_6 * DLOG(EXCH_{EUR}) + \theta_7 \\
 &* DLOG(AGINDEX)
 \end{aligned} \tag{13}$$

Lo valores de los parámetros se detallan en el anexo 5. La ecuación de predicción de $DLOG(IPC)$ muestra un coeficiente de determinación ajustado de 75,21% y su RMSE es de 2,350002, mientras en el grupo de prueba, correspondiente al 20% de los datos (enero-2017 a diciembre-2020) el RMSE de 2,971511. Al transformar los datos a valores del $DLOG(IPC)$ se obtiene el siguiente gráfico:

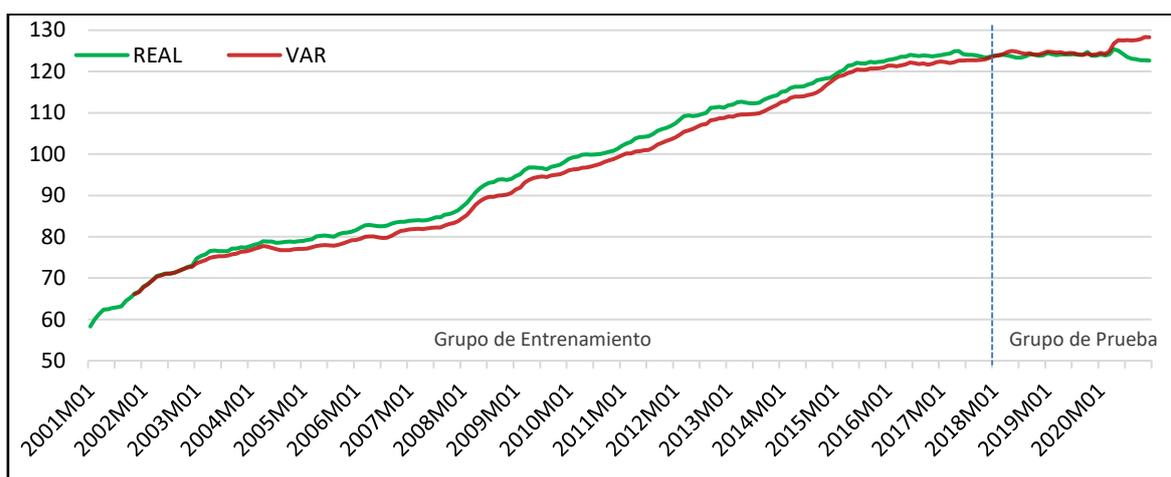


Figura 8 - Pronósticos IPC del modelo VAR

La línea de regresión del modelo VAR para el IPC está consistentemente subestimando el IPC en el grupo de entrenamiento. En el grupo de prueba la predicción es bastante cercana al IPC real, sin embargo, en el 2020 existe una fuerte divergencia entre los valores reales y pronosticados, lo cual coincide con la crisis económica que se sufrió el 2020 ocasionada por la pandemia global del covid-19. La raíz del error cuadrático medio es menor en el grupo de entrenamiento como se resume en la tabla 23:

Tabla 24 - Resultados RMSE del modelo VAR

Indicador	Modelo	Entrenamiento	Prueba
RMSE	VAR	2,350002	2,971511

3.3. Modelo de RNA

El modelo de red neuronal artificial (RNA) se aplicará para la predicción de la inflación y comprobar si interpreta mejor los comportamientos no lineales subyacentes en las series de tiempo económicas [2], en este sentido una RNA puede incorporar ambos comportamientos: lineales y no lineales. Al igual que en los modelos de regresión de los

apartados anteriores, los datos han sido transformados a logaritmos para reducir su volatilidad y se utilizarán sus primeras diferencias para evitar que se produzcan relaciones espurias.

El modelo de RNA perceptrón multicapa de retropropagación es utilizado con frecuencia en estudios estadísticos y económicos para pronosticar el IPC ([2] [34] [39]) por lo cual será el usado en el presente estudio para predecir el IPC ecuatoriano. La topología del modelo de RNA será multivariante que partirá de una capa de entrada para 8 variables, mismas utilizadas en el modelo de MCO, y una neurona para la capa de salida que se corresponde con el pronóstico del $DLOG(IPC)$. Para las capas ocultas de la red se probarán 1 y 2 capas; sin embargo, cabe mencionar que a priori se esperan mejores resultados en los modelos con 2 capas ocultas, dado que “para series económicas usualmente se utiliza un modelo de 2 capas ocultas” [40], o en lo citado por Mora [34] “diferentes estudios [...] señalan que una o máximo dos capas ocultas son ideales para las RNA de series de tiempo macroeconómicas”.

El número de nodos en cada capa se evaluará mediante ensayo y error con un número de 2^k variando k de 1 a 5 evaluando cuál modelo arroja el menor RMSE. De acuerdo con Mora [34], para el pronóstico del IPC “las funciones de activación lineales dieron los mejores resultados que aquellas redes que utilizaron funciones de activación sigmoidea o tangencial hiperbólica”, por tal motivo, en el presente trabajo la topografía de red será construida con la unidad lineal rectificadora (ReLU por sus siglas en inglés). Una vez determinada la topografía se evaluarán diferentes funciones de activación para establecer cuál ofrece mejores resultados, aunque como se mencionó, se esperaba que los modelos con función de activación lineal muestren un mejor desempeño. Así mismo, se evaluó de forma preliminar el RMSE y el número de iteraciones (epochs) y se obtuvo la figura 9:

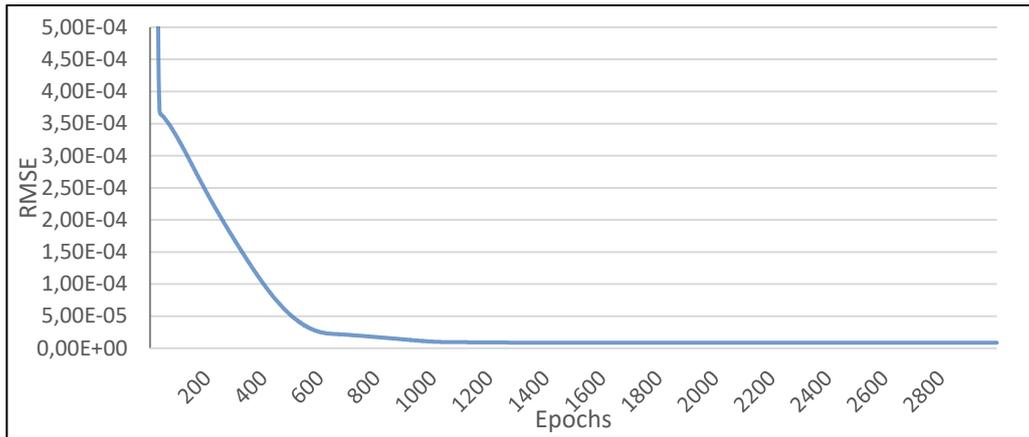


Figura 9 - RMSE vs Epochs

Se puede observar que el error tiende a estabilizarse a partir de las 1200 iteraciones por lo cual los modelos se entrenaron con 2000 iteraciones. Bajo los parámetros descritos en los párrafos precedentes se realizaron las pruebas con las 2 estructuras (1 y 2 capas ocultas) y diferentes números de neuronas obteniendo los siguientes resultados de la tabla 24:

