

ANEXO 02: FORMATO DE APLICACIÓN DE FERIA DE PROYECTOS FIEECS 2019-I

COMPARACIÓN DE MODELOS DE PRONÓSTICO ARIMA Y ARIMAX, PARA RECONOCER EL IMPACTO EN LAS EXPORTACIONES DEL PERÚ, ANTE LA ACTUAL “GUERRA COMERCIAL EEUU-CHINA”

Integrantes: Mendoza Carhuapuma Renso Emerson, Meza Barzena Daniel Gabriel

Docente: Alipio Ordoñez Mercado

Asignatura: Series de Tiempo

Escuela Profesional de Ingeniería Estadística

Facultad de Ingeniería Económica, Estadística y CCSS

Universidad Nacional de Ingeniería

RESUMEN

Un elemento fundamental en la economía de peruana, es tener conocimiento a corto, mediano y largo plazo, el comportamiento de las exportaciones del Perú. En concordancia a este caso, se realizó una evaluación de los niveles de exportaciones, en forma trimestral desde inicios del 2011 hasta la actualidad. Con el objetivo de desarrollar un modelo que permita caracterizar y obtener pronósticos sobre el comportamiento de las exportaciones realizadas en el país; para ello, se usó la metodología Box Jenkins, siguiendo las fases para los modelos ARIMA y los modelos ARIMAX, para ser comparados hasta la elección del mejor modelo de pronósticos de las exportaciones. Los datos fueron tomados de la página web de la SUNAT-PromPeru. La serie bajo estudio mostró un comportamiento ligeramente estacional, y trimestralmente tiene una tendencia moderada a decrecer. Finalmente se halló el mejor, tanto en estimación como en pronóstico, es el modelo ARIMAX(0,1,1) y se obtuvieron los pronósticos entre los años 2019 y 2021, siguiendo un comportamiento estable respecto al periodo de validación de la muestra de la serie. Los datos fueron analizados usando el lenguaje y software RStudio.

Palabras Clave: Metodología Box Jenkins, Modelos ARIMA, Modelos ARIMAX, Exportaciones del Perú

ABSTRACT

A fundamental element in the Peruvian economy, is to have short, medium and long term knowledge, the behavior of Peruvian exports. In accordance with this case, an evaluation of export levels was conducted, quarterly from the beginning of 2011 to the present. With the objective of developing a model that allows to characterize and obtain forecasts about the behavior of exports made in the country; or this, the Box Jenkins method was used, following the phases for the ARIMA models and the ARIMAX models,

to be compared to the choice of the best export forecast model. The data was taken from the SUNAT-PromPeru website. The series under study showed a slightly seasonal behavior, and quarterly it has a moderate tendency to decrease. Finally, the best one was found, both in estimation and forecast, is the ARIMAX model (0,1,1) and the forecasts were obtained between 2019 and 2021, following a stable behavior with respect to the validation period of the sample of the series. The data was analyzed using the RStudio language and software.

Keywords: Box-Jenkins method, ARIMA model, ARIMAX model, Exports of Peru

INTRODUCCIÓN

En este trabajo de investigación se trata de estudiar, aplicar y comparar los resultados encontrados en las técnicas de estimación de pronósticos como son: el Modelo ARIMA y ARIMAX para medir el impacto que trae la Guerra Comercial EEUU-China a las exportaciones del Perú. Estas metodologías han sido usadas en muchos estudios en el exterior como la “Comparación entre un sistema neuro difuso auto organizado y un modelo ARIMAX en la predicción de series económicas volátiles” (Universidad de la Rioja, José A. Avellaneda G., Cynthia M. Ochoa R., Juan Carlos Figueroa García, España) y “Modelo cuantitativo ARIMAX- EGARCH para la predicción de la tasa de cambio colombiana” (Universidad Nacional de Colombia, Mario Alberto Martínez Orozco, Diana Sirley Guzmán Aguilar, Fredy Ocaris Pérez Ramírez, Nini Johana Marín Rodríguez, Colombia), donde se determinó de forma precisa el comportamiento futuro de las variables.

Por lo tanto, es oportuno el uso de estas técnicas en la estimación de pronósticos para las exportaciones peruanas futuras. En el trabajo se realizó el análisis ARIMA y ARIMAX, para su posterior comparación y elección del mejor modelo según sus indicadores. También se realizó el pronóstico de las exportaciones de los años 2019, 2020 y 2021.

