

COMPORTAMIENTO DE LA UVR EN EL LARGO PLAZO

EVELYN LUCÍA TAYLOR CONTO

etaylor@unal.edu.co

*Universidad Nacional de Colombia / Escuela de Estadística
Carrera 65 Nro. 59A - 110, Medellín - Colombia*

ROMARIO ADEMIR CONTO LÓPEZ

romarioconto@itm.edu.co

*Instituto Tecnológico Metropolitano / Departamento de Finanzas
Calle 54A No. 30 – 01, Medellín – Colombia*

Recibido (26/09/2022)

Revisado (28/08/2023)

Aceptado (15/10/2023)

RESUMEN: En este artículo se implementan diferentes técnicas para predecir la unidad de valor real colombiana (UVR), basándose únicamente en el histórico de su comportamiento. La UVR representa el poder adquisitivo basado en la variación del índice de precios al consumidor (IPC) para el mes calendario inmediatamente anterior al calculado, además muchos créditos de vivienda están indexados a la UVR en Colombia, permitiendo que las entidades financieras conserven el poder adquisitivo del dinero prestado. Para el estudio son consideradas técnicas de pronóstico mediante modelos SARIMA, regresión lineal, suavizamiento Holt-Winters aditivo y multiplicativo, y redes neuronales artificiales. De acuerdo con el análisis, se obtiene que la mejor técnica para realizar predicciones al final del ciclo estacional es el suavizamiento Holt-Winters multiplicativo y para predecir a mitad del ciclo estacional es un modelo SARIMA (1,1,1) (0,1,1).

Palabras claves: UVR, series de tiempo, SARIMA, regresión lineal, Holt-Winters, redes neuronales.

ABSTRACT: In this paper, we implemented different techniques to predict the Colombian real value unit (UVR), based solely on the history of its behavior. The UVR represents the purchasing power based on the variation of the consumer price index (CPI) for the calendar month immediately before the calculated one., in addition, many housing credits are indexed to UVR in Colombia, allowing financial institutions to preserve the purchasing power of money borrowed. For the study, forecasting techniques are considered using SARIMA models, linear regression, additive and multiplicative Holt-Winters smoothing, and artificial neural networks. According to the analysis, it is obtained that the best forecasting technique to make predictions at the end of the seasonal cycle was the multiplicative Holt-Winters smoothing and to predict in the middle of the seasonal cycle was the SARIMA (1,1,1) (0, 1,1).

Keywords: UVR, time series, SARIMA, lineal regression, Holt-Winters, neural networks.

1. Introducción

Los pronósticos representan una herramienta que permite reducir la incertidumbre presente en la toma de decisiones en diferentes contextos, por lo que encontrar una metodología adecuada para predecir se convierte en un elemento de gran interés desde el punto de vista estadístico y econométrico. Los métodos de análisis de regresión, suavizamientos, métodos Box-Jenkins y redes neuronales, son ejemplos de técnicas disponibles en la literatura para el cálculo de pronósticos. Estos consideran como variables regresoras sus propios rezagos en el tiempo para estimar el valor futuro de la variable de interés.

En Colombia existe la llamada unidad de valor real (UVR), la cual es una medida certificada por el Banco de la República, con periodicidad mensual y cuya metodología usada para calcularla es establecida por la entidad en la resolución externa 13 de 2000 (Banco de la República de Colombia, 2022). Esta medida fue establecida en 1999 en reemplazo de la antigua unidad de poder adquisitivo constante (UPAC) y es de interés pronosticarla debido a que es usada para calcular los costos de los créditos de vivienda, logrando que las entidades financieras conserven el poder adquisitivo del dinero prestado.

En la literatura se han encontrado pocos estudios donde aplican métodos de pronóstico de la UVR. Por ejemplo, Velásquez y Aguilar (2011) usan un modelo SARIMA para predecirla, tomando como período de estudio los años 2000 a 2007. El modelo usado estima los coeficientes autorregresivos (AR) y de medias móviles (MA) que resultan ser significativos y encuentra que es un modelo que se ajusta adecuadamente a los datos y cumple con los supuestos estadísticos.

En este artículo se busca implementar modelos univariados que consideren las componentes de la UVR a largo plazo para realizar pronósticos eficientes de esta variable. La exploración de diferentes técnicas de pronósticos permite que se pueda identificar una metodología apropiada para el comportamiento histórico de la UVR. En este caso se usan métodos paramétricos y no paramétricos, así como herramientas de Machine Learning, como lo es el uso de redes neuronales.

El resto del documento está organizado de la siguiente manera: la sección 2 presenta la revisión de literatura sobre los modelos de pronóstico disponibles, la sección 3 describe la metodología usada en este trabajo, destacando la formulación teórica de cada una de las técnicas de pronóstico implementadas. En la sección 4 se muestra el análisis y discusión de los hallazgos para luego, en la sección 5, presentarse las principales conclusiones del estudio. Al final del documento, se presentan las referencias consultadas.

2. Revisión de literatura

En la literatura es posible encontrar diversos métodos de pronóstico para series de tiempo univariadas en diferentes campos como la economía, finanzas, producción, mercadeo, entre otros. En este escenario, suele suceder que sea implementada un solo un tipo de técnica de pronóstico o también que se realice un comparativo entre diferentes métodos que se consideran apropiados. Por ejemplo, Velásquez y Aguilar (2011) realizan un estudio en el que buscan predecir la UVR con periodicidad mensual y para ello usan un modelo SARIMA, para el cual obtienen un buen ajuste a los datos y cumplimiento de los supuestos estadísticos.

Tsitsika et al. (2007) implementan modelos ARIMA univariados y multivariados estacionales para modelar y pronosticar la producción mensual de peces en el mar Mediterráneo entre 1990 y 2005. Donde los modelos univariados tuvieron un mejor desempeño. Los resultados del estudio brindaron un apoyo directo para el uso potencial de pronósticos precisos en la toma de decisiones y la gestión de la pesca en el mar Mediterráneo.

Chen et al. (2009) exploran tres métodos de pronóstico con el fin de realizar una predicción precisa de la demanda de transporte aéreo en Taiwán, la cual es una variable de gran interés en este país insular, pues juega como un indicador importante del turismo internacional. En el artículo se implementan el método Holt-Winters, el modelo SARIMA y el modelo "Grey Forecasting" GM (1,1), presentado por Deng (1982), utilizando los datos de series temporales de 1996 a 2007. Los resultados revelan que el modelo SARIMA supera a los modelos Holt-Winters y GM (1,1) en el ajuste a los datos, y en la precisión de sus pronósticos todos los modelos ajustados tienen un buen rendimiento, con valores del MAPE inferiores al 20%.

