

UNIVERSIDAD NACIONAL SAN CRISTÓBAL DE HUAMANGA
FACULTAD DE INGENIERÍA DE MINAS, GEOLOGÍA Y CIVIL
ESCUELA PROFESIONAL DE CIENCIAS
FÍSICO - MATEMÁTICAS



**METODOLOGÍA BOOTSTRAP EN EL ANÁLISIS SERIAL DE LOS
INGRESOS MENSUALES. HOTEL EL CENTRO, AYACUCHO 2011 - 2014**

PRESENTADO POR: Bach. JUAN CARLOS OROSCO GAVILÁN

ASESOR: Lic. NILTON WILFREDO LOYOLA VERDE

TESIS PARA OPTAR EL TÍTULO DE:

Licenciado en Ciencias Físico Matemáticas, Especialidad
de Estadística

AYACUCHO - PERÚ

2017

Dedicatoria

*A mis padres, con amor
por ser un ejemplo para mí
y por su apoyo constante.*

Agradecimiento

- A la Universidad Nacional San Cristóbal de Huamanga y a todos los docentes de la Escuela Profesional de Ciencias Físico Matemáticas por nuestra formación profesional, en especial al Lic. Nilton Loyola Verde.
- A todos los estudiantes y amigos de la Escuela Profesional de Ciencias Físico Matemáticas quienes hicieron ligera la carga de labor de estudiante.

Presentación

Señores miembros del jurado:

En cumplimiento de los requisitos estipulados por el Reglamento de Grados y Títulos, de la Facultad de Ingeniería de Minas, Geología y Civil de la Universidad Nacional de San Cristóbal de Huamanga; me es honroso poner en consideración de vuestro criterio, la presente tesis intitulada: **“Metodología Bootstrap en el análisis serial de los ingresos mensuales. Hotel el Centro, Ayacucho 2011-2014”**. Con la finalidad de obtener el título profesional de Licenciado en Ciencias Físico Matemáticas, Especialidad de Estadística.

Espero sepan comprender los errores involuntarios, los que iré despejando con las experiencias que gane en el futuro.

Respetuosamente:

Bachiller JUAN CARLOS OROSCO GAVILÁN

Índice General

Dedicatoria	I
Agradecimientos	II
Presentación	III
Índice general	IV
Índice de cuadros	VII
Índice de figuras	VIII
Resumen	X
Abstract	XI
1. INTRODUCCIÓN	1
1.1. Objetivos del estudio	4
1.1.1. Objetivo General	4
1.1.2. Objetivos Específicos	4
2. MARCO TEÓRICO	6
2.1. Antecedentes de la investigación	6
2.1.1. Investigaciones acerca de predicciones mediante la metodología de Box y Jenkins	6
2.2. Investigaciones acerca de la metodología Bootstrap en series temporales .	8
2.3. Series Temporales	11
2.3.1. Definición	11

2.3.2.	Procesos Estocásticos	11
2.3.3.	Procesos Estacionarios	12
2.3.4.	Procesos de Ruido Blanco	13
2.3.5.	Estimación de los Momentos de Procesos Estacionarios	13
2.3.6.	Procesos Autorregresivos (AR)	14
2.3.7.	Procesos de Medias Móviles (MA)	14
2.3.8.	Procesos Autorregresivos de Medias Móviles (ARMA)	14
2.3.9.	Procesos Autorregresivos Integrados de Medias Móviles (ARIMA)	15
2.3.10.	Procesos ARIMA Estacionales (SARIMA)	16
2.4.	Metodología de Box y Jenkins	17
2.5.	Metodología Bootstrap	24
2.5.1.	Introducción	24
2.5.2.	Principio del Bootstrap	24
2.5.3.	Aplicaciones del Bootstrap	25
2.5.4.	Bootstrap en Series de Tiempo y Trabajos Relacionados	29
2.5.5.	Bootstrap en Modelos ARIMA	32
2.6.	Hipótesis de la investigación	33
2.6.1.	Hipótesis General	33
2.6.2.	Hipótesis Específicas	33
3.	METODOLOGÍA DE INVESTIGACIÓN	34
3.1.	Metodología	34
3.1.1.	Tipo de investigación	34
3.1.2.	Nivel de investigación	34
3.1.3.	Diseño de investigación	34
3.2.	Universo	34
3.3.	Población	35
3.4.	Muestra	35
3.4.1.	Muestra de entrenamiento	35
3.4.2.	Muestra de prueba	35
3.5.	Unidad de análisis	35
3.6.	Definición de las variables de estudio	35
3.7.	Métodos	35

3.7.1. Técnica de recolección de datos y procesamiento de la información .	35
3.7.2. Métodos de análisis de datos	36
4. RESULTADOS Y ANÁLISIS	38
4.1. Análisis exploratorio	38
4.2. Identificación de los modelos	47
4.3. Estimación y selección del modelo	47
4.4. Validación del modelo	50
4.5. Pronósticos	52
4.6. Comparación	54
5. CONCLUSIONES Y SUGERENCIAS	58
5.1. Conclusiones	58
5.2. Sugerencias	59
6. BIBLIOGRAFÍA	60
7. ANEXOS	60
8. A. Programa del algoritmo Bootstrap mediante el Software <i>R</i> para el cálculo de las predicciones h pasos hacia adelante del modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$ de la serie de ingresos mensuales.	65
9. B. Formato del libro de registro de huéspedes del Hotel El Centro, Ayacucho.	68