PRESENTACIÓN DEL PROBLEMA

El Perú es parte del Bloque Económico Alianza Pacífico, que tiene como objetivo impulsar el crecimiento, desarrollo y competitividad de las economías que la integran, sin embargo, la actual guerra económica entre EEUU y China ha generado la disminución del volumen de exportación, es por esto que el presente trabajo de vital importancia contar con modelos para la evaluación del pronóstico de las exportaciones que nos proporcionen predicciones confiables de las variaciones, con una alta precisión, es por ello que analizaremos la serie temporal de exportaciones del Perú desde el año 2011 hasta la actualidad para aplicar un modelo ARIMA y ARIMAX e identificar cual es el más preciso en la etapa de estimación y la etapa de pronósticos, presentar los valores estimados de las exportaciones para los próximos 3 años.



Gráfico 1: Serie trimestral de las exportaciones peruanas al bloque Alianza Pacífico en valor FOB (millones de US\$).

OBJETIVOS

Objetivo General:

- Determinar el mejor modelo entre el ARIMA y ARIMAX, para la medición del impacto en las exportaciones del Perú al bloque comercial Alianza Pacífico ante la actual Guerra Comercial EEUU – China.

Objetivos Específicos:

- Determinar el valor de las exportaciones del Perú en los próximos 3 años.
- Determinar el mejor modelo en la etapa de Estimación y Pronóstico.

DESCRIPCIÓN DE LA SOLUCIÓN

Un modelo "Autorregresivo Integrado de Media Móvil" (ARIMA) sigue la metodología Box-Jenkins así como un modelo "Autorregresivo Integrado de Media Móvil con variables exógenas (ARIMAX). Ésta metodología es la más usada para encontrar un mejor ajuste para el análisis de una serie temporal de valores, a fin de que los pronósticos sean más acertados.

El motivo de la introducción de los modelos ARIMA y ARIMAX nace del hecho de que no se puede trabajar con una serie temporal no estacionaria. Se dice que una serie es estacionaria cuando su media, varianza y autocovarianza son invariantes en el tiempo. La mayoría de series temporales económicas no son estacionarias pero diferenciándolas un número determinado de veces la serie original se transforma en estacionaria, con lo cual ya se podría aplicar la metodología de Box-Jenkins a los modelos ARIMA y ARIMAX.

La estacionariedad es importante porque, si la serie es no estacionario, todos los resultados típicos del análisis de regresión clásica no son válidos, y para el caso los modelos ARIMAX son imprescindibles ya que incluirán dos variables exógenas en este estudio.

Los pronósticos a largo plazo de una serie estacionaria convergerán a la media incondicional de la serie.

Primera etapa: Análisis Exploratorio, lectura y preparación de los datos

La muestra de observaciones para el estudio de a realizarse, comprende el periodo del primer trimestre del 2011 hasta el cuarto trimestre del 2018, para el caso de validación comprende el mismo periodo.

Con la muestra de estimación el objetivo es proponer un modelo o lista de modelos y a partir de ello escoger el mejor modelo con mayor precisión en el periodo de pronóstico de los siguientes 3 años.

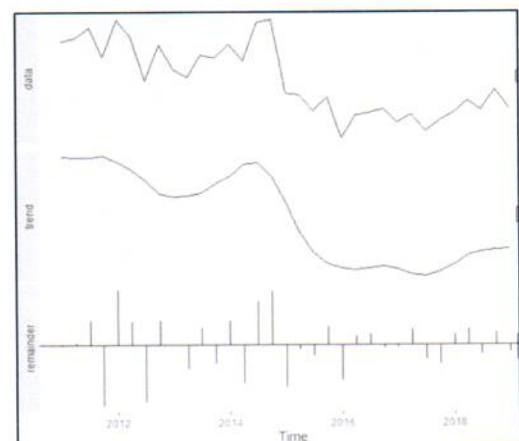


Gráfico 2: Descomposición de la serie original, su tendencia y residual aplicada a los datos trimestrales de las exportaciones peruanas al bloque Alianza Pacífico (2011-2018).

Fuente: Elaboración propia – Rstudio.

