



## Análisis empírico del índice de precios al consumidor de Lima Metropolitana basado en el Modelo Sarima

Empirical analysis of the consumer price index of Metropolitan Lima based on the Sarima Model

Análise empírica do índice de preços ao consumidor da região metropolitana de Lima com base no modelo Sarima

### ARTÍCULO GENERAL

Jaime Yelsin Rosales Malpartida

[jrosalesm@uni.pe](mailto:jrosalesm@uni.pe)

<https://orcid.org/0000-0003-4574-5172>

Universidad Nacional de Ingeniería, Lima, Perú

Recibido 02 de Noviembre 2021 | Arbitrado y aceptado 02 de Noviembre 2021 | Publicado el 16 de Diciembre 2021

#### RESUMEN

Esta investigación selecciona los datos del Índice de Precios al Consumidor (IPC) de la ciudad de Lima Metropolitana de enero 2012 a diciembre 2020 como muestra y establece el modelo de IPC SARIMA(2,1,0)(1,0,2)<sub>12</sub> para el análisis empírico. Se utilizó el software libre R para los análisis. Los resultados muestran que el modelo es exacto. Se pronostican los datos del IPC en 2020, y el valor pronosticado está muy cerca del valor real. Finalmente se utiliza el modelo para hacer los pronósticos del IPC en el primer semestre de 2021.

**Palabras clave:** IPC, SARIMA, Lima Metropolitana.

#### ABSTRACT

This research selects the data from the Consumer Price Index (CPI) of the city of Metropolitan Lima from January 2012 to December 2020 as a sample and establishes the SARIMA IPC model (2,1,0)(1,0,2)<sub>12</sub> for empirical analysis. Free R software was used for the analyzes. The results show that the model is accurate. The CPI data is forecast in 2020, and the predicted value is very close to the actual value. Finally, the model is used to make the CPI forecasts in the first half of 2021.

**Keywords:** IPC, SARIMA, Metropolitan Lima.

#### RESUMO

Esta pesquisa seleciona os dados do Índice de Preços ao Consumidor (IPC) da cidade de Lima Metropolitana de janeiro de 2012 a dezembro de 2020 como amostra e estabelece o modelo SARIMA IPC (2,1,0)(1,0,2)<sub>12</sub> para análise empírica. O software R livre foi usado para as análises. Os resultados mostram que o modelo é preciso. Os dados do CPI são previstos em 2020, e o valor previsto está muito próximo do valor real. Por fim, o modelo é utilizado para fazer as projeções do IPC no primeiro semestre de 2021.

**Palavras-chave:** IPC, SARIMA, Lima metropolitana.

## Introducción

El índice de precios al consumidor (IPC) es un índice estadístico económico muy importante, que refleja el nivel de precios y el nivel de vida de los residentes en un período determinado. Lima Metropolitana es un área metropolitana peruana conformada por cincuenta distritos, donde cuarenta y tres distritos corresponde a la Provincia de Lima y siete distritos a la Provincia constitucional del Callao. Es el área metropolitana más grande, extensa y poblada del Perú, por lo tanto, vale la pena estudiar los problemas económicos de Lima Metropolitana, como el IPC.

En 1970, Box y Jenkins propusieron una metodología rigurosa para la identificación, estimación, validación y pronóstico de modelos dinámicos para datos de series temporales, como los Modelos Autorregresivos Integrados de Promedio Móvil (ARIMA) y los Modelos Autorregresivos Integrados de Promedio Móvil Estacionales (SARIMA). Con esta metodología general, se ha desarrollado en la literatura la metodología de pronósticos para series temporales. Muchas investigaciones realizan análisis empírico de la serie del Índice de Precios al Consumidor en diversos contextos han sido desarrollados por autores. Cerna et al. (2021) [1] analiza el índice de precios al consumidor de Lima Metropolitana según el modelo SARIMA(0,1,1)(1,0,1)<sub>12</sub> con variable dummy. Basándose en los datos mensuales Lima Metropolitana desde enero de 2010 a diciembre de 2020 obtenida de la web del Banco Central de Reserva del Perú. Siesquen y Riojas (2017) [2] tuvieron como objetivo determinar los modelos de pronóstico para el Índice de Precios al Consumidor tanto de la Región Lambayeque como Nivel Nacional, basados en sus series históricas de índices registrados por el INEI, durante el periodo enero 2010 a diciembre 2016. Los modelos estimados, adecuados y óptimos para la serie Índice de Precios al Consumidor Región Lambayeque para el periodo enero 2010 a diciembre del 2016 fue un ARIMA(0, 1, 1); para el caso de la serie Índice de Precios al Consumidor a Nivel Nacional para el periodo enero del 2010 a diciembre 2016 fue un ARIMA(0, 1, 1). Djami et al. (2021) [3] establece el modelo ARIMA (1,1,1) utilizado para pronosticar el IPC en la ciudad de Ambon de Indonesia para el período de enero de 2007 a diciembre de 2020. Xingxiang Qi (2021) [4] Selecciona los datos del IPC de la ciudad de Jinan de China de 2011 y establece el modelo ARIMA (11, 1, 12) de índice de precios al consumidor para el análisis empírico. Los resultados mostraron que el modelo es exacto y se pronostican los datos del IPC en 2020 donde el valor pronosticado está muy cerca del valor real. LI Shu, WANG Fei y