Índice de cuadros

1.	Prueba de raíz unitaria de Dickey y Fuller para la serie ingreso mensual del hotel El Centro periodo 2011 al 2014 (y_t).	40
2.	Prueba de raíz unitaria de Dickey y Fuller para la serie con la primera diferencia regular del ingreso mensual del hotel El Centro periodo 2011 al 2014 (z_t).	43
3.	Prueba de raíz unitaria de Dickey y Fuller para la serie diferenciada regularmente y estacionalmente del ingreso mensual (Dif(ingresos,1,12)) del hotel El Centro durante los años 2011 al 2014 (w_t).	46
4.	Modelo $SARIMA(1,1,0)(0,1,0)_{12}$ estimado y principales medidas estadísticas, para la serie del ingreso mensual del hotel El Centro periodo 2011 al 2014.	48
5.	Modelo $SARIMA(0,1,1)(0,1,0)_{12}$ estimado y principales medidas estadísticas, para la serie del ingreso mensual del hotel El Centro periodo 2011 al 2014.	49
6.	Pronósticos del modelo $SARIMA(0,1,1)(0,1,0)_{12}$ de los ingresos mensuales para el año 2015 para intervalos de confianza al 80 % y valores del error porcentual medio absoluto (EPMA), mediante metodología Bootstrap y metodología de Box y Jenkins.	53
7.	Pronósticos del modelo $SARIMA(0,1,1)(0,1,0)_{12}$ de los ingresos mensuales para el año 2015 para intervalos de confianza al 95 % y valores del error porcentual medio absoluto (EPMA), mediante metodología de Box y Jenkins y metodología Bootstrap.	55
8.	Intervalos de predicción de los ingresos mensuales para el año 2016 mediante la metodología Bootstrap a niveles de confianza del 80 % y 95 %. .	57

Índice de figuras

1.	Algoritmo de la Metodología de Box y Jenkins.	23
2.	Esquema del Algoritmo Bootstrap para la estimación del error estándar .	26
3.	Evolución del ingreso mensual en Nuevos Soles (S/.) del hotel El Centro en el periodo 2011 al 2014 (y_t).	38
4.	Correlogramas y valores de la Función de Autocorrelación (Autocorre- lation) (AC) y Función de Autocorrelación Parcial (Partial Correlation) (PAC) del ingreso mensual del hotel El Centro periodo 2011 al 2014 (y_t). .	39
5.	Evolución de la primera diferencia regular del ingreso mensual en Nuevos Soles (S/.) del hotel El Centro periodo 2011 al 2014 (z_t).	41
6.	Correlogramas y valores de la Función de Autocorrelación (Autocorre- lation) (AC) y Función de Autocorrelación Parcial (Partial Correlation) (PAC); del hotel El Centro periodo 2011 al 2014 (z_t).	42
7.	Evolución de la primera diferencia regular y diferencia estacional del in- greso mensual en Nuevos Soles (S/.) del hotel El Centro durante los años 2011 al 2014 (w_t).	44
8.	Correlogramas y valores de la Función de Autocorrelación (Autocorre- lation) (AC) y Función de Autocorrelación Parcial (Partial Correlation) (PAC) del ingreso mensual del hotel El Centro periodo 2011 al 2014 (w_t). .	45
9.	Correlogramas y valores de la Función de Autocorrelación (Autocorre- lation) (AC) y función de autocorrelación parcial (Partial Correlation) (PAC); y valor del estadístico de prueba de Ljung - Box (Q-Stat) para los residuos del modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$ de la serie del ingreso mensual del hotel El Centro periodo 2011 al 2014.	50

10.	Histograma de frecuencia y principales medidas estadísticas para los residuales del modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$ de la serie del ingreso mensual del hotel El Centro periodo 2011 al 2014.	52
11.	Pronósticos por intervalos de confianza al 80 % para el año 2015 del modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$ de la serie del ingreso mensual, mediante la metodología de Box y Jenkins y metodología Bootstrap.	54
12.	Pronósticos por intervalos de confianza al 95 % de confianza para el año 2015 del modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$ de la serie del ingreso mensual, mediante la metodología de Box y Jenkins y metodología Bootstrap.	56

Resumen

La presente investigación de naturaleza aplicada, de nivel descriptivo inferencial y explicativo de corte longitudinal y titulada “Metodología Bootstrap en el análisis serial de los ingresos mensuales. Hotel el Centro, Ayacucho 2011-2014”. Tuvo el objetivo de analizar y evaluar la metodología Bootstrap y la metodología de Box y Jenkins en el análisis serial de los ingresos mensuales del Hotel El Centro ubicado en el Departamento de Ayacucho, Provincia de Huamanga en el periodo 2011 al 2014. En este trabajo se presentaron, sucintamente, los conceptos básicos de series temporales, la metodología ARIMA de Box y Jenkins, la metodología Bootstrap y sus aplicaciones a la inferencia estadística y a las series temporales. Con los procedimientos mostrados fueron obtenidos predicciones mediante la metodología de Box y Jenkins y metodología Bootstrap, que fueron comparados con valores reales y finalmente fueron utilizadas medidas estadísticas para evaluar el desempeño de ambas metodologías en cuanto a su predicción. El análisis de la serie de ingresos mensuales del Hotel el Centro durante el periodo 2011 al 2014, mostró que el mejor modelo de pronóstico que ajusta a la serie es el modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$ y el estudio comparativo mostró que la aplicación de la metodología Bootstrap permite obtener intervalos de predicciones con menores amplitudes en cada horizonte y un menor error porcentual medio absoluto en comparación con la metodología de Box y Jenkins.