Segunda Etapa: Identificación del modelo

En esta etapa se procederá a realizar la prueba aumentada de Dickey-Fuller (ADF), para ver si nuestra serie temporal es estacionaria.

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \dots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t$$

Dónde: α es una constante, β es el coeficiente sobre una tendencia temporal, p el orden de retraso del proceso autorregresivo.

H_0 : La serie no es estacionaria
 H_1 : La serie es estacionaria

Si el p-valor es mayor a 0.05 se rechaza la H_0 .

Continuamos después con las autocorrelaciones donde las ACF (función de autocorrelación simple) proporcionan información sobre cómo una observación influye en las siguientes y las PACF (función de autocorrelación parcial) proporcionan la relación directa existente entre observaciones separadas por k retardos.

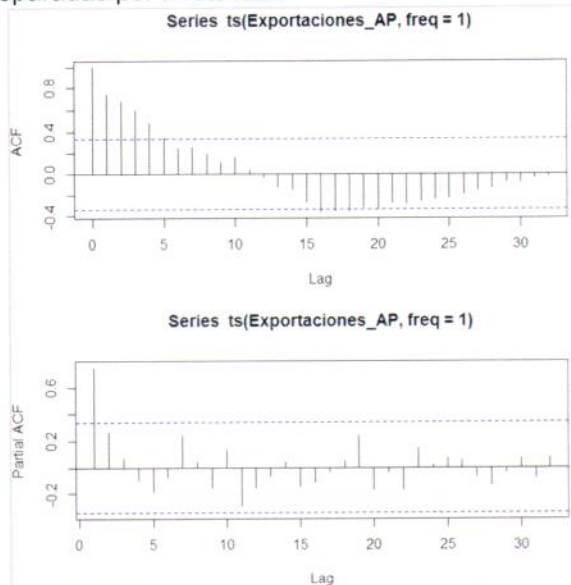


Gráfico 3: Función de Autocorrelación Simple y Función de Autocorrelación Parcial de la serie original.
 Fuente: Elaboración propia - RStudio.

Para realizar un modelo ARIMA y ARIMAX la serie temporal debe ser estacionaria. Para conseguir esta estacionariedad, la diferenciaremos.

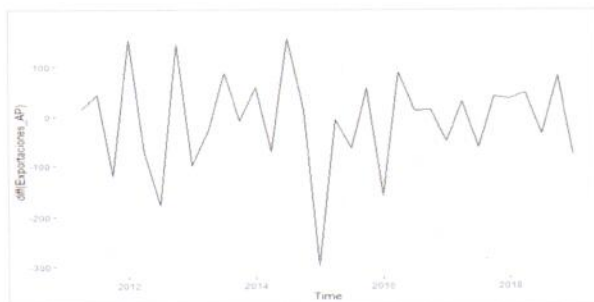


Gráfico 4: Primera diferencia de la serie original de Exportaciones_AP.
 Fuente: Elaboración propia - RStudio.

Para comprobar que la serie es, efectivamente, estacionaria, hacemos de nuevo el test aumentado de Dickey-Fuller.

El p-valor obtenido nos indicará si rechazamos H_0 , lo cual serie temporal es estacionaria.

H_0 : La serie no es estacionaria
 H_1 : La serie es estacionaria

El p-valor calculado es 0.5149518 para la serie diferenciada por lo tanto es mayor a 0.05, entonces se rechaza la H_0 y decimos que la serie es estacionaria.

Al trazar la serie diferenciada, vemos un patrón oscilante alrededor de 0, sin una tendencia fuerte visible. Esto sugiere que la diferenciación de los términos del orden 1 es suficiente y debe incluirse en el modelo.