Tabla 25 - Pruebas de modelos de RNA

Modelo	Capas Ocultas	1ra Capa	2da Capa	RSME Entrenamiento	RSME Prueba
1	1	2 neuronas	NA	10,7269572	19,2490734
2	1	4 neuronas	NA	6,01250106	10,7477486
3	1	8 neuronas	NA	5,08438308	6,51843248
4	1	16 neuronas	NA	7,88305552	13,8176485
5	1	32 neuronas	NA	8,05370712	13,3036762
6	2	2 neuronas	2 neuronas	5,50993013	9,77507645
7	2	2 neuronas	4 neuronas	6,36670703	10,6604896
8	2	2 neuronas	8 neuronas	8,11511895	14,5855796
9	2	2 neuronas	16 neuronas	12,2759013	17,476969
10	2	2 neuronas	32 neuronas	5,14390926	8,76471889
11	2	4 neuronas	2 neuronas	10,7240693	19,1125848
12	2	4 neuronas	4 neuronas	4,6028784	7,76975402
13	2	4 neuronas	8 neuronas	8,61854161	14,0134677
14	2	4 neuronas	16 neuronas	5,45959616	9,69929157
15	2	4 neuronas	32 neuronas	4,16900707	7,99831117
16	2	8 neuronas	2 neuronas	5,27398896	9,32967391
17	2	8 neuronas	4 neuronas	6,27967927	10,3188631
18	2	8 neuronas	8 neuronas	4,01388543	6,56103218
19	2	8 neuronas	16 neuronas	2,39371008	3,04645547
20	2	8 neuronas	32 neuronas	9,79952456	17,5833121

21	2	16 neuronas	2 neuronas	5,40159484	8,34062616
22	2	16 neuronas	4 neuronas	12,6190635	22,5761047
23	2	16 neuronas	8 neuronas	9,73099355	17,5464236
24	2	16 neuronas	16 neuronas	9,83981315	17,4740996
25	2	16 neuronas	32 neuronas	15,4293079	28,0914973
26	2	32 neuronas	2 neuronas	14,7396204	26,6327345
27	2	32 neuronas	4 neuronas	12,3509513	22,0876465
28	2	32 neuronas	8 neuronas	4,67889239	7,46395965
29	2	32 neuronas	16 neuronas	10,2876098	18,4167325
30	2	32 neuronas	32 neuronas	8,14784277	13,1569571

Los RMSE en entrenamiento y en prueba fueron calculados después de la transformación de la serie pronosticada por el modelo de DLOG(IPC) a IPC. En la tabla anterior se puede observar que el modelo con mejores resultados es el 19, es decir el modelo de 2 capas ocultas, con 8 neuronas en la primera capa oculta y 16 en la segunda capa oculta y funciones de activación ReLU. Una vez se ha encontrado la mejor topografía se realizaron pruebas con distintas funciones de activación en las capas ocultas; sin embargo, como se mencionó antes, se espera que las funciones de activación ReLU tengan un mejor desempeño como se observa en la tabla 25:

Tabla 26 - Pruebas de funciones de activación

N°	Función de Activación		RMSE	
	1ra Capa Oculta	2da Capa Oculta	Entrenamiento	Prueba
1	Relu	Relu	2,393710	3,046455
2	Relu	Sigmoid	6,472976	8,819172
4	Relu	Tangente Hiperbólica	7,098465	11,312641
5	Sigmoid	Relu	5,223579	3,943753
6	Sigmoid	Sigmoid	4,587720	3,132680
8	Sigmoid	Tangente Hiperbólica	6,771036	10,518092
9	Tangente Hiperbólica	Relu	6,583764	4,255013
10	Tangente Hiperbólica	Sigmoid	4,730354	3,584044
11	Tangente Hiperbólica	Tangente Hiperbólica	6,295398	6,477504

Los modelos de RNA, al igual que en los modelos de regresión, toman el 80% de los datos correspondientes a los datos de enero-2001 a diciembre-2016 como el grupo de entrenamiento, y el 20% restante, correspondiente al período de enero-2017 a diciembre-2020, como el grupo de prueba. En la tabla anterior se observa que el menor RMSE se da en el modelo 1 (Ver tabla 25), el cual se corresponde con dos funciones de activación ReLU para las 2 capas ocultas. Este modelo, en su primera capa cuenta con 8 neuronas

y la segunda con 16 neuronas. Al realizar la transformación de los pronósticos del DLOG(IPC) a IPC se obtiene la figura 10.

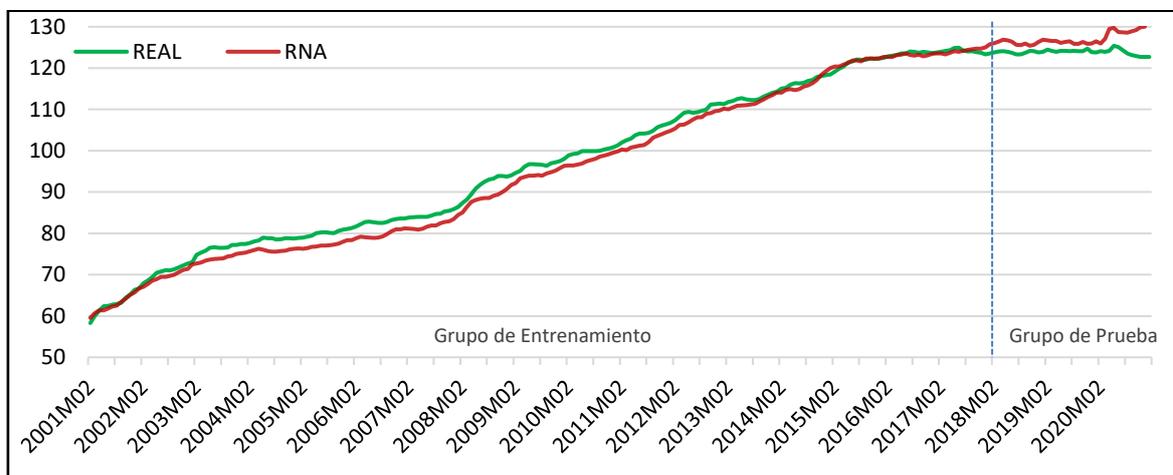


Figura 10 - Pronósticos IPC modelo RNA

En la figura se observa que la línea del pronóstico del modelo RNA para el IPC subestima el IPC real en el grupo de entrenamiento desde el 2003 hasta finales del 2014. Pero, por otro lado, en el grupo de prueba, la predicción del IPC está por encima del IPC real en promedio 1,49 puntos hasta el 2019. En el 2020 existe una fuerte divergencia entre los valores reales y pronosticados, y la diferencia promedio asciende a 4,84 puntos. La raíz del error cuadrático medio es menor en el grupo de entrenamiento con un valor de 2,393710, mientras en el grupo de prueba es 3,046455, tal como se resumen en la tabla 26:

Tabla 27 - Resultados RMSE del modelo RNA

Indicador	Modelo	Entrenamiento	Prueba
RMSE	RNA	2,393710	3,046455

El diagrama del modelo de RNA se lo puede observar en el anexo 7, los pesos de las conexiones entre neuronas se recogen en el anexo 8, mientras que los sesgos de las neuronas en el anexo 9. Un hecho particular que se presentó en las pruebas de los modelos de RNA es que, los modelos podían tener bajos valores de RMSE, la serie DLOG(IPC). Sin embargo, al realizar la transformación a IPC algunos modelos disparaban su RMSE. Esto se debió a que en DLOG(IPC) se podían tener un bajo valor de RMSE en el conjunto de observaciones, pero ocultaba una medición con alto error,

este último en la transformación a IPC se propagaba y mantenía a lo largo de la serie pronosticada lo cual daba como resultado un RMSE alto en la serie transformada.