Palabras clave: Procesos ARIMA, predicción, Bootstrap.

Abstract

The present research of an applied nature, of descriptive level inferential and explanatory of longitudinal section and titled “Bootstrap Methodology in the serial analysis of the monthly income. Hotel el Centro, Ayacucho 2011-2014”. The objective was to analyze and evaluate the Bootstrap methodology and Box and Jenkins methodology in the serial analysis of the monthly income of the Hotel El Centro located in the Department of Ayacucho, province of Huamanga in the period 2011 to 2014. In this work were presented, Briefly, the basic concepts of time series, the ARIMA methodology of Box and Jenkins, the Bootstrap methodology and their applications to statistical inference and time series. With the procedures shown, predictions were obtained using Box and Jenkins methodology and Bootstrap methodology, which were compared with real values and finally measures were used to evaluate the performance of both methodologies in terms of their prediction. The analysis of the monthly income series of the Hotel el Centro during the period 2011 to 2014 showed that the best forecast model is the model $SARIMA(0, 1, 1)(0, 1, 0)_{12}$ and the comparative study showed that the application of the Bootstrap methodology allows to obtain intervals of predictions with smaller amplitudes in each horizon and a smaller error percentage average absolute in comparison with the methodology of Box and Jenkins.

Keywords: ARIMA process, prediction, Bootstrap.

1. INTRODUCCIÓN

Efron (1979) desarrolla y publica el análisis formal del Bootstrap (uno de los procedimientos de remuestreo). Es entonces cuando realmente este enfoque cobra una importante fuerza teórica y capta el interés de toda la comunidad de estadísticos, quienes comienzan a explorarlo y utilizarlo para solucionar una amplia gama de problemas en probabilidades e inferencia. Este proceder ha sido considerado por la American Statistical Association como “el único gran descubrimiento en estadística desde 1970”. (Miranda, 2003).

Para facilitar el proceso de simulación, en 1973 se ideó el software Resampling Stats, soporte técnico que impulsó el proceso de explotación del nuevo método. Aprovechando sus virtudes y haciendo uso del remuestreo se pueden resolver tanto problemas extremadamente sencillos, tales como determinar la probabilidad de obtener al menos una cara en el lanzamiento de dos monedas, como problemas de una naturaleza o envergadura tal que no podrían ser resueltos con los métodos teóricos tradicionales. En tanto el poder de este proceder que se ha llegado a afirmar que la historia de la estadística sería otra si el desarrollo de la computación, imprescindible para aplicar el remuestreo, hubiera llegado antes. (Ricketts y Berry, 1994).

En la actualidad los procedimientos de remuestreo son ampliamente conocidos y utilizados en diversos campos de la investigación y se han ido incorporando como técnicas bastante habituales en el análisis e interpretación de cierto tipo de datos.

Miranda (2003) menciona que, la predicción en series temporales en el ámbito de la economía ha adquirido gran atención dentro de las investigaciones empíricas, tanto que los estadísticos y econometristas han empezado a dedicar mayores esfuerzos a este tipo de información, tal y como se demuestra con el número de libros y artículos publicados en la materia.

Frecuentemente en series temporales se emplean técnicas estadísticas tradicionales para

las estimaciones de parámetros, para tales es necesario conocer determinadas características de la distribución muestral de los estadísticos, sin embargo, ¿qué sucede si tenemos una población estadísticamente pequeña o donde no podemos conocer su distribución a priori? En el pronóstico de series de tiempo es usual el uso de la metodología de Box y Jenkins, las cuales permiten obtener buenas aproximaciones, y de igual manera en dicho procedimiento es necesario conocer determinadas características de la distribución muestral de los estadísticos o la distribución muestral de los residuales. Volviendo al mismo problema de tener que conocer las distribuciones a priori.

El método bootstrap o técnica del remuestreo, reemplaza las derivaciones teóricas del enfoque ya dicho por la evaluación de los estadísticos en submuestras obtenidas a partir de los datos originales, y mediante estos valores se obtienen estimadores de las medidas de exactitud o de la distribución muestral del estadístico. (Tineo, 2005).

Así mismo, el estudio de la variable ingreso nos da a conocer el estado económico dentro de alguna empresa, de esta manera es menester conocer el pasado histórico de dicha variable para poder adelantarse al futuro y hacer predicciones, por tanto, su estudio es vital para la estabilidad y el crecimiento de la empresa; consecuentemente es necesario conocer herramientas y métodos estadísticos adecuados.

Hasta el momento no se tiene conocimiento de la aplicación metodológica bootstrap en el ámbito ayacuchano desde un enfoque en la inferencia estadística y en series temporales, las propuestas en investigaciones se han enfocado al uso de la estadística clásica. Ante las razones expuestas es importante analizar la metodología bootstrap en un análisis serial, para así tener un enfoque mayor al análisis en series temporales.

Frente al planteamiento del problema surge la siguiente interrogante.