A continuación, los picos en rezagos particulares de la serie diferenciada pueden ayudar a informar la elección de p o q para nuestro modelo:

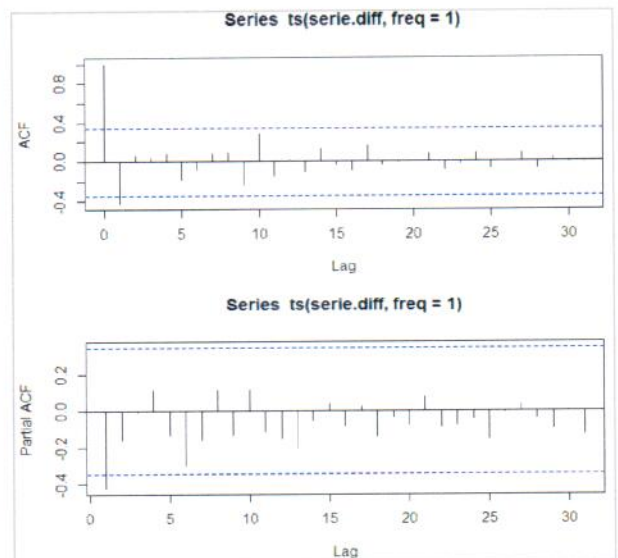


Gráfico 5: Función de Autocorrelación Simple y Función de Autocorrelación Parcial de la serie diferenciada.

Tercera Etapa: Estimación del modelo

Se llevará a cabo la elaboración de los modelos ARIMA que es la abreviatura de "AutoRegressive Integrated Moving Average", consiste en un modelo estadístico que utiliza variaciones y regresiones de datos estadísticos con el fin de encontrar patrones para una predicción hacia el futuro, es decir, las estimaciones futuras vienen explicadas por los datos del pasado y no por variables independientes.

$$Y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q} + e_t$$

$$\phi_p(B)(1-B)^d X_t = \theta_q(B)e_t$$

Dónde: y_d es Y diferenciado, d veces c es una constante
 Componentes: Autoregressive (AR(p)), Integrated ($I(d)$), Moving Average (MA(q)), p es el parámetro autorregresivo y especifica el número de retardos usados en el modelo, el orden q determina el número de términos a incluir en el modelo.

Por otro lado un modelo ARIMA que incluye como variable explicativa otra serie temporal distinta a la que queremos modelar, el modelo recibe el nombre de ARIMA con variable exógena o modelo de función de transferencia, ARIMAX. Con este tipo de modelo, las variables que usaremos para la construcción del mejor modelo serán los términos de Intercambio (TI) y el Tipo de Cambio Nominal (TCN) en función de sus valores pasados y de los valores presentes de éstas variables explicativas que se incluirán.

Supongamos dos series de tiempo denotados por Y_t y X_t , a su vez que ambas sean estacionarias. Entonces el modelo de función de transferencia ARIMAX se puede escribir de la siguiente manera:

$$Y_t = C + v(B)X_t + N_t$$

Dónde:

Y_t : es la serie de salida (variable dependiente).

X_t : es la serie de entrada (variable independiente).

C : es el término constante.

N_t : es la perturbación estocástica, es decir, la serie de ruido del sistema que es independiente de la serie de entrada.

$v(B)X_t$: es la función de transferencia (o la función de respuesta de impulso), que permite X influenciar Y a través de un retardo distribuido.

Por ello se propone un modelo en función de la relación de las exportaciones peruanas hacia la región Alianza Pacífico con otros indicadores, donde se evidencia como existe una correlación positiva considerable entre el las exportaciones peruanas hacia la región de Alianza Pacífico y los Términos de Intercambio, o las exportaciones y el Tipo de Cambio Real (correlación baja), por otra parte existe una correlación negativa (Pero muy fuerte más que en los casos anteriores) entre las exportaciones y el Tipo de Cambio Nominal. Para identificar un modelo con estructura de este tipo, se diseñan modelos con diferentes valores para los parámetros, determinándose posteriormente el más adecuado por medio de algún criterio, como el AIC o BIC.

ESTIMACIÓN	
MODELO	ESTIMACIÓN CON UNA VARIABLE REGRESORA
	MODELO
ARIMA(0,1,0)	ARIMAX(0,1,0)
	ARIMAX(0,1,1)
ARIMA(0,1,1)	
	ESTIMACIÓN CON DOS VARIABLES REGRESORAS
	MODELO
ARIMA(1,1,0)	ARIMAX(0,1,0)
ARIMA(1,1,1)	ARIMAX(0,1,1)
ARIMA(1,0,0)	

Gráfico 6: Comparativo de los modelos ARIMA y ARIMAX propuestos para la etapa de estimación.

Fuente: Elaboración propia.