3.4. Comparación de los modelos

Los modelos para el pronóstico del IPC han sido realizados utilizando técnicas de Machine Learning como los análisis de regresión y las RNA. Todos los modelos pronosticaron la primera diferencia del logaritmo del IPC, esta transformación se dio para que las series sean estacionarias y la correlación entre la variable regresora y las variables regresoras no sea espuria lo cual ocurre con frecuencia en series de tiempo que por inercia tienden a incrementar su valor en el tiempo.

Así mismo, inicialmente se avaluó la influencia de 11 variables sobre el IPC tomando como referencia varios trabajos anteriores, y de su evaluación se determinó que para el período de estudio las variables de mayor relevancia para las variaciones del IPC fueron las de tipo de cambio, precios de materias primas y el salario nominal. Los resultados de los distintos modelos son comparados mediante el indicador RMSE se resumen en la tabla 27:

Tabla 28 - Resumen RMSE por modelo

Modelo	Entrenamiento	Prueba
MCO	2,875303	3,529178
MCO CE	2,162637	2,045096
ARIMA	3,635990	4,618656
SARIMA	3,111389	5,400959
VAR	2,350002	2,971511
RNA	2,393710	3,046455

Los mejores resultados tanto en el grupo de entrenamiento como en el de prueba se dieron en el modelo de mínimos cuadrados ordinarios con corrección de error a corto plazo. Los modelos que peor desempeño mostraron son los autorregresivos de medias móviles especialmente en el grupo de prueba. El modelo de MCO mejoró mucho su capacidad predictiva al incorporar el término de corrección de error a corto plazo y además es el único modelo que tiene un RMSE mejor en el grupo de prueba.

Los modelos autorregresivos y de medias móviles, ARIMA y SARIMA, tuvieron el RMSE más alto de los modelos, una de sus desventajas radica en la pérdida de información al inicio del período ya que se necesitó de 12 grados de libertad para la corrección estacional. Ambos modelos respondieron muy lentamente a cambios en la tendencia del IPC real, lo cual se debe a que la mayoría de rezagos eran de orden 12 o superior, es decir, la recta de regresión de los modelos autorregresivos y medias móviles registraban el mayor peso de los cambios del IPC después de 12 observaciones.

Los modelos VAR y RNA obtuvieron resultados similares en términos del RMSE, en ambos casos su desventaja es que son modelos ateóricos y sus parámetros son numerosos y no tienen interpretación. En el caso del modelo VAR existe poca teoría que explique la relación entre las variaciones del IPC y de los Salario Nominales (Sistema de 2 ecuaciones) que guíen la especificación del modelo. En el caso de las RNA el número de neuronas, los pesos de las conexiones y los sesgos no tienen interpretación en la teoría económica ya que han sido generados de un proceso de ensayo y error de la descomposición de las variables en varias capas.

Todos los modelos mostraron dificultades para el pronóstico del IPC el año 2020 dada la coyuntura económica especial que va más allá de las variables explicativas consideradas. El año 2020 presentó un contexto económico complicado debido a la crisis sanitaria y la reducción abrupta de la demanda de petróleo a nivel mundial como resultado de los confinamientos y cuarentenas impuestas por los gobiernos para frenar los contagios de covid-19. Indiscutiblemente, la pandemia marca un punto de inflexión en la trayectoria del IPC porque inicia un período de bajos precios del petróleo y retroceso de la demanda global de materias primas. Antes de la crisis, en 2019, el IPC mantuvo un promedio de 124,14 puntos mientras el WTI cotizó en promedio a USD 57,04/barril, pero durante el 2020, el IPC mantuvo un promedio de 123,72 puntos y el WTI se cotizó en USD 39,34/barril en promedio, con una tendencia a la baja el primer semestre del año presionada por la perspectiva negativa sobre la demanda de crudo debido al rebrote de casos de Covid-19 en EE.UU. y Europa. El segundo semestre, los precios del crudo se recuperaron y se reactivó el comercio mundial de materias primas con la aprobación de las primeras vacunas y expectativas de que iniciaran los procesos de vacunación antes de terminar el año. De acuerdo a estimaciones de la Organización Mundial de Comercio (OMC) y el Fondo Monetario Internacional (FMI) en 2020 el comercio mundial se redujo 5,3% [41]. La disminución de precios internacionales de materias primas en 2020 así

como el desplome de la demanda interna y externa afectó los precios productos relacionados con la agricultura, silvicultura y pesca [42], lo cual pesó en el resultado final, cerrando 2020 con una inflación acumulada de -0,93% [43].

CAPÍTULO IV: DISCUSIÓN

Se han realizado varios estudios sobre los determinantes de la inflación en Ecuador desde su dolarización, que al ser un país principalmente exportador de materia prima es vulnerable a las variaciones de los precios internacionales de *commodities*. De acuerdo con el trabajo de Gachet *et al.*, los niveles de inflación en Ecuador en 2008 “son principalmente causados por los precios internacionales de commodities, tipos de cambio y política pública” [15], lo cual ha sido confirmado por el modelo de MCO donde las variables de mayor significancia para las variaciones del $DLOG(IPC)$, al nivel del 5%, fueron los tipos de cambio, precios de materia primas y salario nominal. Los resultados del modelo MCO también confirman el trabajo realizado en la Universidad del Azuay, el cual determinó que, para el período enero-2000 a abril-2018, los principales determinantes de la inflación en Ecuador fueron el tipo de cambio real y los salarios [16].

Similar a los resultados del modelo VAR de Morán [32], el modelo VAR incluyó 9 rezagos consiguiendo que las perturbaciones sean heterosedásticas y el modelo sea estable (anexo 6). En su trabajo, las determinantes de la inflación fueron las de tipo de cambio y gasto público seguido por salario nominal y base monetaria. No obstante, las pruebas realizadas en el modelo MCO mostró que el gasto público y la oferta monetaria no eran determinantes para el IPC en el período 2001-2020. Además, a diferencia de los resultados de Mora [34] período 2004-2017, los modelos ARIMA y SARIMA no mostraron los mejores resultados, en la tabla 28 se resume la comparación de los resultados:

Tabla 29 - Comparación con resultados de Mora [34]

Autor	Período	Modelo	RMSE Prueba
Mora [34]	2000-2017	RNA	1,805315
Mora [34]	2000-2017	SARIMA	0.952456
Mora [34]	2000-2017	VAR	3.000885
Villarreal	2001-2020	RNA	3,046455
Villarreal	2001-2020	SARIMA	5,400959
Villarreal	2001-2020	VAR	2,971511

En discrepancia con los resultados de Mora [34] período 2004-2017, los modelos ARIMA y SARIMA no tuvieron buenos resultados, esta diferencia se debe a que el período de análisis considerado en ambos trabajos es distinto. Desde 2016 el IPC ecuatoriano (base 2014) comienza a estabilizarse alrededor de los 123 puntos y esto provoca que la inercia de la serie pronosticada diverja notoriamente en el grupo de entrenamiento. Al mismo tiempo, los modelos autorregresivos, al incorporar rezagos interanuales, responden con mayor lentitud a estos cambios en el IPC y por ello sus altos valores de RMSE en períodos en que el IPC tiende a estabilizarse.

Por otro lado, similar al trabajo realizado por Mora [34], el modelo de RNA para pronosticar no es mejor que el modelo VAR, no obstante, señala que el modelo de RNA es mejor pronosticando el IPC en el mediano plazo.