¿Cómo es el comportamiento serial de los ingresos mensuales del Hotel El Centro, Huamanga - Ayacucho en el periodo 2011 al 2014?

A partir del cual se plantean los siguientes problemas específicos:

- ¿Cómo es el comportamiento de predicción de los ingresos mensuales del Hotel El Centro, Huamanga - Ayacucho en el periodo 2011 al 2014, mediante la metodología de Box y Jenkins?
- ¿Cómo es el comportamiento de predicción de los ingresos mensuales del Hotel El

Centro, Huamanga - Ayacucho en el periodo 2011 al 2014, mediante la metodología Bootstrap?

- ¿Será que la Metodología Bootstrap presenta mejores predicciones de los ingresos mensuales del Hotel El Centro, Huamanga - Ayacucho en el periodo 2011 al 2014, en contraste con la metodología de Box y Jenkins.

El estudio de fenómenos económicos, dentro de ella el estudio del promedio de los ingresos, constituye una variable fundamental asociado a las acciones que se materializan en decisiones. Como tarea de planeamiento futuro es necesario el análisis de estas variables, por tanto, es relevante comprender el pasado, emplear los datos históricos y el buen juicio para elaborar planes inteligentes que ayuden en un futuro al buen desarrollo de la empresa.

En la presente investigación se ha preferido obviar, en datos de corte longitudinal las distribuciones muestrales de los residuales. Así a manera de contribuir con la mejora en el análisis estadístico, tengo a bien estudiar y utilizar la técnica bootstrap, con el objetivo de divulgar el uso de esta técnica en el campo de la investigación.

La técnica Bootstrap, se basa en la reproducción de los datos originales mediante un remuestreo. Si nuestras observaciones tienen una estructura de dependencia, ésta debe estar reflejada en los nuevos datos. (Tineo, 2005). Tal y como acontece en el análisis de las series temporales, donde los residuos del modelo propuesto original son reproducidos mediante un remuestreo así obteniendo un número elevado de predicciones obtenidos del modelo original. Por lo tanto, los métodos variarán en función de la estructura temporal existente y como ésta se refleje en su predicción.

Los métodos estadísticos clásicos se apoyan en modelos matemáticos de naturaleza estocástica, de tal forma que los resultados que de ellos se derivan, requieren en muchas ocasiones, complejos desarrollos analíticos, lo que ha supuesto un obstáculo para su utilización comprensiva en muchas áreas científicas, básicamente para utilizar aquellos modelos estadísticos que se ajusten a características de los datos de cada disciplina. En términos más precisos, los modelos estadísticos incorporan distintos supuestos que establecen restricciones sobre las variables aleatorias analizadas.(Alvarez y Rubio, 2000).

Dicho esto es imprescindible conocer métodos estadísticos adecuados a cumplir nuestros objetivos, ayudándonos de preferencia a reducir el tiempo, el costo, y dando un resultado

óptimo a nuestros requerimientos.

Por estas razones se pretende hacer uso de esta metodología bootstrap en la aplicación de estimaciones de parámetros, predicciones puntuales y por intervalos con datos de corte longitudinal obtenidos del Hotel El Centro de la ciudad de Ayacucho durante el periodo 2011 al 2014.

La presente investigación, analiza la serie temporal de ingresos mensuales del Hotel el Centro ubicado en la provincia de Huamanga, en el departamento y distrito de Ayacucho en los años 2011 al 2014 mediante la metodología Bootstrap, y describe e implementa un programa para el método Bootstrap, orientado a series de tiempo. En el capítulo 1 se explica los fundamentos para la realización de la tesis, así como el planteamiento del problema, justificación y objetivos de la investigación. En el capítulo 2 se presentan los antecedentes de la investigación, así mismo los diversos conceptos necesarios para el entendimiento de la tesis que consiste en: base teórica y explicación del procedimiento de la metodología Bootstrap y el procedimiento para el análisis de predicción haciendo un énfasis en la metodología de Box y Jenkins, que será una herramienta más para el cálculo de análisis bootstrap, Las cuales nos permitirán a conocer cuál es el comportamiento serial de los ingresos mensuales en el Hotel El Centro. En el capítulo 3, se describen los materiales y métodos para el modelo de pronóstico y bootstrap en series temporales, para los ingresos del Hotel El Centro durante los años 2011 al 2014. En el capítulo 4 se muestran los resultados y análisis, y por último en el capítulo 5 se presentan las conclusiones y recomendaciones.

1.1. Objetivos del estudio

1.1.1. Objetivo General

Analizar el comportamiento serial de los ingresos mensuales mediante la metodología Bootstrap y de Box y Jenkins, Hotel El Centro, Huamanga - Ayacucho en el periodo 2011 - 2014.

1.1.2. Objetivos Específicos

- Proponer un modelo de predicción mediante la metodología de Box y Jenkins de los ingresos mensuales del Hotel El Centro, Huamanga - Ayacucho en el periodo

2011 - 2014.

- Proponer un modelo de predicción mediante la metodología Bootstrap de los ingresos mensuales del Hotel El Centro, Huamanga - Ayacucho en el periodo 2011 - 2014.
- Determinar si mediante la metodología Bootstrap se obtienen mejores predicciones de los ingresos mensuales en contraste con la metodología de Box y Jenkins del Hotel El Centro, Huamanga - Ayacucho en el periodo 2011 - 2014.