Braimllari y Sala (2016) realizan un estudio donde modelan el nivel de las importaciones de alimentos en Albania con el fin de realizar un pronóstico. Su objetivo principal es encontrar el modelo más apropiado para esta serie mensual, analizando información desde enero de 2005 a octubre de 2016. Los dos métodos usados son el Holt-Winters y el modelo ARIMA estacional. Se encontró que el Holt-Winters fue más preciso para realizar los pronósticos y el ARIMA estacional fue el mejor en el ajuste.

Akpinar y Nejat (2016) aplican diferentes métodos de pronóstico para estudiar la demanda estacional mensual de gas natural en la provincia de Sakarya en Turquía, durante los años de 2011 al 2014. Los métodos estudiados son el Holt-Winters, el suavizamiento exponencial, el ARIMA estacional y la descomposición de la serie. Los datos fueron separados en dos series, la primera de 2011 a 2013 para estudiar el comportamiento histórico y la segunda durante 2014 para estudiar la precisión de los pronósticos. Se obtuvo que todos los métodos proporcionaban resultados satisfactorios, siendo el ARIMA estacional quien obtuvo pronósticos más consistentes. Por lo que, si se dispone de los conocimientos y el software especializado para su implementación, el modelo ARIMA estacional sería el implementado. En el caso de no disponer de esto, los otros métodos pueden ser aplicados de manera sencilla en una hoja de cálculo.

Al-Gounmeein y Ismail (2020) realizan la implementación de modelos Box-Jenkins, con el ajuste de un modelo ARIMA y un modelo SARIMA para la serie mensual del tipo de cambio entre el Dinar Jordano y el Dólar Estadounidense, entre el 2008 y 2018. Ellos encontraron que el modelo SARIMA (1,0,1) (1,0,0)₁₂ era un modelo idóneo para modelar y pronosticar la serie, pues pasó las pruebas diagnósticas, lo cual indicaba que la serie temporal era estacional. Para encontrar el mejor modelo se usó el AIC y el BIC. El diagnóstico del modelo se realizó mediante las pruebas de Ljung-Box y la prueba de normalidad de Kolmogorov-Smirnov. La precisión de los pronósticos fue evaluada mediante el MAE, MAPE y RMSE.

Barría-Sandoval et al. (2021) modelan los diagnósticos y las muertes por COVID-19 en Chile usando los métodos ARIMA, técnicas de suavizamiento exponencial y modelos Poisson para datos de conteo que son dependientes en el tiempo, para ello usan un conjunto de datos de entrenamiento y otro de prueba para evaluar la precisión de los pronósticos. Se obtuvo que al analizar el comportamiento de estas series temporales diarias entre el 2 de marzo y el 14 de julio de 2020 el modelo que resulta ser más apropiado para predecir el número de casos confirmados es el ARIMA y el método de suavizado usado que permite amortiguar las tendencias erráticas fue el mejor para predecir el número de muertes. Los autores concluyen que dependiendo de las características de la serie temporal puede resultar ser mejor un método u otro.

Yamacli D. y Yamacli S. (2023) estudian la eficiencia de usar modelos ARIMA y redes neuronales artificiales para pronosticar la tasa de desempleo mensual en Turquía. Para ello, analizaron el período comprendido entre enero de 2008 y agosto de 2022, y obtuvieron que ambos modelos arrojaban pronósticos cercanos a los valores reales de desempleo, pero fue el modelo ARMA (2,1) el más acertado según los criterios del RMSE, MAE y MAPE. Los autores resaltan que para el período de pandemia 2020-2021 el error de pronóstico fue menor para la red neuronal ajustada, indicando que, en momentos de incertidumbre económica, el uso de la red neuronal proporciona más información.

Youness y Driss (2022, May) estiman un modelo ARIMA (0,1,1) para predecir el comportamiento de la tasa de desempleo en Marruecos entre 1971 y 2019, el cual exhibe buenos resultados, pero presenta volatilidad, por lo que se recomienda evaluar el uso de métodos de aprendizaje automático en comparación con los modelos ARIMA clásicos. Posteriormente, Youness y Driss (2023) comparan el uso de regresión con máquinas de soporte vectorial (SVR) y modelos ARIMA en términos de capacidad de pronóstico mediante para la serie del nivel de inversión extranjera directa estadounidense, entre 1972 y 2021. Los autores encontraron que, para pronósticos a corto plazo, ambos métodos reducen significativamente el error de pronóstico, pero que es la técnica de aprendizaje automático usada, la regresión SVM, que funciona mejor, debido a que sus medidas de precisión son más bajas.

Bilgili et al. (2023) analizan diferentes enfoques que permitan predecir el consumo bruto de electricidad en Turquía en función de su histórico y otras variables. Para ello utilizaron un modelo de red neuronal de memoria a corto plazo (LSTM) y un modelo SARIMA, obteniéndose resultados similares, pero fue el método LSTM que superó al SARIMA, con valores del MAPE, MAE y RMSE más bajos.

Otros trabajos como los de Goh & Law (2001) usan un modelo SARIMA y un MARIMA para predecir la demanda de turismo en Hong Kong. Gil (2016) realiza el pronóstico de la demanda mensual de electricidad en Colombia mediante un modelo regresión con variables indicadoras, pues la serie tenía componentes de tendencia y estacionalidad. Villazón-Bustillos et al. (2016) proponen un modelo ARIMA estacional y redes neuronales artificiales para predecir el próximo evento de sequía en el noroeste de México.

Esta revisión permite ver la gran aplicación del análisis de las series de tiempo en el pronóstico de variables de todo tipo. Donde, en la mayoría de los casos, las investigaciones analizan el ajuste de los métodos a los datos y su capacidad de pronóstico, debido a que no necesariamente el modelo que mejor ajusta es el que mejor pronostica, pero cuya decisión final suele ser por el que sea mejor prediciendo.