Cuarta Etapa: Diagnóstico del modelo

Si los parámetros de orden del modelo y la estructura se especifican correctamente, no esperaríamos que existieran autocorrelaciones significativas.

Hemos ajustado un modelo que puede producir un pronóstico, pero comenzamos examinando los gráficos ACF y PACF para los residuos del modelo. Si los parámetros de orden del modelo y la estructura se especifican correctamente, no esperaríamos que existieran autocorrelaciones significativas.

Para el caso de los residuales tanto individual como el global resulto ser: aleatorio. Donde se usó la prueba de Ljung - Box para contrastar dicho supuesto y se puede definir de la siguiente manera:

H_0 : Los datos se distribuyen de forma independiente (es decir, las correlaciones en la población de la que se toma la muestra son 0, de modo que cualquier correlación observada en los datos es el resultado de la aleatoriedad del proceso de muestreo).

H_1 : Los datos no se distribuyen de forma independiente.

Donde el estadístico de Prueba es:

$$Q = n(n+2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n-k}$$

Donde n es el tamaño de la muestra, $\hat{\rho}_k$ es la autocorrelación de la muestra en el retraso k y h es el número de retardos que se están probando. Por nivel de significación α , la región crítica para el rechazo de la hipótesis de aleatoriedad es:

$$Q > \chi_{1-\alpha, h}^2$$

Donde $\chi_{1-\alpha, h}^2$ es la α -cuantil de la distribución chi-cuadrado con h grados de libertad.

VALIDACIÓN		CON UNA VARIABLE REGRESORA:	
MODELO	Correlación Prueba Ljung-Box	MODELO	Correlación Prueba Ljung-Box
ARIMA(0,1,0)	0.0193	ARIMAX(0,1,0)	0.2440
ARIMA(0,1,1)	0.7316	ARIMAX(0,1,1)	0.9622
ARIMA(1,1,0)	0.5339		
ARIMA(1,1,1)	0.8207	CON DOS VARIABLES REGRESORAS:	
ARIMA(1,0,0)	0.1728	MODELO	Correlación Prueba Ljung-Box
		ARIMAX(0,1,0)	0.7824
		ARIMAX(0,1,1)	0.9872

Gráfico 7: Comparativo del p-valor en la validación de los modelos estimados. Fuente: Elaboración propia.

Quinta Etapa: Pronóstico

Se usarán los siguientes Indicadores para comparar los modelos estimados anteriormente, donde estos indicadores sirven para comparar la efectividad de diferentes modelos utilizados en Pronóstico. Siempre se busca el valor menor en los indicadores MAPE (Porcentaje promedio absoluto del error), MAE (Promedio absoluto de los errores) y MSE (Desviación cuadrática media de los errores anormales de pronóstico) ya que esto representan un mejor ajuste del modelo.

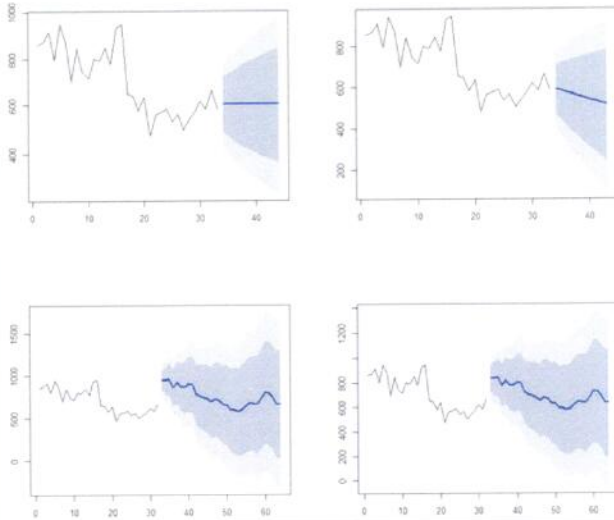


Gráfico 8: Comparativo de los modelos en pronósticos Arima y el mejor Arima así como el modelo Arimax y el mejor modelo Arimax, con variable dependiente EXPO_AP y variable exógena TI y TCN. Fuente: Elaboración propia - RStudio

RESULTADOS

Se procedió a realizar las pruebas de los modelos ARIMA propuestos para la estimación en R-studio; la tabla siguiente muestra los resultados de las pruebas de estos modelos, con sus indicadores de AIC (Criterio de Información de Akaike) y de la desviación estándar de los residuales en los (5) modelos probados.