2. MARCO TEÓRICO

2.1. Antecedentes de la investigación

En esta sección primeramente se revisaron investigaciones nacionales e internacionales acerca de la metodología Box y Jenkins, desde un enfoque de predicción en distintas aplicaciones, principalmente en el área de la economía en particular en la predicción en ventas. En seguida se revisaron investigaciones acerca de la metodología Bootstrap en series temporales, teniendo en cuenta que no se hallaron investigaciones nacionales acerca de esta metodología aplicada a en las series temporales, por tanto se consideraron solamente investigaciones internacionales.

2.1.1. Investigaciones acerca de predicciones mediante la metodología de Box y Jenkins

Mirando y Ramos (2010) realizaron un estudio con el objetivo de pronosticar a nivel nacional y regional la tendencia de las enfermedades diarreicas agudas (EDA) en menores de cinco años de edad para el año 2010 mediante un modelo ARIMA con el enfoque Box y Jenkins, obteniendo como resultados las tendencias de las EDAs en población menor de cinco años para el año 2010, fue ascendente en relación a los años 2008 y 2009. Se estimó para los meses de enero y febrero del 2010 un incremento de tres episodios por cada mil menores de 5 años respecto al año 2009; para los meses de Febrero y Julio del 2010 una tasa máxima de 24 episodios por cada mil menores de cinco años. El modelo ARIMA con el enfoque BoxJenkins mostró para el año 2010 una tendencia ascendente de las EDAs en relación a los años 2008 y 2009 determinándose seis regiones con tendencia al incremento en las que debería priorizarse las intervenciones en salud.

Urquiza (2011) realizó una investigación con el objetivo de predecir el consumo de energía eléctrica residencial en la Región de Piura, partiendo de los datos históricos mensuales del

consumo residencial en la Región, correspondientes al periodo 2001 a junio 2013, con un horizonte de predicción a mediano plazo. Como resultado, permitió obtener un modelo $ARIMA(2,1,1)(0,1,1)$, que explica eficientemente el consumo histórico residencial y la predicción del mismo para los dos años siguientes, permitiendo tener un sustento técnico para las decisiones de planificación e inversión en la ampliación de redes de media y baja tensión en la zona de concesión, evitando el exceso o defecto de inversiones que resultan perjudiciales para los intereses de la empresa eléctrica. El procedimiento ARIMA o de Box y Jenkins permitió encontrar un modelo óptimo con mejores resultados predictivos frente al método de alisado exponencial Winters.

Medina y Vizconde (2011) realizaron un estudio titulado “modelo SARIMA para la llegada mensual de visitantes extranjeros por el Aeropuerto Internacional Jorge Chavez”, con el objetivo de determinar un modelo de pronóstico para la llegada mensual de visitantes extranjeros por el Aeropuerto Internacional “Jorge Chávez” de Perú, basado en la serie histórica en el periodo enero 1997 - junio 2009. Se utilizó la metodología de Box-Jenkins (identificación, estimación, prueba de adecuacidad, y pronóstico y validación) y el programa Eviews, versión 5.0. El periodo de pronóstico fue julio - diciembre 2009. El modelo estimado fue un modelo SARIMA $[0,1,1] [(12,36),0,(12,36,48)]$, apropiado y con validez de pronóstico, y las llegadas mensuales para el periodo de pronóstico fueron estimadas con desviación absoluta media de 3972 pasajeros al mes y 3.61 % de porcentaje de error absoluto.

Alarcón (2012) realizó un estudio con la finalidad pronosticar la inflación mensual del Ecuador y evaluar la capacidad predictiva de los modelos utilizados, Se utilizó la modelación econométrica de la serie a través de las metodologías de Box y Jenkins obteniendo el modelo $SARIMA(1,0,2)(1,0,0)_{12}$ con un error cuadrático medio menor.

Guevara (2012) Empleó la metodología de Box y Jenkins para crear una propuesta de diseño de pronóstico, para la demanda de llamadas telefónicas en un centro de atención de llamadas, en la ciudad de Lima. El estudio permitió la resolución del problema de planificación del recurso humano, haciendo uso de las técnicas de optimización, la aplicación de la metodología de Box y Jenkins tuvo el objetivo de resolver problemas de atención en centro de llamadas, de manera que la solución es un arma estratégica en la gestión del recurso humano, se hizo uso de los modelos SARIMA como modelos más adecuados

para predecir la demanda de llamadas del centro de llamadas en estudio.

Carvalho y cols. (2013) realizó una aplicación de un modelo SARIMA en la predicción de ventas de motocicletas, con el objetivo de estudiar un modelo de predicción de ventas, basado en la metodología de Box y Jenkins, fue utilizado la serie de ventas mensuales de los años 2006 al 2010 de las motocicletas más comerciales en Brasil. El modelo seleccionado fue un modelo SARIMA, teniendo como criterio de selección los valores del error porcentual medio absoluto (EPMA) y el AIC, fueron comparados con valores reales permitiendo concluir que el modelo SARIMA es una alternativa viable para la serie temporal analizada.