3. Metodología

3.1. Cálculo de la unidad de valor real (UVR)

El valor en moneda legal colombiana de la UVR según el artículo 3 de la ley 546 de 1999 es hallada con periodicidad diaria de acuerdo con la Ec. (1).

$$UVR_t = UVR_{15}(1 + i)^{\frac{t}{d}} \quad (1)$$

donde UVR_t es el valor en moneda legal colombiana de la UVR en el día t para período de cálculo. UVR_{15} es el valor en moneda legal colombiana de la UVR el día 15 de cada mes. i representa la variación mensual del índice de precios al consumidor certificada por el DANE (Departamento Administrativo Nacional de Estadística colombiano) durante el mes calendario inmediatamente anterior al mes del inicio del período de cálculo. t es el número de días calendario transcurridos desde el inicio de un período de cálculo hasta el día de cálculo de la UVR; por tanto, t tendrá valores entre 1 y 31, de acuerdo con el número de días calendario del respectivo periodo de cálculo. Finalmente, d es igual al número de días calendario del respectivo periodo de cálculo comprendido entre el día 16 de un mes hasta el día 15 del mes siguiente (ambos inclusive).

3.2. Regresión lineal

Mediante el análisis de regresión es posible modelar las componentes de una serie temporal, cuya estructura puede ser de componentes aditivas o multiplicativas, dependiendo de si los picos estacionales presentan alturas cambiantes o constantes (Diebold, 1999). Si se denota Y_t a la serie temporal, T_t la componente de tendencia, S_t la componente estacional y E_t el error aleatorio, el modelo aditivo considera que la serie temporal puede ser modelada como $Y_t = T_t + S_t + E_t$, y el multiplicativa como $Y_t = T_t \times S_t \times E_t$. Esta última requiere una transformación para linealizar la relación, tal que $\ln(Y_t) = T_t^* + S_t^* + E_t^*$. Mediante variables indicadoras se puede modelar la componente estacional aditiva y tendencia lineal de Y_t como:

$$Y_t = \beta_0 + \beta_1 t + \sum_{i=1}^{s-1} \delta_i I_{i,t} + \varepsilon_t \quad (2)$$

con $\varepsilon_t \sim IIDN(0, \sigma^2)$. Donde δ_i con $i = 1, 2, \dots, s-1$, es el efecto estacional al estar en la estación i respecto a la última estación s . $I_{i,t}$, con $i = 1, 2, \dots, s-1$; la indicadora de la estación i . Mediante funciones trigonométricas es posible representar a la serie temporal Y_t como:

$$Y_t = T_t + \sum_{i=1}^k \left[\alpha_j \text{Sen} \left(\frac{2\pi jt}{s} \right) + \gamma_j \text{Cos} \left(\frac{2\pi jt}{s} \right) \right] + \varepsilon_t \quad (3)$$

con $\varepsilon_t \sim IIDN(0, \sigma^2)$. Donde $k = s/2$ son la cantidad de ciclos estacionales posibles dado el período estacional de tamaño s .

3.3. Métodos de suavización

Los métodos de suavización permiten generar pronósticos basándose en los valores pasados de la serie temporal ponderándolos de manera exponencialmente decreciente (Hanke y Wichern, 2010). Dentro de estos métodos están el suavizamiento exponencial simple, el suavizamiento Holt, el suavizamiento de Brown y el suavizamiento Holt-Winters aditivo y multiplicativo. Una serie temporal con tendencia lineal local y componente estacional aditiva puede ser pronosticada mediante un suavizamiento Holt-Winters aditivo, implementando las siguientes cuatro ecuaciones (Hanke y Wichern, 2010)

- i. Estimación del nivel de la serie: $L_t = \alpha(Y_t - S_{t-s}) + (1 - \alpha)(L_{t-1} + T_{t-1})$
- ii. Estimación de la tendencia de la serie: $T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$
- iii. Estimación de la estacionalidad de la serie: $S_t = \gamma(Y_t - L_t) + (1 - \gamma)S_{t-s}$
- iv. Pronóstico para p períodos futuros: $\hat{Y}_{t+p} = (L_t + pT_t) + S_{t-s+p}$

donde Y_t es el valor de la serie actual, L_t es el estimado del nivel actual, α es la constante de suavización del nivel, T_t es la estimación de la tendencia, β la constante de suavización para el estimado de la tendencia, S_t es el estimado de la componente estacional, γ es la constante de suavizamiento para el estimado estacional, p los períodos futuros a pronosticar, s la longitud del período estacional y \hat{Y}_{t+p} el pronóstico para el período p futuro.

En el caso en que la serie tiene tendencia lineal local con componente estacional multiplicativa, es posible aplicar el Holt-Winters multiplicativo mediante las siguientes ecuaciones:

- i. Estimación del nivel de la serie: $L_t = \alpha(Y_t/S_{t-s}) + (1 - \alpha)(L_{t-1} + T_{t-1})$
- ii. Estimación de la tendencia de la serie: $T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$
- iii. Estimación de la estacionalidad de la serie: $S_t = \gamma(Y_t/L_t) + (1 - \gamma)S_{t-s}$
- iv. Pronóstico para p períodos futuros: $\hat{Y}_{t+p} = (L_t + pT_t)S_{t-s+p}$

3.4. Métodos Box-Jenkins

Box & Jenkins en 1976 publicaron esta metodología orientada a la construcción de modelos para series temporales, con el fin de encontrar el proceso generador de los datos. La metodología está basada en un procedimiento de cuatro pasos: identificación, estimación, diagnóstico y pronósticos. En la etapa de identificación se proponen modelos tentativos para la serie, en la estimación son hallados los coeficientes estimados, en el diagnóstico se realizan pruebas para determinar si el o los modelos son adecuados y posiblemente proponer uno mejor. Finalmente, en la etapa de pronóstico se estiman los valores futuros de la serie (Bowerman, et al., 2009; Kumar y Anand, 2015; Stellwagen y Tashman, 2013)

Los modelos Box-Jenkins son empleados en el análisis de series temporales estacionarias. Una serie estacionaria es aquella cuya media y varianza son constantes en el tiempo (Hanke y Wichern, 2010). La estacionariedad de una serie puede ser estudiada mediante las pruebas Dickey y Fuller (1979), y Kwiatkowski, et al (1992) para la parte regular y con la prueba Hylleberg, et al (1990) o Canova y Hansen (1995) para la parte estacional.

Una serie temporal estacional puede ser modelada a partir de los métodos Box-Jenkins mediante el llamado modelo ARIMA estacional o SARIMA. El modelo considera términos autorregresivos y de medias móviles para explicar las correlaciones temporales en períodos estacionales y no estacionales. Para la implementación del modelo es necesario que la serie temporal sea estacionaria tanto en su parte regular (dentro de los períodos estacionales) como en su parte estacional (entre períodos estacionales), en cuya ACF se notará un decrecimiento rápido de los coeficientes de autocorrelación a cero. Debido a esto, si la

serie no es estacionaria, se hace necesario diferenciarla tanto regular como estacionalmente para especificar completamente el modelo (Hanke y Wichern, 2010).