MODELO	ESTIMACIÓN 2011 - 2018		Validación
	AIC	$\hat{\sigma}_\epsilon$	
ARIMA(0,1,0)	368.08	97.7343	individual y general no son aleatorios
ARIMA(0,1,1)	382.3	89.02809	individual y general son aleatorios
ARIMA(1,1,0)	383.11	90.21641	individual y general son aleatorios
ARIMA(1,1,1)	384.22	88.91007	individual y general son aleatorios
ARIMA(1,0,0)	398.85	91.82592	individual y general son aleatorios

Gráfico 9: Comparativo de los modelos ARIMA en estimación y la selección del mejor modelo ARIMA en cuanto a sus indicadores AIC y desviación estándar residual. Fuente: Elaboración propia - RStudio

A partir de lo anterior, el mejor modelo en estimación es el **ARIMA (0,1,1)**, según **AIC (Criterio de Información de Akaike)**, así mismo de la validación individual y general de que los residuales sean aleatorios.

MODELO	PRONÓSTICO 2019 - 2021		
	RMSE	MAE	MAPE
ARIMA(0,1,0)	96.241	72.8408	10.6703
ARIMA(0,1,1)	87.6716	65.8192	9.6857
ARIMA(1,1,0)	88.8405	66.4658	9.7598
ARIMA(1,1,1)	87.5531	65.6937	9.6693
ARIMA(1,0,0)	91.8285	71.236	10.2652

Gráfico 10: Comparativo de todos los modelos ARIMA en pronóstico para los próximos 3 años. Fuente: Elaboración Propia - RStudio

Por el lado del pronóstico, el mejor modelo resulto ser el **ARIMA (1,1,1)**, según los indicadores de: RMSE, MAE, MAPE, para los próximos 3 años.

Mientras que por otro lado tenemos la prueba de los modelos ARIMAX propuestos para la estimación en R-studio, las tablas siguientes mostraran los resultados de las pruebas con sus indicadores de AIC, AICc, BIC y de la desviación estándar de los residuales en los (4) modelos probados.

Modelos de Estimación ARIMAX con 1 variable exógena

MODELO	ESTIMACIÓN 2011 - 2018				Validación
	AIC	AICc	BIC	$\hat{\sigma}_\epsilon$	
ARIMAX(0,1,0)	368.94	369.37	371.81	88.555	individual y general son aleatorios
ARIMAX(0,1,1)	368.99	369.88	373.29	87.1378	individual y general son aleatorios

Gráfico 10: Resumen de todos los indicadores de los modelos ARIMAX estimados y su respectiva validación. Fuente: Elaboración Propia - RStudio

Modelos de Estimación ARIMAX con 2 variables exógenas

MODELO	ESTIMACIÓN 2011 - 2018				Validación
	AIC	AICc	BIC	$\hat{\sigma}_\epsilon$	
ARIMAX(0,1,1)	381.93	383.41	387.79	87.37	individual y general son aleatorios
ARIMAX(1,0,0)	389.33	390.76	395.32	81.47	individual y general son aleatorios

Gráfico 11: Resumen de todos los indicadores de los modelos ARIMAX estimados y su respectiva validación. Fuente: Elaboración Propia - Rstudio

A partir de los cuadros de resumen de los indicadores de estimación tenemos que los modelos ARIMAX son mucho mejores que los modelos ARIMA evaluados, a su vez el modelo **ARIMAX (0,1,1)** con una variable exógena posee mejores indicadores frente a un modelo ARIMAX con dos variables exógenas en cuanto a estimación.

Por el lado del pronóstico, el mejor modelo resultó ser el **ARIMAX (0,1,1)** con una variable exógena, según los indicadores de: RMSE, MAE, MAPE, para los próximos 3 años.

MODELO	PRONÓSTICO 2019 - 2021		
	RMSE	MAE	MAPE
ARIMAX(0,1,1)	82.953	64.6316	9.617
ARIMAX(0,1,0)	85.743	66.172	9.956

Gráfico 4: Resumen de los indicadores para los modelos de Pronóstico Arimax con una sola variable exógena.