CAPÍTULO V: CONCLUSIONES Y RECOMENDACIONES

5.1 Conclusiones

Al evaluar varios conjuntos de variables se determinó que los principales determinantes de la inflación en Ecuador para el período 2001-2020, fueron las de tipo de cambio: Tipo de cambio real y tipos de cambio bilaterales: USDCLP, USDCNY, USDCOP, USDPEN y USDEUR, los cuales se corresponden a los principales socios económicos del país que, sin considerar Estados Unidos, fueron el destino del 60% de las exportaciones en 2019(anexo 10). Además, se debe considerar que la variable tipo de cambio efectivo real, calculada por el BCE, incorpora de manera simultánea el comportamiento de los tipos de cambio bilaterales con los 29 principales socios comerciales de Ecuador, por cual se constituyó en una variable muy potente para la construcción de los modelos.

Para captar el comportamiento de los precios de las materias primas se utilizó el índice de precios de productos agrícolas del Banco Mundial, el cual actuó muy bien como representante de los precios internacionales de materias primas, y se mantuvo como variable significativa al nivel del 1% en todas las pruebas. Por otro lado, el WTI a pesar de ser significativo al nivel del 5% no lo fue al nivel del 1%, por tal motivo y para evitar el sobreajuste fue retirada del modelo. La interacción marginal del WTI con el coeficiente de

determinación ajustado se debe a que las variaciones en su precio están ya capturadas por los tipos de cambio en los pares de divisas con el USD. Esto se debe a que, el WTI al negociarse en dólares se ve afectado por la revaluación y depreciación del dólar, esto es, un dólar más fuerte encarece el WTI a compradores en moneda extranjera y viceversa.

El modelo de mínimos cuadrados ordinarios con corrección al corto plazo mostró los mejores resultados, seguido del modelo VAR y el modelo de RNA. Si se comparan los coeficientes de determinación ajustados del modelo MCO CE (72,78%) y del modelo VAR (75,21%), el modelo VAR es mejor en la predicción del DLOG(IPC) pero reduce su capacidad predictiva al realizar la transformación a IPC. Así mismo, si calculamos el coeficiente de determinación ajustado para los resultados del modelo de RNA mediante el procedimiento del anexo 11 se obtiene un $R_a^2 = 80,06\%$, es decir, el modelo de RNA es mejor prediciendo las variaciones del DLOG(IPC) que los modelos MCO CE y VAR. Sin embargo, el interés del presente estudio es pronosticar el IPC y en ese sentido el mejor modelo resultó ser el de MCO CE, porque, en términos del RMSE, su error es 9,65% menor en el grupo de entrenamiento y 32,87% en el grupo de prueba que el modelo de RNA.

Todos los modelos estudiados mostraron dificultad en ajustarse a los datos del 2020 que por su coyuntura económica especial en el contexto de la crisis económica derivada de la pandemia del covid-19 merecen un trato especial y quizá la incorporación de variables dicótomas que ayuden a corregir la gran divergencia entre los valores pronosticados para este escenario particular.

5.2 Recomendaciones

Parte de un ejercicio adicional podría consistir en normalizar las series para poder comparar los valores de los parámetros y determinar cuáles variables tienen mayor influencia sobre las variaciones del IPC, ya que las variables al encontrarse en distintos rangos no pueden compararse directamente.

Los modelos ARIMA y SARIMA deberían modelarse por períodos en los cuales el IPC se mantiene más estable, estos son: 2001-2003, 2004-2016 y 2016-2019. Esto permitiría que sus predicciones sean más ajustadas y se obtengan mejores resultados.

REFERENCIAS BIBLIOGRÁFICAS

- [1] P. Zambrano, *Macroeconomía*, Quito: Editorial Jurídica del Ecuador, 2011.
- [2] J. Horák y T. Krulický, «Comparison of exponential time series alignment and time series alignment using artificial neural networks by example of prediction of future development of stock prices of a specific company,» *Innovative Economic Symposium 2018 - Milestones and Trends of World Economy (IES2018)*, 2019.
- [3] C. Granger, «Developments in the Nonlinear Analysis of Economic Series,» *Scandinavian Journal of Economics*, vol. 2, n° 93, pp. 263-276, 1991.
- [4] Constitución de la República del Ecuador, Montecristi, 2008. Artículo 302.
- [5] M. Parkin y E. Loria, *Macroeconomía*, México: Pearson Educación, 2010.
- [6] P. A. Samuelson y W. D. Nordhaus, *Macroeconomía. Con aplicaciones a Latinoamérica*, México: McGrawHill, 2010.
- [7] J. Santacruz, «Definición de la inflación según las escuelas económicas,» 2013.
- [8] O. Gutiérrez y A. Zurita, «Sobre la inflación,» *Perspectivas*, vol. IX, n° 3, pp. 339-346, 2009.
- [9] Instituto Nacional de Estadística y Censos, «EcuadorEnCifras,» Febrero 2021. [En línea]. Available: <https://www.ecuadorencifras.gob.ec/indice-de-precios-al-consumidor/>.
- [10] Instituto Nacional de Estadística y Censos, «Índice de Precios al Consumidor (IPC) - Metodología,» Quito, 2019.
- [11] I. Rivera, *Principios de Macroeconomía: Un enfoque de sentido común*, Lima: Fondo Editorial PUCP, 2017.
- [12] A. Roura, «20 años de la dolarización en Ecuador: ¿por qué es tan popular la dolarización entre los ecuatorianos? (y cuál es su lado oscuro),» *BBC News Mundo*, 09 Enero 2020.
- [13] Banco Mundial, «worldbank.org,» 2019. [En línea]. Available: https://databank.worldbank.org/views/reports/reportwidget.aspx?Report_Name=CountryProfile&Id=b450fd57&tbar=y&dd=y&inf=n&zm=n&country=ECU.

- [14] Ekos , «Ekos Negocios,» 01 enero 2020. [En línea]. Available: <https://www.ekosnegocios.com/articulo/10-principales-productos-de-exportacion-del-ecuador>.
- [15] I. Gachet, D. Maldonado y W. Pérez, «Determinantes de la Inflación en una Economía Dolarizada: El Caso Ecuatoriano,» *Cuestiones Económicas*, pp. 5-28, 2008.
- [16] A. Calderón, E. Quinde, I. Orellana y M. Reyes, «Determinantes de la inflación en Ecuador,» *Economía y Sociedad*, vol. 31, pp. 35-64, 2019.
- [17] O. T. Bedri Kamil y T. Selahattin, «Efectos de la dolarización o fi cial en una pequeña economía abierta : el caso de Ecuador,» *Investidación Económica*, vol. LXXIII, pp. 51-86, 2014.
- [18] F. Mochón, Principios de Economía, Madrid: McGraw Hill, 2006.
- [19] C. Mattmann, Machine Learning with TensorFlow, Shelter Island: Manning Publications Co., 2020.
- [20] J. Vergara, «¿Qué son "entrenamiento" e "inferencia" en inteligencia artificial?,» Neuroons, 2021.
- [21] P. Harrington, Machine Learning in Action, Shelter Island,: Manning Publications Co., 2012.
- [22] S. Athey, «The Impact of Machine Learning on Economics,» de *The Economics of Artificial Intelligence*, Chicago, University of Chicago Press, 2019, pp. 507-552.
- [23] A. Lampropoulos y G. Tsihrintzis, Machine Learning Paradigms, London: Springer, 2015.
- [24] S. Skaar, «The Perceptron Model ANNs,» de *A comprehensive guide to Neural Network Modeling*, New York, Nova Science Publishers, 2020.
- [25] L. Deng y D. Yu, Foundations and Trends in Signal Processing, Redmond: Microsoft, 2014.
- [26] J. I. Bagnato, Aprende Machine Learning en Español, LeanPub, 200.
- [27] M. Vochozka, «Using Kohonen's neural networks to identify the bankruptcy of enterprises: Case study based on construction companies in South Bohemian region,»