El operador diferencia regular (∇) es tal que $\nabla Y_t = Y_t - Y_{t-1}$; efectúa la diferencia entre el valor de la serie en t y $t - 1$. El operador diferencia estacional (∇_s) es tal que $\nabla_s Y_t = Y_t - Y_{t-s}$; realiza la diferencia entre el valor de la serie en t con el de $t - s$. El operador diferencia tiende a volver estacionaria una serie que inicialmente no lo era, ya sea en la parte regular o estacional.

El modelo $SARIMA(p, d, q)(P, D, Q)_s$ se puede escribir de manera general como: (Box et al., 2008):

$$\phi_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D Y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \quad (4)$$

donde,

- $\Phi_P(B^s) = 1 - \phi_1 B^s - \phi_2 B^{2s} - \dots - \phi_P B^{Ps}$ el polinomio autorregresivo estacional de orden P y período estacional s , denotado $SAR(P)$.
- $\Theta_Q(B^s) = 1 - \theta_1 B^s - \theta_2 B^{2s} - \dots - \theta_Q B^{Qs}$ el polinomio de medias móviles estacional de orden Q y período estacional s , denotado $SMA(Q)$.
- $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ el polinomio autorregresivo regular de orden p , denotado $AR(p)$.
- $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ el polinomio de medias móviles regular de orden q , denotado $MA(q)$

con $\varepsilon_t \sim IIDN(0, \sigma^2)$. El factor B es el operador de rezagos, tal que $B^j Y_t = Y_{t-j}$. ∇^d es el operador diferencia de orden d y ∇_s^D es el operador diferencia estacional de orden D . Los coeficientes asociados a los polinomios autorregresivos y de medias móviles serán estimados mediante el conjunto de observaciones históricas disponibles.

3.5. Redes neuronales artificiales

Las redes neuronales permiten obtener pronósticos basados en la modelación matemática del comportamiento del cerebro, consiguiendo estudiar relaciones más complejas entre la respuesta y las variables predictoras. En este contexto, las predictoras forman la llamada capa de entrada y los pronósticos son la capa de salida. También es posible que se tengan capas intermedias, denominadas “ocultas” con un conjunto de “neuronas ocultas”. La Figura 1, muestra la estructura de una red neuronal con cinco neuronas en la capa de entrada cuatro en la capa oculta y una capa de salida.

Una red neuronal estima unos coeficientes adjuntos a cada entrada, los cuales se denominan “pesos” y los pronósticos son obtenidos considerando una combinación lineal de las entradas. La red neuronal usa un “algoritmo de aprendizaje” buscando minimizar una función como el MSE, donde las salidas de los nodos en las capas anteriores se convierten en las entradas de la capa siguiente, ponderándose mediante una combinación lineal, para luego el resultado ser modificado con una función no lineal. (Hyndman, y Athanasopoulos, 2021).

En el contexto de series de tiempo, los rezagos se convierten en entradas para la red, construyendo el llamado modelo de autorregresión de red neuronal o modelo NNAR. Para datos estacionales, conviene usar la estructura $NNAR(p, P, k)_m$, la cual implica que, dado un período estacional m , se usan p rezagos de la serie $y_{t-1}, y_{t-2}, \dots, y_{t-p}$, P rezagos estacionales $y_{t-m}, y_{t-2m}, \dots, y_{t-pm}$ y k neuronas en la capa oculta. Para el cálculo de pronósticos, la red se aplica de manera iterativa, debido a que para predecir un paso adelante, se usan todos los datos históricos y para dos pasados adelante se usa el pronóstico de un paso más el histórico, y así sucesivamente (Hyndman, y Athanasopoulos, 2021).

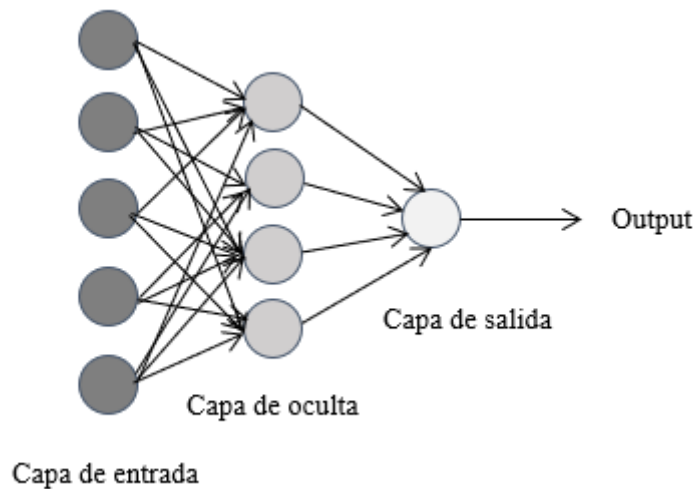


Figura 1 Red neuronal con cinco entradas, una capa oculta con cuatro neuronas. Elaboración propia.

3.6. Estrategia empírica

Para este estudio se toman 264 datos correspondientes a la UVR mensual desde enero de 2000 a diciembre de 2021 y se procede a analizar sus componentes y/o estructura de autocorrelación histórica. De esta manera es posible definir el tipo de modelos factibles para ajustar y predecir el comportamiento futuro de la serie. Luego, estos modelos son estimados usando la estrategia de validación cruzada, escogiendo una parte de los datos para entrenamiento de los modelos y otra para la validar su capacidad de pronóstico. En este caso se consideraron dos escenarios: dejar el último medio año (6 observaciones) y dejar todo el último año (12 datos) para medir la capacidad de pronóstico de cada técnica. Finalmente, son usados los criterios de evaluación MAPE, RMSE y MAE para comparar los métodos seleccionados. Cabe resaltar que, para los métodos Box-Jenkins y métodos de regresión, es necesario validar el cumplimiento de los supuestos sobre los cuales se basa la estimación del modelo, ya que estos son métodos paramétricos que asumen una distribución de probabilidad para sus estimaciones. En este caso se usará la prueba creada por Shapiro y Wilk (1965) para chequear la normalidad de los errores, la prueba Ljung-Box para probar independencia y el gráfico de residuales contra valores estimados de la serie para evaluar la homocedasticidad de los errores. Las metodologías implementadas en este artículo son: regresión lineal, suavizamiento Holt-Winters aditivo y multiplicados, modelo SARIMA y redes neuronales artificiales. Todo esto con ayuda del software R studio.