Quiroz (2014) realizó un estudio con la finalidad de pronosticar la demanda de energía eléctrica residencial en la zona de Lima Sur mediante un enfoque estocástico ARIMA y método de Holt Winters, como resultado obtuvo que la metodología ARIMA o de Box y Jenkins proporciona mejores resultados predictivos, tanto gráficamente como en margen de error y en complejidad de variable de estudio, demostrando la Hipótesis Nula formulada para la investigación, de ofrecer resultados con mejor eficiencia y bondad predictiva. Así con el método ARIMA se obtuvieron resultados con menor error absoluto porcentual promedio.

2.2. Investigaciones acerca de la metodología Bootstrap en series temporales

Alvarez y Rubio (2000) presentaron la metodología Bootstrap como una alternativa para construir intervalos de predicción en series temporales, realizaron un estudio de los métodos bootstrap cuando las series analizadas son heterocedásticas, haciendo especial énfasis en la predicción de modelos heterocedásticos condicionalmente autorregresivos (ARCH). Presentaron un estudio empírico sobre el índice IBEX-35 con datos diarios desde 1992 hasta 1998 y se comparó la metodología bootstrap con los métodos tradicionales utilizados en la construcción de intervalos de predicción. En los modelos de rentabilidad-riesgo de tipo GARCH-M la predicción bootstrap mostró un excelente comportamiento en el corto y medio plazo.

Henrique (2003) estudió el tráfico aéreo en dos regiones de Portugal, con el objetivo de

obtener predicciones para el planeamiento y tomar decisiones anticipadamente, respecto al flujo aéreo. Aplicando la metodología Bootstrap por bloques en modelos de predicción por el método de Holt Winters y ARIMA, y realizando remuestras a los residuales de modelos obtenidos con anticipación, se obtuvieron predicciones por intervalos adecuados para el flujo aéreo.

Kim. y cols. (2009) evaluaron el rendimiento de intervalos de predicción de series mensuales para el número de llegadas de turistas a Hong Kong y Australia, empleando el modelo autorregresivo (AR), el modelo AR usando el Bootstrap corregido por sesgo, los modelos estacionales ARIMA, las innovaciones de los modelos de estado-espacio para el suavizado exponencial, y modelos de series de tiempo estructurales de Harvey. Se encontraron que todos los modelos producen intervalos de predicción satisfactorios, excepto para el modelo autorregresivo (AR). En particular, los basados en el Bootstrap corregido por sesgo funcionan mejor en general, obteniendo intervalos con tasas de cobertura precisas, especialmente cuando el horizonte de pronóstico es largo.

Marques (2013) realizó un estudio de la metodología Bootstrap en métodos de alisamiento exponencial, principalmente, al método no paramétrico de Holt Winters para la predicción de variables ambientales de calidad de agua de superficie, de marzo de 2000 a diciembre de 2011. Las predicciones fueron comparadas con valores reales, utilizando el error cuadrático medio como medida para evaluar el desempeño de la metodología Bootstrap en cuanto a la calidad de las predicciones. El estudio comparativo mostró que las aplicaciones del Bootstrap en asociación al método de Holt Winters permitió obtener intervalos de confianza de las predicciones con menor amplitud a comparación de los obtenidos por los métodos usuales.

Stine (1987) propuso el primer método bootstrap para estimar el error cuadrático medio del predictor lineal clásico cuando el proceso es un modelo autorregresivo de orden p (AR(p)). El método se basó en la obtención de la función de distribución empírica de los errores centrados. A partir de ésta extraía muestras para reproducir un proceso AR(p) con los parámetros estimados inicialmente y fijando los primeros p valores de la serie. Dada la serie bootstrap se estiman los parámetros y se obtiene el predictor clásico para evaluar el error cuadrático medio de dicha predicción. Repitiendo un número elevado de veces este proceso podemos aproximar el error cuadrático medio de la predicción puntual.

Thombs y Schucany (1990) propusieron la construcción de intervalos de predicción en modelos $AR(p)$, propusieron realizar una representación hacia atrás del proceso para posteriormente utilizar la representación habitual para obtener los valores futuros. Estos autores demostraron la validez asintótica de este método bootstrap y realizaron un estudio de simulación donde comparan con los intervalos clásicos para diferentes distribuciones y coberturas.

McCullough (1994) sugirió que el método percentil de obtener los intervalos de predicción no era el más adecuado cuando existe asimetría en el error de predicción y, por lo tanto, propuso un método de corrección de sesgo. Además, advirtió que la representación backward utilizada por Thombs y Schucany (1990) no asegura que los errores sean independientes e idénticamente distribuidas y utilizó un procedimiento más sofisticado para obtener dichos errores hacia atrás.

La validez asintótica del método de Thombs y Schucany (1990) para modelos AR fue demostrada por Jurado, Manteinga, y Sánchez (1995). La idea general se basa en obtener la serie diferenciada y aplicar el método TS para construir la réplica bootstrap de la serie que tiene la estructura autoregresiva. Posteriormente, se integra para tener réplicas bootstrap de la serie original.

Un inconveniente importante del método Thombs y Schucany (1990) consiste en la representación backward de la serie, no sólo por el gasto computacional sino porque no es fácil cuando los procesos no tienen distribución normal. Cao y cols. (1997) proponen un método alternativo, que es consistente, para construir intervalos de predicción sin tener que realizar réplicas hacia atrás de la serie y demostraron la consistencia de este método tanto en el caso de utilizar la función de distribución empírica de los residuos directamente como en el de utilizar un bootstrap suavizado, es decir, cuando se extrae la muestra a partir de una suavización no paramétrica de la distribución.