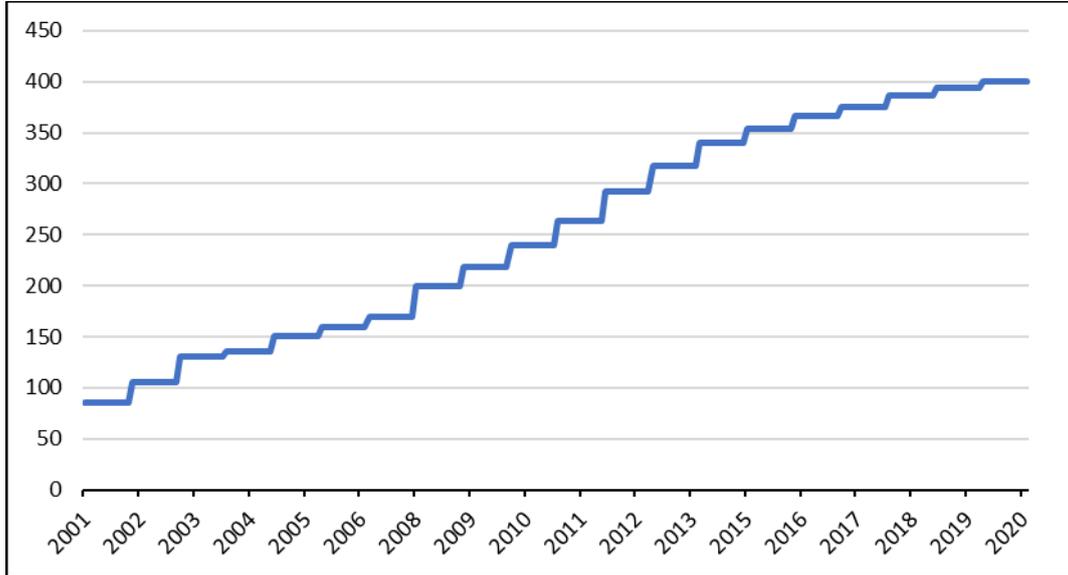
Proceedings of the 5th International Conference Innovation Management, Entrepreneurship and Sustainability, nº 7, pp. 232-239, 2018.

- [28] E. Y. Li, «Artificial Neural Networks and Their Business Applications,» *Information and Management*, 1994.
- [29] D. Gujarati y D. Porter, *Econometría*, Mexico: McGraw-Hill, 2010.
- [30] C. Granger y R. Engle, «Econometría de las series de tiempo, cointegración y heteroscedasticidad condicional autoregresiva,» *Cuestiones Económicas*, vol. 20, nº 2:3, pp. 83-118, 2004.
- [31] A. Novales, *Modelos vectoriales autoregresivos*, 2016.
- [32] D. Morán, «Determinantes de la inflación en Ecuador Un análisis econométrico utilizando modelos VAR,» *Economía y Sociedad*, vol. 18, nº 31, pp. 53-70, 2014.
- [33] S. Guerrero, *Determinantes de la inflación en Ecuador 2000-2016*, Buenos Aires: Universidad de Buenos Aires, 2017.
- [34] A. Mora, «Utilización de Redes Neuronales Artificiales (RNA) como alternativa para el pronóstico del Índice de Precios al Consumidor (IPC) y contraste de resultados con modelos estadísticos tradicionales (ARIMA y VAR) para el Ecuador periodo 2000-2017,» Cuenca, 2018.
- [35] G. Mankiw, *Principios de Economía*, México: Cengage Learning, 2009.
- [36] CEPII, «BACI,» 2019. [En línea]. Available: http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=37.
- [37] D. Mas y D. Ahlfeld, «Comparing artificial neural networks and regression models for predicting faecal coliform concentrations,» *Hydrological Sciences Journal ISSN:*, vol. 52, nº 4, pp. 713-731, 2007.
- [38] G. Eguez, *Análisis del desempeño de redes neuronales artificiales en la reconstrucción de datos pluviométricos de la ciudad de Quito*, Quito, 2020.
- [39] J. Zapata, «Aplicación de las redes neuronales para determinar lo pronósticos mensuales de inflación para el 2018.,» 2019.
- [40] F. Aminia, D. Suarez, M. Aaminian y D. T. Walz, «Forecasting Economic Data with Neural Networks,» *Computational Economics*, vol. 28, pp. 71-88, 2006.

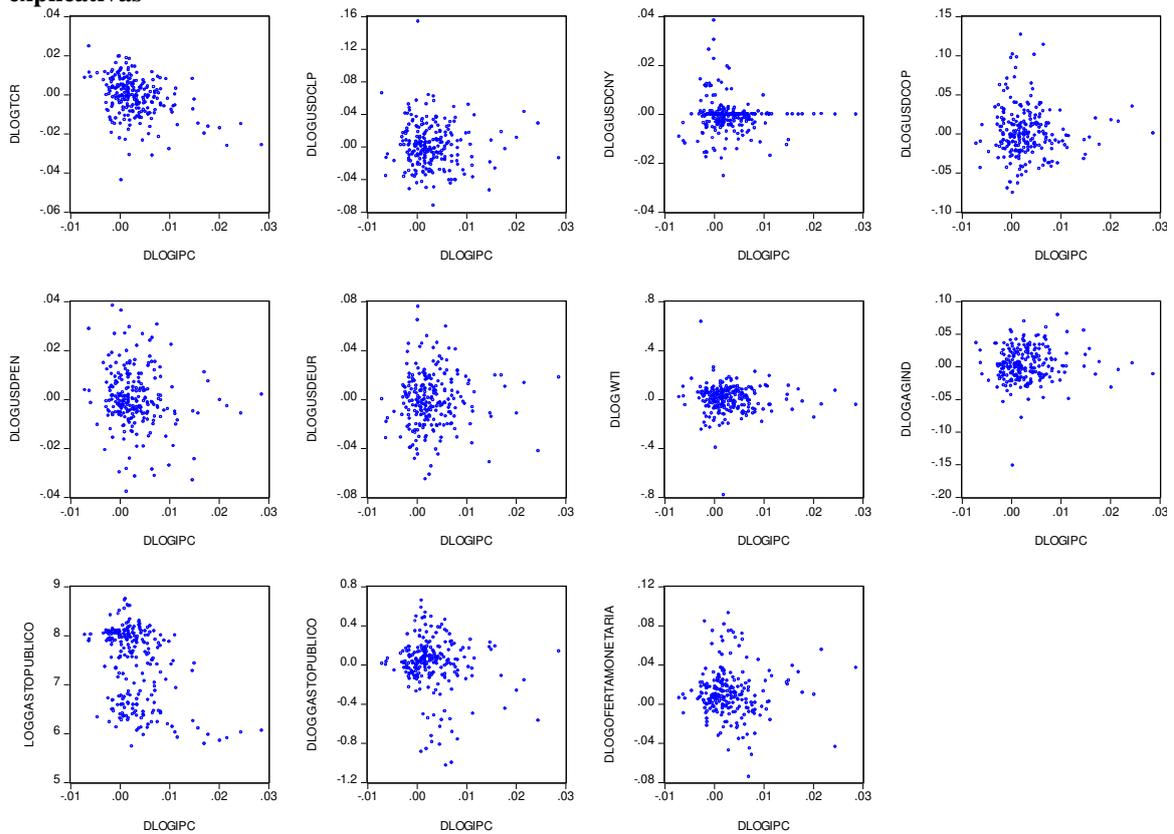
- [41] R. Morales, «UNCTAD, OMC y FMI prevén recuperación del comercio en forma de “paloma”,» *El Economista*, 10 Junio 2021.
- [42] El Universo, «Ecuador cierra el año 2020 con inflación negativa de 1,50%, dice el INEC,» *El Universo*, 7 Enero 2021.
- [43] Banco Central del Ecuador, «Ecuador: Reporte Mensual de Inflación,» BCE, Quito, 2021.
- [44] L. Koskela, Application of the new production philosophy to construction, Finland: VTT Building Technology, 1992, p. 13.
- [45] Instituto Nacional de Estadística y Censos, «Boletín Técnico N°01-2021-IPC,» Quito, 2021.
- [46] J. R. Cuadrado, Política Económica, México: McGraw-Hill, 2010.
- [47] G. Julián, «Las redes neuronales: qué son y por qué están volviendo,» *xataka*, 2014.
- [48] E. Nakamura, «Inflation forecasting using a neural network,» *Economics Letters*, vol. 86, p. 373–378, 2005.
- [49] Banco Central del Ecuador, «1.1.1 OFERTA MONETARIA (M1) Y LIQUIDEZ TOTAL (M2),» *Información Estadística Mensual No. 2030 - Abril 2021*, n° 2030, 2021.
- [50] Ministerio de Producción, Comercio Exterior, Inversiones y Pesca, *Informe Mensual de Comercio Exterior Agosto 2020*, vol. 10, 2020.
- [51] C. Menacho, «Modelos de Regresión Lineal con Redes Neuronales,» *Anales Científicos*, vol. 72, n° 2, pp. 253-260, 2013.