4. Análisis y discusión de resultados

Las medias descriptivas básicas para la UVR se presentan en la Tabla 1. Valores como su media, desviación estándar, coeficientes de asimetría, curtosis, y valores mínimo y máximo son reportados. En este caso se observa que la UVR histórica presenta media mayor a la mediana, por lo que hay asimetría a la derecha y una curtosis por negativa, implicado que los datos no siguen una distribución normal, pues esta tiene un coeficiente de asimetría igual a cero y curtosis de 3. Adicionalmente, la prueba Shapiro-Wilk, rechaza la hipótesis de normalidad, pues su p-valor es 6.953×10^{-5} (menor a un nivel de significancia $\alpha = 0.05$). La Figura 2, muestra el comportamiento histórico de la UVR (izquierda) y su ACF (derecha), donde es claro que presenta una tendencia global creciente con un decrecimiento muy lento en sus coeficientes de autocorrelación. La Figura 3 (izquierda) presenta el gráfico de la descomposición aditiva de la serie, notándose una componente estacional que inicialmente no se revelaba. La componente aleatoria, en este caso, presenta mayor variabilidad en los últimos períodos, indicando heterocedasticidad. La Figura 3 (derecha) muestra cómo la descomposición multiplicativa deja ver una componente estacional y una

componente aleatoria que ya es homocedástica, lo cual es evidencia de una posible componente estacional multiplicativa.

Tabla 1. Medidas descriptivas para la UVR mensual histórica.

	<i>UVR</i>
Nobs	264
Minimum	103.85
Maximum	288.62
1. Cuartil	151.865
3. Cuartil	240.6325
Media	194.145341
Mediana	191.31
Varianza	2655.970632
Desv. Estándar	51.53611
Asimetría	0.107504
Curtosis	-1.097880

Fuente: Elaboración propia programa R.

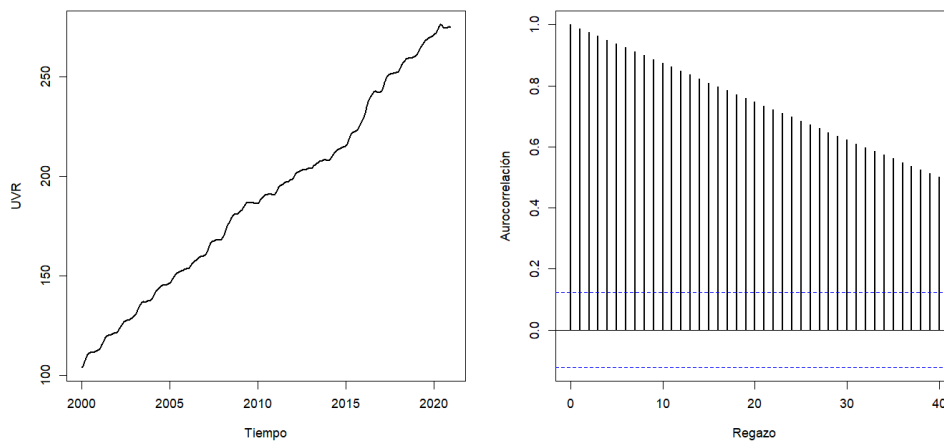


Figura 2. Gráfico de la UVR histórica (izquierda) y ACF d la UVR histórica. Elaboración propia.

Se aplicó la prueba KPSS de raíz unitaria regular y la prueba SEAS para raíz unitaria estacional y ambas concluyeron que la serie histórica de la UVR no es estacionaria ni en la parte regular (su tendencia), ni en la parte estacional, ya que sus p-valores fueron mayores al nivel de significancia del 5%.

En esta parte se tendrán en cuenta dos procedimientos para el ajuste y pronóstico: primero se consideran para el período de entrenamiento 252 datos y para el período de testeo las últimas 12 observaciones (escenario 1). Después se consideran para entrenamiento 258 datos y para el período de testeo las últimas 6 observaciones (escenario 2). Analizando el escenario 1, la primera metodología implementada es el modelo SARIMA o ARIMA estacional, para la cual en este caso se requiere la diferenciación regular y estacional de la serie con el objetivo de eliminar la no estacionariedad de los datos debida a la tendencia y la estacionalidad. Luego de hacer dichas diferencias, se analizaron las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) para identificar los órdenes AR y MA regulares y estacionales. De acuerdo con la Figura 4, se identifica un modelo SARIMA (1,1,1) (0,1,1). Pues en la parte intra-período estacional (rezago del 1 al 11) se presentan colas decrecientes a las bandas con pico en el primer rezago, indicando un ARMA (1,1) y entre períodos estacionales, se nota un patrón de decrecimiento en PACF y

otro de corte bastante significativo en el rezago $k=12$ (primer rezago estacional) en la ACF. El resultado de la estimación de este modelo se puede ver en la Tabla 2.

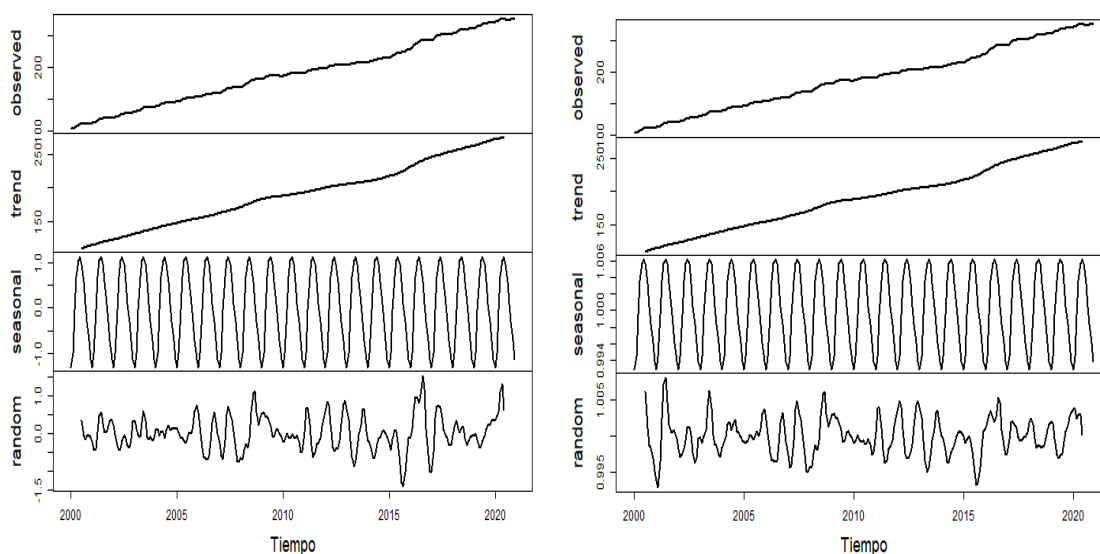


Figura 3. Descomposición de la UVR histórica: aditiva (izquierda) y multiplicativa (derecha). Elaboración propia.