Finalmente, cabe destacar que Pascual, Romo, y Ruiz (1999) propusieron un método que capta el error de predicción debido a la incertidumbre en la estimación pero que no es necesaria realizar una representación backward de la serie.

2.3. Series Temporales

2.3.1. Definición

Una serie de tiempo es una secuencia cronológica de observaciones de una variable particular. Conseguimos series de tiempo en los distintos campos del saber: en economía, mercado, demografía, meteorología, ingeniería, etc. Muchos son los ejemplos de series de tiempo que podrían citarse, tales como: Las ventas mensuales de una empresa en la última década, El número de automóviles producidos por año de determinada marca en período 1985-2000, La temperatura diaria promedio en los últimos 6 meses, etc.

Según Alpuim (1998), los objetivos por los cuales se puede querer analizar una serie de tiempo son:

- Descripción: Cuando tenemos una serie de tiempo, el primer paso en el análisis es graficar los datos y obtener medidas descriptivas simples de las propiedades principales de la serie.
- Explicación: Cuando las observaciones son tomadas sobre dos o más variables, es posible usar la variación en una serie para explicar la variación en otras series.
- Predicción: Dada una serie de tiempo se puede querer predecir los valores futuros de la serie. Este es el objetivo más frecuente en el análisis de series de tiempo.
- Control: Cuando una serie de tiempo se genera por mediciones de calidad de un proceso, el objetivo del análisis puede ser el control de proceso.

2.3.2. Procesos Estocásticos

Un proceso estocástico es una colección de variables aleatorias ordenadas en el tiempo, así, sea C un conjunto arbitrario y una colección $y = \{y_t, t \in C\}$, tal que para cada $t \in C$, y_t es una variable aleatoria, es decir para cada valor de t del conjunto C está definida una variable aleatoria, y_t , y los valores observados de las variables aleatorias en distintos instantes forman una serie temporal.

El proceso queda caracterizado si se define la distribución de probabilidad conjunta de las variables aleatorias $(y_1, \dots, y_t, \dots, y_T)$, para cualquier valor de T . Estas distribuciones se denominan las *distribuciones finito-dimensionales* del proceso. Se dice que se conoce la

estructura probabilística de un proceso estocástico cuando se conozcan estas distribuciones, que determinan la distribución de cualquier subconjunto de variables y, en particular, las distribuciones marginales de cada variable que tienen las siguientes propiedades:

- Función de medias del proceso $E(y_t) = \mu_t$,
- Función de varianza del proceso $Var(y_t) = \sigma_t^2$,
- Función de autocovarianzas del proceso

$$\gamma(t, t+j) = Cov(y_t, y_{t+j}) = E[(y_t - \mu_t)(y_{t+j} - \mu_{t+j})],$$

en particular $\gamma(t, t) = Var(y_t) = \sigma_t^2$, y

- Coeficiente de autocorrelación de orden $(t, t+j)$

$$\rho(t, t+j) = \frac{Cov(t, t+j)}{\sigma_t \sigma_{t+j}} = \frac{\gamma(t, t+j)}{\gamma^{1/2}(t, t) \gamma^{1/2}(t+j, t+j)}$$

2.3.3. Procesos Estacionarios

Se dice que un proceso estocástico es estacionario en *sentido estricto* si:

- Las distribuciones marginales de todas las variables son idénticas,
- Las distribuciones finito-dimensionales de cualquier conjunto de variables sólo dependen de los retardos entre ellas.

Se dice que un proceso estocástico es estacionario en *sentido débil*, si para todo t :

- $\mu_t = \mu = cte$,
- $\sigma_t^2 = \sigma^2 = cte$,
- $\gamma(t, t-k) = E[(y_t - \mu)(y_{t-k} - \mu)] = \gamma_k, k = 0, \pm 1, \pm 2, \dots$

Las dos primeras condiciones indican que la media y la varianza son constantes. La tercera, que la covarianza entre dos variables depende sólo de su separación.

2.3.4. Procesos de Ruido Blanco

Se dice que $\{\varepsilon_t \in t \in \mathbb{Z}\}$ es un ruido blanco discreto si:

- $E(\varepsilon_t) = 0, \quad t=1,2,\dots$
- $Var(\varepsilon_t) = \sigma^2, \quad t=1,2,\dots$
- $Cov(y_t, y_{t-k}) = 0, \quad k = 0, \pm 1, \pm 2, \dots$

La condición primera establece que la esperanza es siempre constante e igual a cero, la segunda condición que la varianza es constante y la tercera condición que las variables del proceso están incorreladas para todos los retardos.

2.3.5. Estimación de los Momentos de Procesos Estacionarios

Supongamos un proceso estacionario con media μ , varianza σ^2 y covarianza γ_k del que se observa una realización $(y_1, y_2, \dots, y_T)$. A continuación se presenta los estimadores de la media, la varianza, las covarianzas y las autocorrelaciones del proceso a partir de esta única realización disponible.

- El estimador de la media poblacional es la media muestral $\bar{y} = \frac{\sum_{t=1}^T y_t}{T}$
- El estimador de la autocovarianza poblacional es dado por:

$$\hat{\gamma}_k = \frac{1}{T} \sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})$$

Donde la varianza del proceso se estima por $\hat{\gamma}_0$.