ANEXOS

Anexo 1 - Salario Nominal Ecuador (USD)



Anexo 2 - Diagramas de dispersión de la primera diferencia del logaritmo de las variables explicativas



Anexo 3 - Test de Walt para modelo MCO

Wald Test: Equation: Untitled			
Test Statistic	Value	df	Probability
F-statistic	93.59420	(8, 231)	0.0000
Chi-square	748.7536	8	0.0000
Null Hypothesis: C(1)=0,C(2)=0,C(3)=0,C(4)=0,C(5)=0,C(6)=0,C(7)=0,C(8)=0 Null Hypothesis Summary:			
Normalized Restriction (= 0)	Value	Std. Err.	
C(1)	-0.807398	0.037002	
C(2)	-0.046155	0.008979	
C(3)	-0.156210	0.028630	
C(4)	-0.115082	0.009476	
C(5)	-0.066192	0.018427	
C(6)	-0.122532	0.011387	
C(7)	0.031131	0.008542	
C(8)	0.000369	3.53E-05	

Restrictions are linear in coefficients.

Anexo 4 - Prueba de raíz unitaria en residuos

Null Hypothesis: RESIDL has a unit root Exogenous: Constant Lag Length: 0 (Automatic - based on SIC, maxlag=14)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-9.910551	0.0000
Test critical values:	1% level		-3.457747	
	5% level		-2.873492	
	10% level		-2.573215	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation Dependent Variable: D(RESIDL) Method: Least Squares Date: 07/08/21 Time: 13:06 Sample (adjusted): 2001M03 2020M12 Included observations: 238 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
RESIDL(-1)	-0.569005	0.057414	-9.910551	0.0000
C	-2.03E-06	0.000161	-0.012554	0.9900
R-squared	0.293876	Mean dependent var		-3.87E-05
Adjusted R-squared	0.290884	S.D. dependent var		0.002956
S.E. of regression	0.002489	Akaike info criterion		-9.145484
Sum squared resid	0.001462	Schwarz criterion		-9.116305
Log likelihood	1090.313	Hannan-Quinn criter.		-9.133724
F-statistic	98.21902	Durbin-Watson stat		1.913300
Prob(F-statistic)	0.000000			

Anexo 5 - Valores de los parámetros del modelo VAR

Vector Autoregression Estimates		
Date: 06/29/21 Time: 19:13		
Sample (adjusted): 2001M11 2016M12		
Included observations: 182 after adjustments		
Standard errors in () & t-statistics in []		
	DLOG(B100_201 0)	LOG(SALARION OMINAL)
DLOG(B100_2010(-1))	0.232209 (0.05985) [3.87964]	0.168454 (0.68946) [0.24433]
DLOG(B100_2010(-2))	0.208022 (0.05904) [3.52362]	-0.186910 (0.68006) [-0.27484]
DLOG(B100_2010(-3))	-0.081732 (0.05974) [-1.36817]	-0.768243 (0.68814) [-1.11641]
DLOG(B100_2010(-4))	0.080707 (0.05700) [1.41591]	4.313778 (0.65659) [6.56993]
DLOG(B100_2010(-5))	0.050210 (0.06578) [0.76327]	0.240820 (0.75777) [0.31780]
DLOG(B100_2010(-6))	0.015414 (0.05927) [0.26005]	-1.203428 (0.68277) [-1.76257]
DLOG(B100_2010(-7))	-0.109471 (0.05988) [-1.82823]	-0.647966 (0.68975) [-0.93943]
DLOG(B100_2010(-8))	-0.214966 (0.05706) [-3.76711]	-1.880854 (0.65733) [-2.86134]
DLOG(B100_2010(-9))	0.062556 (0.04956) [1.26236]	1.419929 (0.57084) [2.48745]
LOG(SALARIONOMINAL(-1))	0.005845 (0.00756) [0.77263]	0.802942 (0.08714) [9.21413]
LOG(SALARIONOMINAL(-2))	-0.011616 (0.00939) [-1.23722]	-0.013739 (0.10815) [-0.12703]
LOG(SALARIONOMINAL(-3))	0.026098 (0.00926) [2.81724]	0.094891 (0.10671) [0.88924]
LOG(SALARIONOMINAL(-4))	-0.034205 (0.00950) [-3.60178]	-0.147160 (0.10940) [-1.34520]
LOG(SALARIONOMINAL(-5))	-0.008074 (0.00958) [-0.84309]	0.000666 (0.11032) [0.00604]
LOG(SALARIONOMINAL(-6))	0.023378 (0.00882) [2.64911]	0.154211 (0.10166) [1.51698]