WANG Feng-xiao (2020) [5] toman como muestra los datos mensuales del IPC de Qingdao desde enero de 2010 hasta junio de 2019, y establecen el modelo ARIMA (11,1,12). CHU Rui (2018) [6] selecciona los datos mensuales del IPC de enero de 2005 a junio de 2017 en Beijing como muestra, y se construyó el modelo ARIMA (0,1,12). Li Xuan y Huang Dongdong [7] Con base en los datos mensuales del IPC de enero de 2000 a diciembre de 2012, construyeron un modelo ARIMA para analizar y pronosticar el IPC en la primera mitad de 2013. Los resultados empíricos muestran que el modelo ARIMA.

(3,1,3) puede muy bien describir el IPC y proporcionar una mejor predicción. Zhang Yangyang, Deng Wei y Song Changyu (2020) [8] utilizan el modelo ARIMA (2,1,2) para ajustar los datos del IPC de la provincia de Shandong de enero de 1995 a mayo de 2020 y predecir el IPC de la provincia de Shandong de junio a octubre de 2020. Zhou Mei-ying (2011) [9] establece el modelo ARIMA (10,1,3) y analiza la serie temporal del IPC de la provincia de Hubei. GUO Xiao-feng (2012) [10] selecciona los datos mensuales del IPC de China desde enero de 2001 a octubre de 2011 como muestras y establece el modelo ARIMA (12,1,20). CUI Wen-yan y XU Feng-hua (2016) [11] establecen el modelo ARIMA (0,1,2) sobre los datos anuales del IPC de la provincia de Shandong durante 1995-2011. En las referencias de [1] - [11], aunque los parámetros del modelo son distintos, todos usan ARIMA o SARIMA para ajustar y predecir los datos del IPC en diferentes períodos. En esta investigación se selecciona los datos del IPC de la ciudad de Lima Metropolitana de enero 2012 a diciembre 2020 como muestra y utiliza el modelo SARIMA(2,1,0) (1,0,2)<sub>12</sub> para ajustar y predecir el primer semestre de 2021.

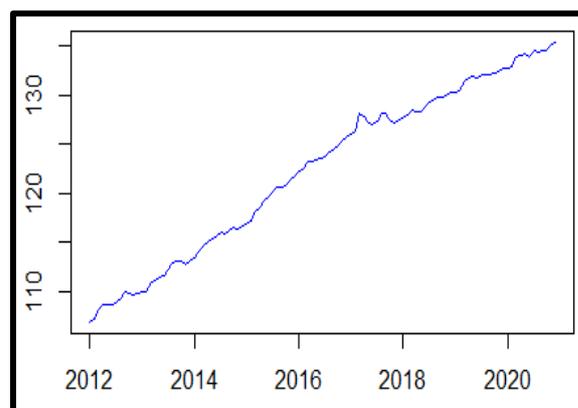


Figura 1. Índice de precios al consumidor de Lima metropolitana, enero 2012 - diciembre 2020

## Material y Métodos

En la presente investigación de series temporales la población de interés está constituida por la serie histórica del Índice Mensual de Precios al Consumidor de Lima Metropolitana (año base 2012=100). La muestra comprendió desde el periodo enero 2012 a diciembre 2020, la serie se obtuvo de la página web del BCRP. La metodología utilizada para estimar el modelo econométrico de pronóstico satisfactorio es la de Box-Jenkins, consistente en: identificación, estimación, validación y pronóstico. La serie histórica se divide en dos partes; de enero 2012 a diciembre 2019 para la identificación, estimación y validación del modelo (etapa de estimación) y el periodo de enero 2020 a diciembre 2020 para la validación del pronóstico (prueba o Testing). Las pruebas estadísticas correspondientes a cada una de estas etapas se realizaron en el software libre R.