Fuente: Elaboración propia – Rstudio

En cambio, el mejor modelo con dos variables exógenas resultó ser el **ARIMAX (1,0,0)**, según los indicadores de: RMSE, MAE, MAPE, para los próximos 3 años.

MODELO	PRONÓSTICO 2019 - 2021		
	RMSE	MAE	MAPE
ARIMAX(0,1,1)	82.907	63.798	9.420
ARIMAX(1,0,0)	77.681	60.052	8.771

Gráfico 4: Resumen de los indicadores para los modelos de Pronóstico Arimax con dos variables exógenas.

Fuente: Elaboración propia - RStudio

CONCLUSIONES

En este trabajo, hemos construido (5) modelos ARIMA y (4) modelos de función de transferencia ARIMAX para el pronóstico de las exportaciones peruanas a la integración regional Alianza Pacífico medido en valor FOB en millones de dólares US\$ en los próximos 3 años como una serie de salida y los Términos de Intercambio además del Tipo de Cambio Nominal como series de entradas.

El mejor modelo en cuanto a pronóstico encontrado fue el **ARIMAX (1,0,0)** con dos variables exógenas fue suficiente y los residuos eran ruido blanco.

El estadístico R-cuadrado indicó que el modelo equipado explica 67,03% de la variabilidad en de las exportaciones peruanas a la integración regional Alianza Pacífico medido en valor FOB en millones de dólares US\$ para los próximos 3 años. Para ello se calcularon el valor de las exportaciones trimestrales para los años 2019 (US\$ 3 282 millones), 2020 (US\$ 3 173 millones) y 2021 (US\$ 2 910 millones) con el mejor modelo que fue ARIMAX con 2 variables regresoras (en cuanto a pronóstico), mediante el cual se compararon con el valor de las exportaciones calculadas por los mejores modelos ARIMA para las exportaciones peruanas a la integración regional Alianza Pacífico medido en valor FOB en millones de dólares US\$ y se pudo ver que el valor de éstas eran ligeramente diferentes y ambos asumieron caída de las exportaciones en los siguientes 3 años.

BIBLIOGRAFÍA

- [1] **Box, G. E. P. – Jenkins, G. M. – Reinsel, G. C.:** Time Series Analysis: Forecasting and Control. John Wiley & Sons Inc., New York, 2008.
- [2] **Box, G. E. P. – Tiao, G. C.:** Intervention analysis with applications to economic and environmental problems. Journal of the American Statistical Association. 1975, vol. 70, no. 349, p. 70-79.
- [3] **Marek, L.:** Transfer function models. Acta Oeconomica Pragnesia. 2000, vol. 8, no. 3, p. 83-94
- [4] **Pankratz, A.:** Forecasting with Dynamic Regression Models. Wiley-Interscience, 1991.
- [5] **Rublíková, E. – Marek, L.:** Linear transfer function model for outflow rates. Ekonomickérozhľad. 2001, vol. 30, no. 4, p. 457-466

ANEXOS

Modelos ARIMA - INEI

https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitaes/Lib0516/Libro.pdf

Análisis de Series de Tiempo

https://www.cepal.org/sites/default/files/courses/files/01_2_arima.pdf

Revista Espacios: Modelo cuantitativo ARIMAX – EGARCH para la predicción de la tasa de cambio colombiana (COP/USD).

<https://www.revistaespacios.com/a18v39n07/a18v39n07p16.pdf>

Análisis de series temporales: Modelos ARIMA

<https://addi.ehu.es/bitstream/handle/10810/12492/04-09gon.pdf;jsessionid=33EAF9A61207F3979821636268CAF80A?sequence=1>

ARIMA vs. ARIMAX – which approach is better to analyze and forecast macroeconomic time series?
http://mme2012.opf.slu.cz/proceedings/pdf/024_Durka.pdf

What's the difference between ARMA, ARIMA, and ARIMAX, in layman's terms? What exactly do P, D, Q mean and how do you know what to put in for them in say R (1,0,2) or (2,1,1)?
<https://www.quora.com/Whats-the-difference-between-ARMA-ARIMA-and-ARIMAX-in-laymans-terms-What-exactly-do-P-D-Q-mean-and-how-do-you-know-what-to-put-in-for-them-in-say-R-1-0-2-or-2-1-1>