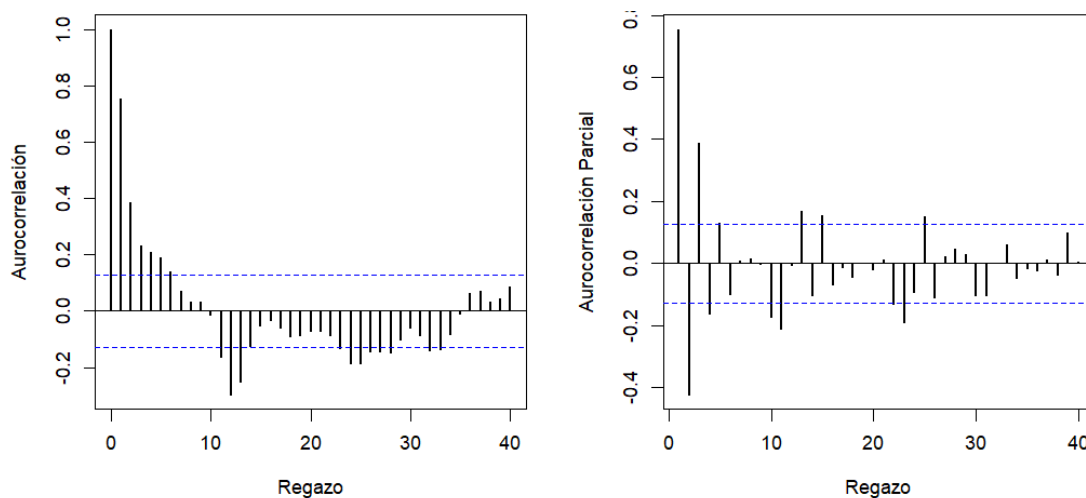


Figura 4. ACF (izquierda) y PACF (derecha) de la UVR histórica diferenciada regular y estacionalmente. Elaboración propia.

También se aplica el modelo de regresión lineal considerando variables indicadoras y otro con funciones trigonométricas, con el fin de explicar la estacionalidad. Pero ninguno de estos modelos arrojó significancia en los parámetros estacionales. Por lo que se optó por una regresión solo contra la tendencia temporal lineal, ya que el coeficiente de determinación fue muy alto (99.26%). La Tabla 2. también relaciona los resultados de dicha estimación, indicando que el modelo tiene parámetros significativos o distintos de cero, pues sus p-valores para las pruebas de significancia son menores al 5%. Posteriormente se aplican los suavizamientos Holt-Winter Aditivo y Multiplicativo, cuyos resultados se pueden ver en la Tabla 2. Los parámetros para el nivel, pendiente y factor estacional fueron optimizados usando la función `HoltWinters()` en R studio.

Tabla 2. Estimaciones para los modelos en el escenario 1

Modelo SARIMA (1,1,1) (0,1,1)		
	<i>Estimación</i>	<i>P-valor</i>
ϕ_1	0.5806	0
θ_1	0.8016	0
Φ_1	-0.8922	0
Modelo de regresión lineal		
	<i>Estimación</i>	<i>P-valor</i>
<i>Intercepto</i>	105.4	0
<i>Pendiente</i>	0.668	0
<i>R cuadrado</i>	99.26%	-
Suavizamiento Holt-Winters		
	<i>Aditivo</i>	<i>Multiplicativo</i>
α	1	1
β	0.3091	1
γ	1	1

Fuente: Elaboración propia programa R.

Finalmente, fue aplicada una red neuronal autorregresiva mediante la función `NNETAR()` en R. La cual estima que un modelo factible para la serie de la UVR es un NNAR (1,1,2)₁₂, usando el primer rezago regular, el primer rezago estacional y 2 neuronas en la capa oculta.

Con cada modelo se procedió a predecir los 12 períodos siguientes, correspondientes a los 12 meses del año 2021. Los pronósticos resultantes con los cinco métodos y los valores reales para 2021 de la UVR se pueden ver en la Figura 5. También, es posible ver las medidas de capacidad de pronóstico en la Tabla 3, obteniéndose que el modelo que mejor predice la UVR para todo el siguiente año es el Holt-Winters Multiplicativo; debido a que es el método con menor MAPE, RMSE y MAE. Se destaca que, en este caso, la red neuronal es el método con menor eficiencia en sus pronósticos, notándose predicciones estacionarias cuando la serie tiene una marcada tendencia creciente. Esto se debe posiblemente a que durante los últimos meses de cada año la serie tiende a volverse estacionaria y este es el patrón que finalmente predice la red.

Tabla 3. Medidas de capacidad de pronóstico a 12 meses para los modelos.

	<i>SARIMA</i>	<i>Regresión</i>	<i>HWA</i>	<i>HWM</i>	<i>Red Neuronal</i>
<i>MAPE</i>	1.1792	1.6391	1.3793	0.7983	2.3693
<i>RMSE</i>	4.0678	5.1371	4.9751	2.6926	7.8396
<i>MAE</i>	3.3682	4.6669	3.9442	2.2665	6.7602

Fuente: Elaboración propia programa R.

Al validar los supuestos distribucionales de los errores para el modelo SARIMA, se obtuvo que, para la prueba Shapiro-Wilk el p-valor (0.3809) es mayor al nivel de significancia del 5%, indicando que se cumple el supuesto de normalidad. El gráfico de residuales contra predichos no muestra desviación grave del supuesto de homocedasticidad y tanto la ACF, la PACF y la prueba Ljung-Box con 40 rezagos aplicados a los residuales indican incorrelación temporal, y bajo normalidad se concluye independencia en los errores del modelo. Para el modelo de regresión lineal se obtienen errores autocorrelacionados, heteroscedásticos y no normales según las pruebas, por lo que no es un modelo válido estadísticamente.