## Análisis empírico

### Pruebas de estacionariedad

Se utilizó la prueba Augmented Dickey Fuller (ADF) para probar la raíz unitaria de los datos del IPC y sus datos de diferencia de primer orden.

Prueba de hipótesis de ADF de los datos del IPC.

H0: la serie no es estacionaria, tiene raíces unitarias

H1: la serie es estacionario, no tiene raíces unitarias

Dickey-Fuller = -0.70794, Lag order = 4,

p-value = 0.9668

Según la prueba de ADF, el valor p del IPC de la serie original es mucho mayor que 0.05, por lo que debemos aceptar la hipótesis nula de que la serie tiene raíz unitaria y es una serie de tiempo no estacionaria. Por lo tanto, necesitamos hacer un procesamiento diferencial.

Resultados de la prueba de hipótesis de ADF de los datos del IPC en primera diferencia.

Dickey-Fuller = -4.0577, Lag order = 4,

p-value = 0.01

El valor p de la secuencia de diferencias de primer orden  $d(\text{IPC})$  es menor que 0.05, por lo que no se acepta la hipótesis nula, es decir, no se considera que la secuencia tenga raíz unitaria y es una serie de tiempo estacionaria. Por tanto, la secuencia de diferencias de primer orden es una serie de tiempo estacionaria.

Del análisis anterior, podemos ver que la serie de tiempo después de la diferencia de primer orden es una serie estacionaria sin ruido blanco, que puede ajustarse al modelo SARIMA(p,1,q) (P,0,Q)<sub>12</sub>. Para obtener el modelo de ajuste relativamente óptimo, los valores de p, q, P y Q se determinan según el coeficiente de autocorrelación simple y el coeficiente de autocorrelación parcial, estableciendo de este modo diferentes modelos SARIMA(p,1,q) (P,0,Q)<sub>12</sub>.

### Establecimiento del modelo ARIMA

La función de autocorrelación parcial (PACF) determina los valores de p y P; la función de autocorrelación (ACF) determina los valores de q y Q. Los resultados del análisis de la función de autocorrelación simple y la función de autocorrelación parcial se muestran en la Figura 2 y 3 respectivamente.

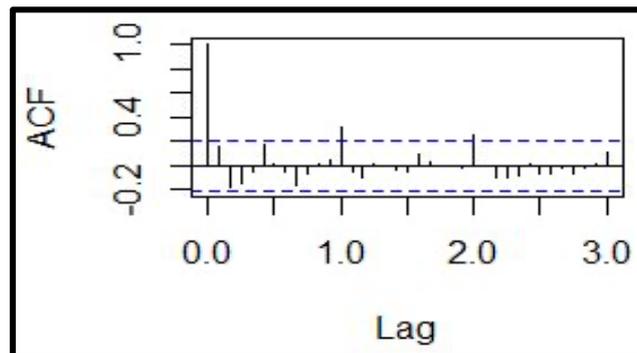


Figura 2. Análisis de autocorrelación simple de  $d(\text{IPC})$ .

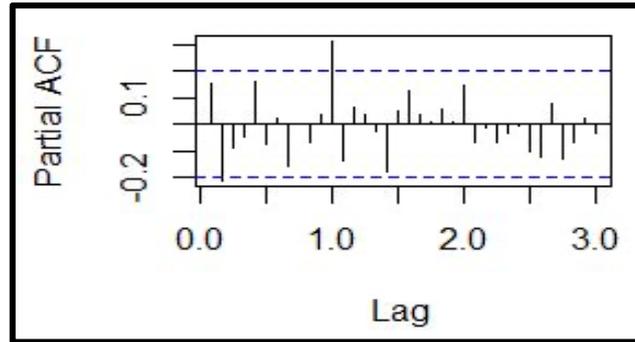


Figura 3. Análisis de autocorrelación parcial de d(IPC).

De acuerdo con las características estadísticas de la función de autocorrelación y la función de autocorrelación parcial en la Figura 2 y 3 se determinan los valores de  $p$ ,  $q$ ,  $P$  y  $Q$  para el modelo SARIMA. Para obtener un modelo de mejor ajuste se intenta establecer diferentes modelos SARIMA( $p, d, q$ )( $P, D, Q$ )<sub>12</sub> para la estimación de parámetros. Los resultados de los mejores modelos encontrados se muestran en la tabla 1.