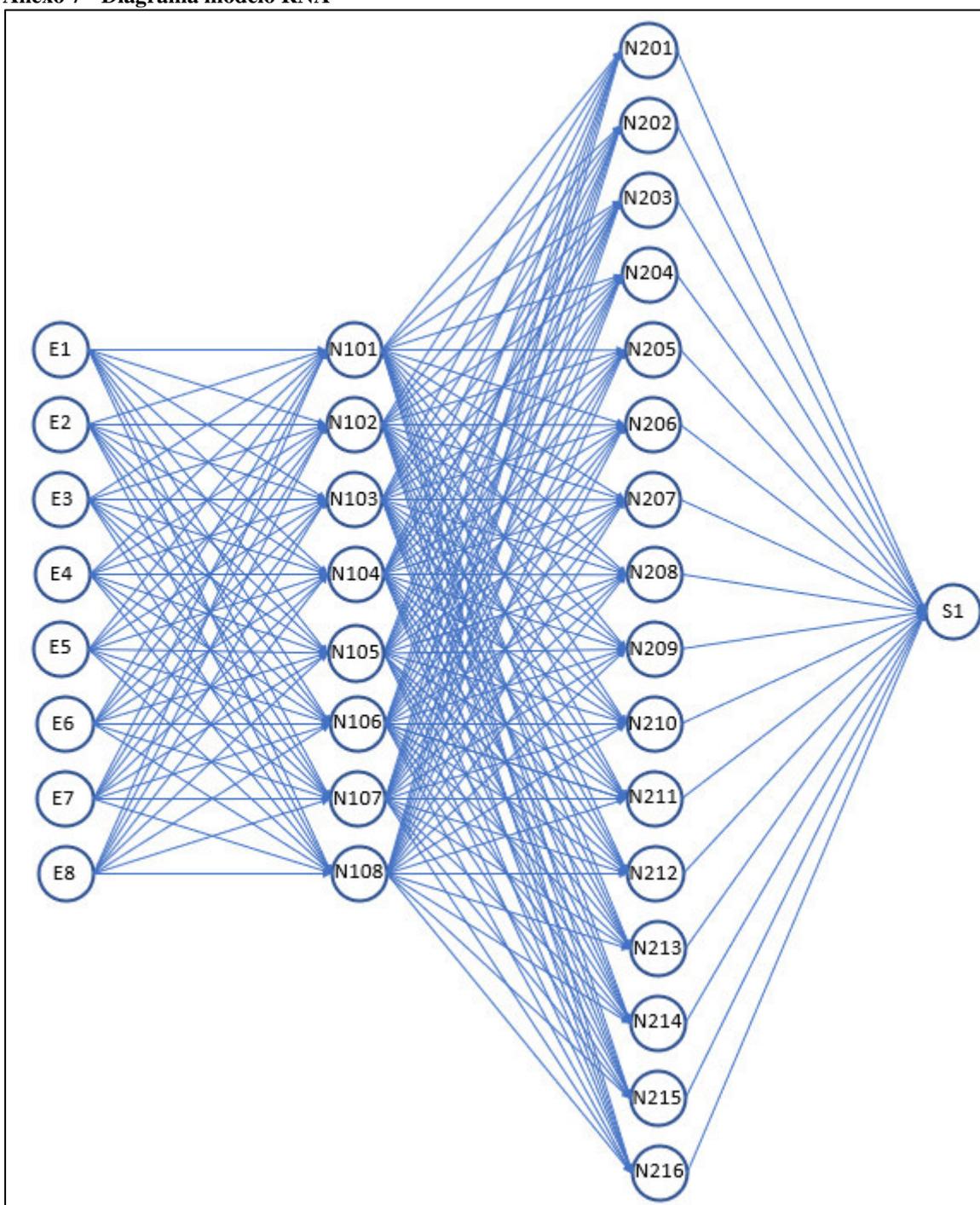
LOG(SALARIONOMINAL(-7))	0.006584 (0.00838) [0.78584]	-0.102155 (0.09651) [-1.05847]
LOG(SALARIONOMINAL(-8))	0.023243 (0.00877) [2.65172]	0.197958 (0.10097) [1.96057]
LOG(SALARIONOMINAL(-9))	-0.030983 (0.00699) [-4.43549]	0.014139 (0.08047) [0.17571]
DLOG(TCRB2014)	-0.630173 (0.04713) [-13.3721]	-2.254581 (0.54286) [-4.15319]
DLOG(EXCH_CHL)	-0.033717 (0.00823) [-4.09667]	-0.113304 (0.09481) [-1.19510]
DLOG(EXCH_CHN)	-0.141475 (0.03404) [-4.15573]	-0.705304 (0.39215) [-1.79854]
DLOG(EXCH_COL)	-0.092856 (0.00878) [-10.5794]	-0.367770 (0.10111) [-3.63748]
DLOG(EXCH_PER)	-0.033491 (0.01637) [-2.04544]	-0.040594 (0.18861) [-0.21522]
DLOG(EXCH_EUR)	-0.103970 (0.01144) [-9.08989]	-0.355579 (0.13176) [-2.69874]
DLOG(AGINDEX)	0.029858 (0.00724) [4.12351]	0.163094 (0.08341) [1.95536]
R-squared	0.784978	0.996986
Adj. R-squared	0.752108	0.996525
Sum sq. resids	0.000664	0.088160
S.E. equation	0.002057	0.023697
F-statistic	23.88156	2164.019
Log likelihood	881.1317	436.3204
Akaike AIC	-9.408041	-4.520005
Schwarz SC	-8.967930	-4.079894
Mean dependent	0.003521	5.351994
S.D. dependent	0.004132	0.402012
Determinant resid covariance (dof adj.)		1.90E-09
Determinant resid covariance		1.42E-09
Log likelihood		1337.586
Akaike information criterion		-14.14929
Schwarz criterion		-13.26907
Number of coefficients		50

Anexo 6 – Prueba de estabilidad modelo VAR

Roots of Characteristic Polynomial Endogenous variables: DLOG(B100_2010) LOG(SALARIONOMINAL) Exogenous variables: C DLOG(TCRB2014) DLOG(EXCH_CHL) DLOG(EXCH_CHN) DLOG(EXCH_COL) DLOG(EXCH_PER) DLOG(EXCH_EUR) DLOG(AGINDEX) Lag specification: 1 9 Date: 07/07/21 Time: 16:45	
Root	Modulus
0.995248	0.995248
0.434722 - 0.820706i	0.928731
0.434722 + 0.820706i	0.928731
0.758145 - 0.499756i	0.908042
0.758145 + 0.499756i	0.908042
0.072339 + 0.862103i	0.865132
0.072339 - 0.862103i	0.865132
-0.435775 + 0.745261i	0.863316
-0.435775 - 0.745261i	0.863316
-0.742212 - 0.440872i	0.863276
-0.742212 + 0.440872i	0.863276
0.790201 + 0.269592i	0.834924
0.790201 - 0.269592i	0.834924
-0.819982 - 0.104510i	0.826615
-0.819982 + 0.104510i	0.826615
-0.266118 + 0.716494i	0.764319
-0.266118 - 0.716494i	0.764319
0.397902	0.397902

No root lies outside the unit circle.
 VAR satisfies the stability condition.

Anexo 7 - Diagrama modelo RNA



Anexo 8 - Pesos de conexiones modelo de RNA

Entrada/1raC	N101	N102	N103	N104
E1: DLOG(TCR)	0,495563	-0,965090	-0,495019	0,971678
E2: DLOG(EXCH_CHL)	0,592072	0,273292	-0,373024	-0,175694
E3: DLOG(EXCH_CHN)	-0,567888	0,305623	0,432699	0,676215
E4: DLOG(EXCH_COL)	0,285960	-0,537442	0,223871	0,023180
E5: DLOG(EXCH_PER)	-0,038223	0,088505	-0,084844	-0,093511
E6: DLOG(EXCH_EUR)	0,458942	-0,253707	0,323686	-0,139181
E7: DLOG(AGINDEX)	0,284033	-0,146340	0,289545	0,364243
E8: LOG(SALNOM)	-0,016467	0,474929	-0,276892	0,111042
Entrada/Neurona	N105	N106	N107	N108
E1: DLOG(TCR)	-0,995019	0,412817	-1,022456	0,694210
E2: DLOG(EXCH_CHL)	-0,293785	0,208443	-0,396434	0,145881
E3: DLOG(EXCH_CHN)	0,162628	0,240323	-0,382414	0,411244
E4: DLOG(EXCH_COL)	-0,375164	-0,151250	-0,286530	-0,025717
E5: DLOG(EXCH_PER)	-0,115140	0,550120	-0,443790	-0,188602
E6: DLOG(EXCH_EUR)	-0,366285	-0,239702	0,287717	0,528423
E7: DLOG(AGINDEX)	0,574838	-0,468331	0,007632	0,160151
E8: LOG(SALNOM)	0,543014	0,076032	0,206096	0,434233