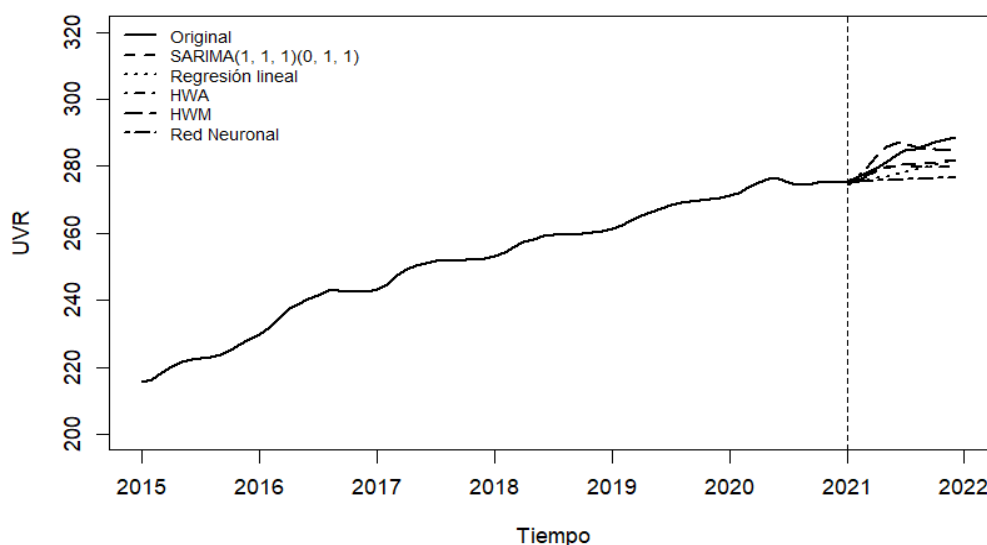


Figura 5. Gráfico comparativo de pronóstico para todo el año 2021 con todas las técnicas. Elaboración propia.

Se procedió a estimar los pronósticos para el escenario 2, dejando como datos para validación los últimos 6 meses, junio a diciembre de 2021. Se emplea primero el modelo SARIMA (1,1,1) (0,1,1), que también resultó ser el mejor en este caso. Luego es estimado el modelo de regresión lineal considerando tendencia lineal, para después ser aplicados los suavizamientos Holt-Winters aditivo y multiplicativo, y la red neuronal autorregresiva con los mismos órdenes que para el escenario 1. Las estimaciones realizadas para todos estos métodos se pueden ver en la Tabla 4.

Tabla 4. Estimaciones para los modelos en el escenario 2

Modelo SARIMA (1,1,1) (0,1,1)		
	Estimación	p-valor
ϕ_1	0.5806	0
θ_1	0.8016	0
Φ_1	-0.8922	0
Modelo de regresión lineal		
	Estimación	p-valor
Intercepto	105.3	0
Pendiente	0.669	0
R cuadrado	99.31%	-
Suavizamiento Holt-Winters		
	Aditivo	Multiplicativo
α	1	1
β	0.3375	1
γ	1	1

Fuente: Elaboración propia programa R.

Con cada modelo se realizaron pronósticos para los últimos 6 meses del año 2021. Dichos pronósticos y los valores reales para el segundo semestre 2021 de la UVR se muestran en la Figura 6 y las medidas de

capacidad de pronóstico en la Tabla 5. El modelo que mejor predice el período en cuestión es el SARIMA, pues tiene las menores medidas de capacidad de pronóstico. La prueba Shapiro-Wilk aplicada a los residuales del modelo seleccionado obtuvo un p-valor mayor al 1% y las pruebas de homocedasticidad e incorrelación validan el cumplimiento de los supuestos con el mismo nivel de significación, lo que convierte al modelo en válido estadísticamente. La regresión lineal realizada es la que peores pronósticos arroja, debido a que la tendencia global que asume está alejada de la que realmente está presente en serie para los últimos períodos observados, además este modelo obtiene errores autocorrelacionados, heteroscedásticos y no normales de acuerdo con la aplicación de las mismas pruebas de validación implementadas en el escenario 1.

Tabla 5. Medidas de capacidad de pronóstico a 6 meses para los modelos.

	<i>SARIMA</i>	<i>Regresión</i>	<i>HWA</i>	<i>HWM</i>	<i>Red Neuronal</i>
<i>MAPE</i>	0.2859	2.1703	0.4492	0.43519	0.9712
<i>RMSE</i>	0.9287	6.2319	1.3651	1.2496	3.0457
<i>MAE</i>	0.8178	6.2217	1.2895	1.3701	2.7896

Fuente: Elaboración propia programa R.

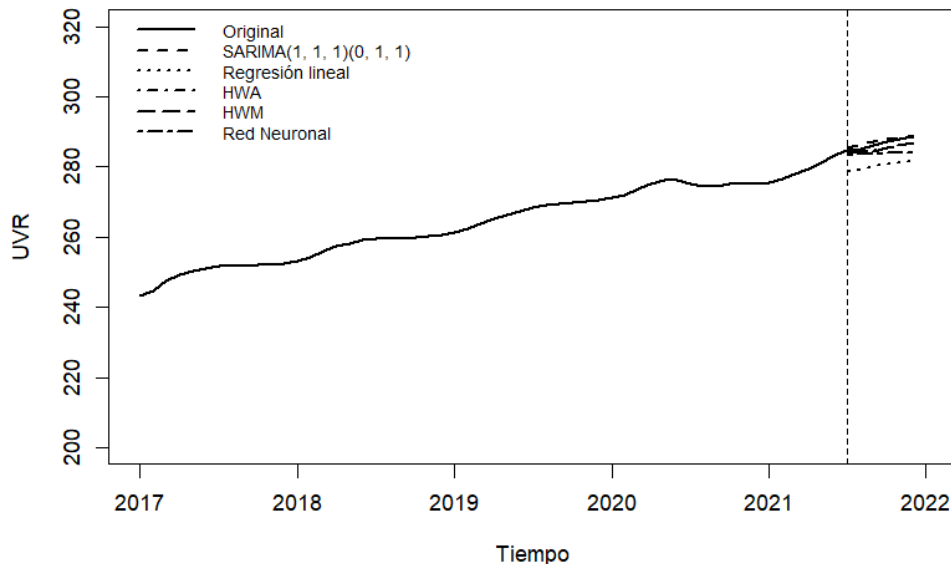


Figura 6. Gráfico comparativo de pronóstico para el segundo semestre de 2021 con todas las técnicas. Elaboración propia.