Tabla 1. Presentación de resultados del mejor modelo SARIMA

Modelos	Criterio	Testing
SARIMA(2,1,0)(1,0,0) <sub>12</sub>	RMSE	0.703
	MAE	0.593
	MAPE	0.441
SARIMA(0,1,0)(0,0,2) <sub>12</sub>	RMSE	0.695
	MAE	0.635
	MAPE	0.473
SARIMA(2,1,0)(1,0,2) <sub>12</sub>	RMSE	0.429
	MAE	0.347
	MAPE	0.259
SARIMA(1,1,0)(0,0,2) <sub>12</sub>	RMSE	0.625
	MAE	0.556
	MAPE	0.414
SARIMA(2,1,0)(1,0,1) <sub>12</sub>	RMSE	0.451
	MAE	0.368
	MAPE	0.274
SARIMA(1,1,0)(1,0,1) <sub>12</sub>	RMSE	0.444
	MAE	0.364
	MAPE	0.271

Un modelo con un valor de RMSE, MAE y MAPE más bajo es mejor [12], en ese sentido podemos ver que el modelo SARIMA(0,1,0)(0,0,1)<sub>12</sub> tiene valores de RMSE,

MAE y MAPE más pequeños siendo el modelo simple y factible dado que solo se hizo una diferenciación simple ( $d=1$ ). Por lo tanto, se selecciona este modelo.

### Prueba del mejor modelo SARIMA

La comparación del valor real con el valor de ajuste del modelo SARIMA(2,1,0)(1,0,2)<sub>12</sub> se muestra en la Figura 4 y el valor de los residuales del modelo SARIMA(2,1,0)(1,0,2)<sub>12</sub> se muestra en la Figura 5.

El análisis de autocorrelación simple y autocorrelación parcial de los residuales se muestra en la Figura 6 y 7 respectivamente.

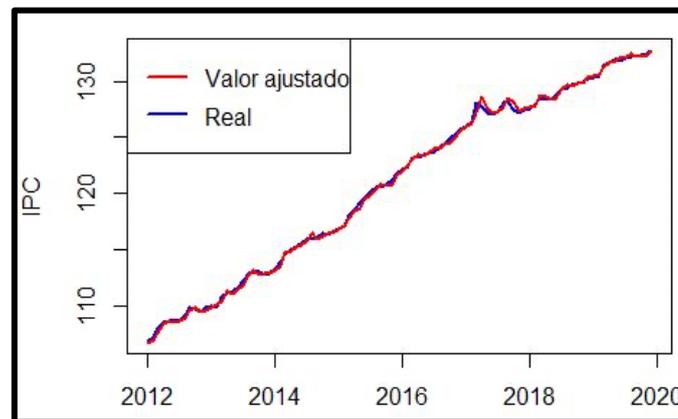


Figura 4. Valor real y valor ajustado del modelo SARIMA(2,1,0)(1,0,2)<sub>12</sub>.

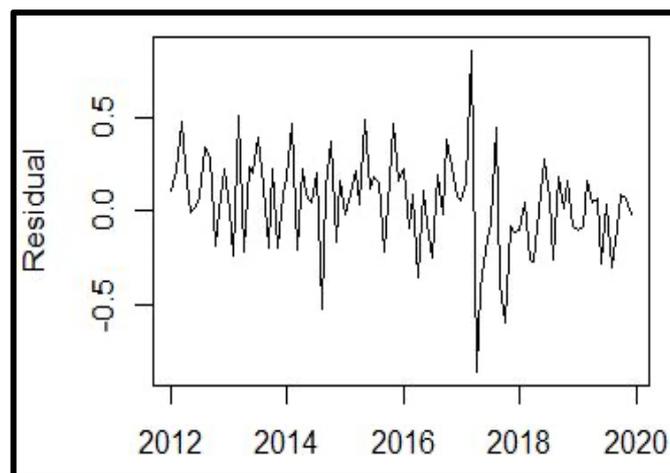


Figura 5. Valor residual del modelo SARIMA(2,1,0)(1,0,2)<sub>12</sub>.

En la Figura 4, el cambio del valor de ajuste y el valor real del modelo es consistente, y en la Figura 5 el residuo del modelo fluctúa alrededor del valor medio

cero. Por otra parte, se muestra en la Figura 6 que el coeficiente de autocorrelación simple ACF básicamente cae dentro de la banda de confianza con nivel de significancia de 0.05, es decir, la secuencia residual del modelo es una secuencia de ruido blanco aleatorio, y el modelo

SARIMA(2,1,0)(1,0,2)<sub>12</sub> establecido es más razonable.