1daC/2daC	201	202	203	204
101	-0,095477	-0,360226	0,186172	0,031801
102	-0,410634	-0,363078	-0,161093	-0,274323
103	0,434124	0,406898	0,492273	0,306491
104	-0,342359	-0,159364	-0,143562	0,344998
105	0,139634	0,173906	-0,146988	0,453080
106	0,024624	0,409770	0,176655	-0,340639
107	0,169364	0,304923	0,081815	-0,107329
108	-0,283301	-0,164311	-0,123933	-0,408486
1daC/2daC	205	206	207	208
101	0,020648	-0,182158	-0,037138	-0,086405
102	-0,169390	-0,395959	0,219932	0,181597
103	-0,146003	0,463857	0,039505	0,427431
104	-0,364923	0,161192	-0,194581	0,469933
105	0,017793	0,290443	0,419707	0,492271
106	-0,036316	-0,243407	-0,209733	-0,031556
107	0,055580	-0,315554	-0,407173	0,304031
108	0,080441	0,212580	-0,320093	0,504582
1daC/2daC	209	210	211	212
101	-0,468742	0,399011	0,106164	0,247682
102	-0,484460	-0,006629	0,176841	-0,097894
103	-0,365869	-0,423822	0,004793	0,403479

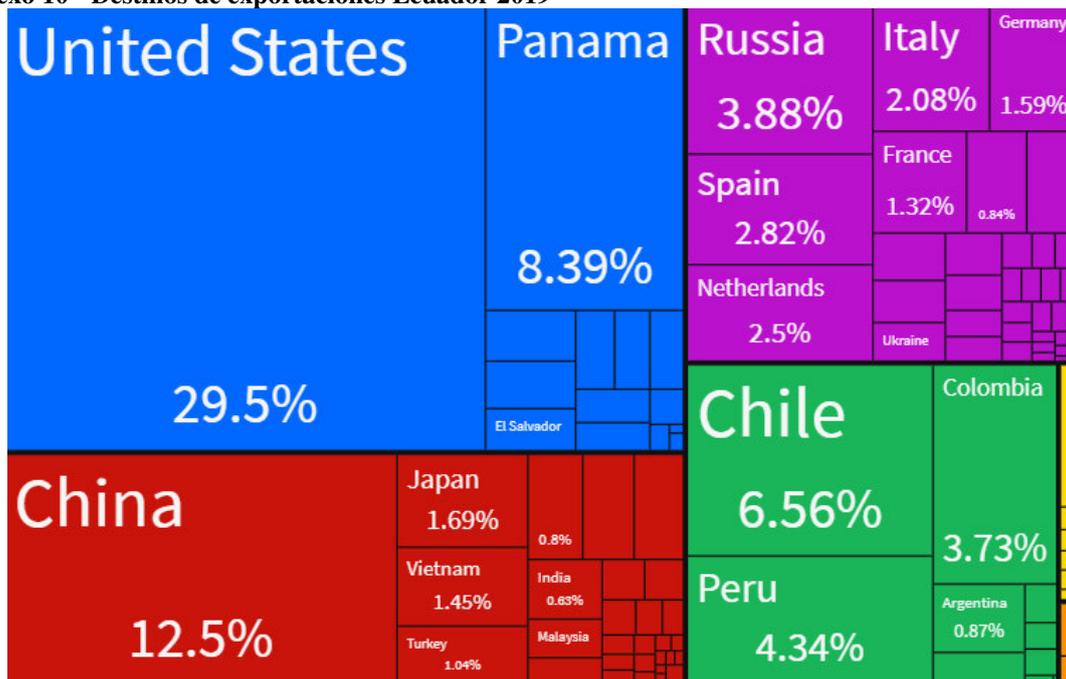
104	-0,207423	-0,250218	-0,011436	0,020470
105	-0,368411	-0,159550	0,344232	0,070919
106	0,069762	0,037002	-0,359045	0,136863
107	0,497987	0,036095	0,387178	-0,065510
108	0,239995	0,375974	-0,333189	-0,254043
1daC/2daC	213	214	215	216
101	-0,442221	0,438319	0,212310	0,491483
102	-0,419625	0,419166	0,341708	0,298331
103	-0,002229	-0,251287	0,470620	-0,128269
104	-0,016233	0,216395	-0,571464	-0,008413
105	0,246425	-0,028503	0,109811	0,000970
106	0,437135	-0,364786	0,368377	0,436413
107	0,182645	0,038849	-0,048548	-0,490615
108	0,470567	-0,151695	0,314302	-0,324201

2daC/Salida	S1 DLOG(IPC)
201	0,266253
202	0,175943
203	-0,077675
204	0,515597
205	0,215076
206	0,352810
207	0,218772
208	-0,119493
209	-0,541469
210	0,286001
211	0,507199
212	0,127813
213	-0,213495
214	-0,121434
215	0,062631
216	0,554317

Anexo 9 - Sesgos neuronas modelo RNA

Neurona	Sesgo
101	-0,003714
102	0,010265
103	0,000000
104	0,011821
105	-0,006649
106	0,001447
107	0,024176
108	0,019453
201	0,000000
202	0,000000
203	0,000000
204	0,000000
205	0,000000
206	-0,040696
207	0,003193
208	-0,000401
209	0,000000
210	0,003380
211	0,003353
212	0,000000
213	-0,000916
214	0,000423
215	0,001447
216	0,000000

Anexo 10 - Destinos de exportaciones Ecuador 2019



Anexo 11 - Coeficiente de determinación ajustado modelo RNA

De acuerdo con Gujarati y Porter [29] el coeficiente de determinación ajustado es la medida de bondad de ajuste más común de una línea de regresión, para ello se puede utilizar la siguiente fórmula:

$$r^2 = \frac{(\sum Y_i \hat{Y}_i)^2}{(\sum Y_i^2)(\sum \hat{Y}_i^2)}$$

Donde, Y_i es el Y_{real} y \hat{Y}_i es el $Y_{estimado}$. Sin embargo, este coeficiente de determinación tiene la desventaja de que se incrementa con el número de variables explicativas, aunque las variables añadidas no tengan relevancia estadística, sobreestimando el modelo. Por ello, se utilizará la siguiente fórmula para calcular el coeficiente de determinación ajustado:

$$R_a^2 = 1 - \left[\left(\frac{n-1}{n-k-1} \right) \right] (1 - r^2)$$

Donde, n es el número de observaciones, k el número de variables independientes y r^2 el coeficiente de determinación. Para el cálculo del coeficiente de determinación se

utilizaron los siguientes valores calculados a partir de los valores reales de DLOG(IPC) y los pronosticados por el modelo de RNA:

$(\sum Y_i \hat{Y}_i)^2$	0,006116506
$(\sum Y_i^2)$	0,006142756
$(\sum \hat{Y}_i^2)$	0,00752793

El resultado tras aplicar las fórmulas es un $R_a^2 = 0,800641993\%$.