5. Conclusiones

En este estudio fueron aplicadas cinco técnicas de pronóstico: SARIMA, regresión lineal, suavizamiento Holt-Winters aditivo y multiplicativo y redes neuronales autorregresivas, con la finalidad de predecir la unidad de valor real (UVR) colombiana, usada para calcular los costos de los créditos de vivienda. De acuerdo con los métodos aplicados y considerando una ventana de observación de 264 datos mensuales de la UVR entre el 2000 y 2021, se obtuvo para el escenario 1 que la técnica que arrojó pronósticos más eficientes o cercanos a los valores reales de la UVR para el año 2021, fue el suavizamiento Holt-Winters multiplicativo. Esto debido a que logró captar la tendencia y estacionalidad histórica de los datos y proyectarla en los pronósticos futuros. Mientras en el escenario 2, pronósticos a 6 meses desde mitad de año, la técnica que obtuvo mejores pronósticos fue el modelo SARIMA, el cual parece predecir mejor la estructura de autocorrelación de la UVR cuando se está a mitad del ciclo estacional. Lo anterior, se debe a que la serie a final de año presenta un decaimiento en su tendencia, lo que hace que métodos como las redes

neuronales no logren arrojar pronósticos con la tendencia adecuada, mientras a mitad de año, los métodos logran proyectar mejor las componentes, siendo el Holt-Winter multiplicativo el más eficiente. Este estudio permite ver que dependiendo del momento del ciclo estacional en que se esté, la eficiencia de los pronósticos realizados por una técnica puede cambiar. Siendo así, este trabajo aporta a la discusión de la importancia de la selección adecuada de la técnica de pronóstico para una variable como la UVR, pues de ello puede depender la decisión de tomar o no un préstamo de vivienda en Colombia. Como trabajo futuro es posible realizar un análisis multivariado de series de tiempo, con el fin de estudiar si otras variables financieras y macroeconómicas pudieran ser significativas para predecir el comportamiento de la UVR colombiana.

Referencias Bibliográficas

- Akpinar, M., & Nejat, Y. (2016) Year ahead demand forecast of city natural gas using seasonal Time Series Methods. *Energies*. 9(9), 727.
- Al-Gounmeein, R. S. & Ismail, M. T. (2020) Forecasting the Exchange Rate of the Jordanian Dinar versus the US Dollar Using a Box-Jenkins Seasonal ARIMA Model. *International Journal of Mathematics & Computer Science*. 15(1) 27-40.
- Barría-Sandoval C, Ferreira G, Benz-Parra K, López-Flores P (2021) Prediction of confirmed cases of and deaths caused by COVID-19 in Chile through time series techniques: A comparative study. *PLoS ONE* 16(4): e0245414
- Bilgili, M., & Pinar, E. (2023). Gross electricity consumption forecasting using LSTM and SARIMA approaches: A case study of Türkiye. *Energy*, 128575.
- Bowerman, B., O'Connell, R. & Koehler, A. (2006) Pronósticos, series de Tiempo y Regresión. CENGAGE Learnig. 4ta Edición. (2006).
- Box, G., Jenkins, G. & Reinsel, G. (2008) Time series analysis forecasting and control, 4th Ed., Wiley & Sons Inc.
- Braimllari, A. & Sala, E. (2016) Modeling and Forecasting of Food Imports in Albania. *Albanian Journal of Agricultural Sciences*. 15(4), 200-205.
- Canova, F., & Hansen, B. E. (1995). Are seasonal patterns constant over time? A test for seasonal stability. *Journal of Business & Economic Statistics*, 13(3), 237-252.
- Chen, C., Chang, Y. & Chang, Y. (2009) Seasonal ARIMA forecasting of inbound air travel arrivals to Taiwan. *Transportmetrica*. 5, 125–140.
- Dickey, D. & Fuller, W. (1979) Distribution of the estimators for autoregressive time series with a unit root. *Journal of American Statistical Association* 74(366), 427-431.
- Diebold, F. (1999) Elementos de Pronóstico. International Thomson Editores, S.A. de C.V.
- Gil, V. D. (2016). Pronóstico de la demanda mensual de electricidad con series de tiempo. *Revista EIA*. 13(26), 111-120.
- Goh, C. & Law, R. (2001) Modeling and forecasting tourism demand for arrivals with stochastic nonstationary seasonality and intervention. *Tourism Management*. 23(5), 499–510.
- Hankey, J. & Wichern, D. (2010) Pronósticos en los Negocios. Pearson. 9na edición. Ed. Pearson.
- Hylleberg, S., Engle, R. F., Granger, C. W., & Yoo, B. S. (1990). Seasonal integration and cointegration. *Journal of econometrics*, 44(1-2), 215-238.
- Hyndman, R. & Athanasopoulos, G. (2021) Forecasting: principles and practice, 3rd edition, OTexts: Melbourne, Australia. Web: OTexts.com/fpp3.
- Kumar, M. & Anand, M. (2015) An application of time series ARIMA forecasting model for predicting sugarcane production in India. *Studies in Business and Economics*. 9(1), 81-95.
- Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?. *Journal of econometrics*, 54(1-3), 159-178.

- Shapiro, S. & Wilk, M. (1965) An analysis of variance test for normality (complete samples). *Biometrika*, 52(3), 591–611.
- Stellwagen, E. & Tashman, L. (2013) ARIMA: The models of Box & Jenkins. *International Journal of Applied Forecasting*, 30, 28-33.
- Tsitsika, E. C., Maravelias, C. and Haralabous, V J. (2007) Modeling and forecasting pelagic fish production using univariate and multivariate ARIMA models. *Fisheries Science*, 73, 979-988.
- Velásquez J. & Aguilar, S. (2011). Un análisis de la dinámica de largo plazo de la UVR. *Revista Ingenierías Universidad de Medellín*, 10, 97-106.
- Villazón-Bustillos, D., Rubio-Arias, H., Ortega-Gutiérrez, J., Rentería-Villalobos, M., González-Gurrola L. & Pinales-Munguia A. (2016) Análisis en series de tiempo para el pronóstico de sequía en la región noroeste del estado de Chihuahua. *Ecosistemas y recursos agropecuarios*, 3(9), 307-315.
- Yamacli, D. S., & Yamacli, S. (2023). Estimation of the unemployment rate in Turkey: A comparison of the ARIMA and machine learning models including Covid-19 pandemic periods. *Heliyon*, 9(1) 1-10.
- Youness, J., & Driss, M. (2022). An ARIMA model for modeling and forecasting the dynamic of univariate time series: the case of moroccan inflation rate. In *2022 International Conference on Intelligent Systems and Computer Vision (ISCV)* (pp. 1-5). IEEE.
- Youness J., & Driss, M. (2023). Comparing The Forecasting Accuracy Metrics of Support Vector Regression and ARIMA Algorithms for Non-Stationary Time Process. *Mathematics and Statistics*, 11(2), 294 - 299.