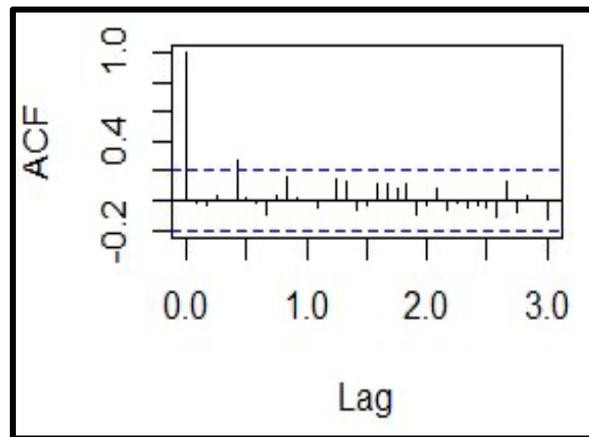


Figura 6. Análisis de autocorrelación simple de los residuales de SARIMA(2,1,0)(1,0,2)<sub>12</sub>.

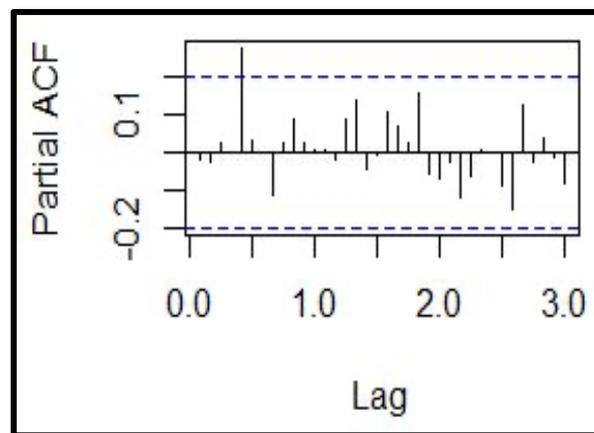


Figura 7. Análisis de autocorrelación parcial de los residuales de SARIMA(2,1,0)(1,0,2)<sub>12</sub>.

El modelo planteado es SARIMA(2,1,0) (1,0,2)<sub>12</sub> que es significativamente eficaz y se puede utilizar directamente para la predicción cuya fórmula es la siguiente:

$$\Delta IPC_t = 0.321IPC_{t-1} - 0.152IPC_{t-2} + 0.999IPC_{t-12} + \varepsilon_t - 0.932\varepsilon_{t-12} - 0.032\varepsilon_{t-24}$$

Donde:

$\Delta$ : Operador de primera diferencia.

IPC: Índice de Precios al Consumidor.

$\varepsilon_t$ : Error.

### Predicción del modelo ARIMA

El modelo SARIMA(2,1,0) (1,0,2)<sub>12</sub> se aplica para pronosticar el IPC mensual de Lima Metropolitana en 2020 y se compara con el valor real (ver Tabla 2) y se ve que los datos del IPC mensual de Lima Metropolitana en 2020 ajustados por el modelo SARIMA(2,1,0)(1,0,2)<sub>12</sub> tienen un error muy pequeño con el valor real, y el error en mayo es casi cero, lo que indica que el modelo se ajusta bien. y el resultado de la predicción es ideal, que se puede utilizar para la predicción de datos futuros. Utilizando el modelo SARIMA(2,1,0)(1,0,2)<sub>12</sub> para pronosticar los datos del IPC mensual de Lima Metropolitana en el primer semestre de 2021, los resultados se muestran en la Tabla 3.

Tabla 2. El valor real y el valor de pronóstico del IPC mensual en Lima Metropolitana en 2020.

Fecha	Valor real	Valor pronosticado	Error
Ene-20	132.77	132.90	0.13
Feb-20	132.96	133.16	0.20
Mar-20	133.82	133.99	0.17
Abr-20	133.96	134.17	0.21
May-20	134.23	134.30	0.07
Jun-20	133.87	134.42	0.54
Jul-20	134.49	134.74	0.25
Ago-20	134.35	135.08	0.73
Set-20	134.53	135.24	0.71
Oct-20	134.55	135.31	0.76
Nov-20	135.25	135.34	0.09
Dic-20	135.32	135.62	0.30

Tabla 3. Resultados de los pronósticos de los datos mensuales del IPC de Lima Metropolitana en el primer semestre de 2021.

Fecha	Valor pronosticado
Ene-21	135.3997
Feb-21	135.6337
Mar-21	136.4461
Abr-21	136.5744
May-21	136.7131
Jun-21	136.7131

A partir de los resultados del pronóstico, los datos del IPC de enero a junio de 2021 en comparación con el mismo período del año pasado, han aumentado todos los meses y la tendencia es creciente. Desde enero de 2020, la tendencia creciente del IPC no ha cambiado.

### Conclusiones

Hasta cierto punto, el modelo SARIMA puede utilizar el análisis de valores pasados y presentes para predecir el desarrollo futuro de series de tiempo. El modelo SARIMA es más adecuado para pronósticos a corto plazo. Con la extensión del período de pronóstico, el error de pronóstico aumentará en consecuencia. Utilizando los datos del IPC mensual de Lima Metropolitana en los últimos 9 años, se utiliza el modelo SARIMA(2,1,0)(1,0,2)<sub>12</sub> para ajustar y pronosticar. Los datos del IPC del primer semestre de 2021 pueden ser utilizados como referencia por los departamentos de gestión económica para tomar medidas de control a nivel macro con anticipación para garantizar la estabilidad y el desarrollo económico a largo plazo.

### Referencias

CERNA, Branco Ernesto Arana; FARFÁN, Karl Vladimir Mena; CABALLERO, Ana María Núñez. Modelo sarima para el pronóstico del índice mensual de precios al consumidor de Lima metropolitana. TZHOECOEN, 2021, vol. 13, no 1, p. 101-120.

CHU Rui, Pronóstico y análisis del IPC en Beijing basado en el modelo ARIMA, Ingeniería de valor, Shijiazhuang, Hebei, 2018, págs. 21-22. DOI: 10.14018 / j.cnki.cn13-1085 / n.2018.05.009

- CUI Wen-yan, XU Feng-hua, La aplicación del modelo ARIMA en la predicción del IPC en la provincia de Shandong, Revista de la Universidad Zaozhuang, vol. 33, Zaozhuang, Shandong, 2016, págs. 24-28.
- E. Hopali y A. Cakmak., Rev. Internacional de Información, Negocios y Gestión, vol. 12 (2020).
- GUO Xiao-feng, Predicción y análisis de la tendencia del IPC de China basada en el modelo ARIMA, Estadísticas y decisiones, Wuhan, Hubei, 2012, págs. 29-32. DOI: 10.13546 / j.cnki.tjyjc.2012.11.004
- LI Shu, WANG Fei, WANG Feng-xiao, Aplicación del modelo ARIMA en el pronóstico del índice de precios al consumidor de Qingdao, Revista de la Universidad de Artes y Ciencias de Lanzhou (Ciencias Naturales), Lanzhou, Gansu, 2020, págs. 11-16 . DOI: 10.13804 / j.cnki.2095-6991.2020.06.003
- Li Xuan, Huang Dongdong, Análisis y pronóstico del IPC basado en el modelo ARIMA, Revista de la Universidad de Shenyang (Ciencias Sociales), vol. 15, Shenyang, Liaoning, 2013, págs. 306-310.
- QI, Xingxiang. Análisis empírico del índice de precios al consumidor de Jinan basado en el modelo ARIMA. En VII Congreso Internacional de Investigación en Humanidades y Ciencias Sociales (ICHSSR 2021) . Atlantis Press, 2021. p. 235-238.
- RADJABAYCOLE, Jefry Esna T.; DJAMI, Ronald John; HAUMAHU, Gabriella. Forecasting the Ambon City Consumer Price Index Using Arima Box-Jenkins. Tensor: Pure and Applied Mathematics Journal, 2021, vol. 2, no 2, p. 87-96.
- SIESQUÉN ACOSTA, César Nicanor; RIOJAS YAMUNAQUÉ, José Jhosimar. Modelos de pronóstico del índice de precios al consumidor región lambayeque ya nivel nacional. Enero 2010 a diciembre 2016. 2018.
- Zhang Yangyang, Deng Wei, Song Changyu, Predicción del IPC en la provincia de Shandong basada en análisis de series de tiempo, Revista de la Universidad Normal de Shandong (Ciencias Naturales), Jinan, Shandong, 2020, págs. 329-336. DOI: 10.3969 / j.issn.1001-4748. 2020.03.010

ZHOU Mei-ying, Análisis de series de tiempo y predicción del IPC en la provincia de Hubei Basado en el modelo ARIMA, Kunming, Yunnan, Times Finance, 2011, págs. 66,110.