- La autocorrelación se estima por, $\hat{\rho} = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$

Además de estudiar la correlación de una forma global, también nos interesa estudiar la correlación parcial entre $y(t)$ y $y(t+k)$, cuando se fijan las variables intermedias $y(t+1), y(t+2), \dots, y(t+k-1)$, esto es la correlación simple entre $y(t)$ y $y(t+k)$ después de eliminar el efecto que sobre estas producen las variables intermedias, esta es llamada *función de autocorrelación parcial* (FACP).

Y podemos representar la función de autocorrelación (FAC) y la función de autocorrelación parcial (FACP) dando los coeficientes de correlación muestral o estimados

en función del retardo. A esta representación se le denomina *correlograma*, o *función de autocorrelación muestral*.

2.3.6. Procesos Autorregresivos (AR)

Los procesos autorregresivos constituyen un tipo importante de procesos estacionarios. Son obtenidos al imponer una dependencia de tipo lineal entre el valor del proceso en un determinado instante y los valores del pasado. Se dice que y_t sigue un proceso estocástico autorregresivo de orden p , o $AR(p)$ si se escribe como:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t$$

Donde $\phi_i, i = 1, \dots, p$ son constantes fijas y ε_t es una secuencia de variables no correlacionadas con media cero y varianza σ^2 , esto es, un proceso de ruido blanco.

Definiendo el operador autorregresivo estacionario de orden p ,

$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$, entonces el proceso $AR(p)$ se puede escribir también como:

$$\phi(B)y_t = \varepsilon_t$$

2.3.7. Procesos de Medias Móviles (MA)

El proceso de medias móviles consiste en expresar y_t en términos de un proceso más simple, como es el proceso de ruido blanco. Así, un proceso de medias móviles de orden q se define en cada instante t , como la media ponderada de las $q + 1$ observaciones de un proceso de ruido blanco ε_t . Estos procesos son siempre estacionarios.

Un proceso se dice de medias móviles de orden q , o $MA(q)$, si:

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

siendo $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ el operador de medias móviles de orden q , entonces el proceso $MA(q)$ se puede expresar también como:

$$y_t = \theta(B)\varepsilon_t$$

2.3.8. Procesos Autorregresivos de Medias Móviles (ARMA)

Alpuim (1998) menciona que, tanto los procesos autorregresivos como los procesos de medias móviles constituyen buenos modelos para las series temporales que se pueden ob-

servar en la “vida real”. Así es natural considerar procesos más generales que son contruidos a partir de la combinación entre estos dos modelos.

Decimos que un proceso es un proceso autorregresivo de medias móviles de orden (p, q) , o $ARMA(p, q)$, si satisface la ecuación:

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

Donde ε_t es un proceso de ruido blanco, y si $\phi(B)$ y $\theta(B)$ son los operadores autorregresivos y de medias móviles respectivamente, entonces el proceso $ARMA(p, q)$ se puede también expresar en la forma:

$$\phi(B)y_t = \theta(B)\varepsilon_t$$

2.3.9. Procesos Autorregresivos Integrados de Medias Móviles (ARIMA)

Los modelos presentados en los capítulos anteriores son apropiados para describir series estacionarias, sin embargo muchas series encontradas en la práctica no son estacionarias, si una serie de tiempo no es estacionaria en el sentido antes definido, se denomina serie de tiempo no estacionaria.

Para el tratamiento de la no estacionariedad en la media se propone la diferenciación sucesiva de la serie, aprovechando la propiedad de que gozan una gran parte de los procesos estocásticos, de convertirse en estacionarios al diferenciarlos cierto número de veces. Las primeras diferencias de los valores de la serie de tiempo $(y_1, y_2, \dots, y_n)$ son:

$$\nabla y_t = y_t - y_{t-1} = (1 - B)y_t$$

para $t = 2, 3, \dots, n$

Las segundas diferencias

$$\nabla^2 y_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = (1 - B)^2 y_t$$

para $t = 3, 4, \dots, n$

Generalizando para $d \geq 0$

$$\nabla^d Y_t = (1 - B)^d y_t$$

El objetivo es determinar la serie estacionaria obtenida por la menor diferenciación, una vez que la varianza aumenta con la diferenciación.

La diferenciación no elimina todos los casos de no estacionariedad. La estacionariedad puede ser conseguida a través de transformaciones que permiten estabilizar la varianza. Podemos entonces recurrir a transformaciones logarítmicas de la serie o por la transformación de Box y Cox. ver Murteira y Muller (1990).

Una metodología bastante utilizada en el análisis de modelos paramétricos es conocida como metodología ARIMA de Box y Jenkins. Tal metodología consiste en ajustar modelos auto-regresivos integrados de medias móviles, $ARIMA(p, d, q)$, a un conjunto de datos.

Si $z_t = \nabla^d y_t$ fuese estacionaria, podemos representar z_t por un modelo $ARMA(p, q)$, es decir:

$$\phi(B)z_t = \theta(B)\varepsilon_t$$

decimos que z_t sigue un modelo autorregresivo, integrado, de medias móviles, o modelo $ARIMA$:

$$\phi(B)\Delta^d y_t = \theta(B)\varepsilon_t$$

de orden (p, d, q) y se denota como $ARIMA(p, d, q)$, donde p y q son ordenes de $\phi(B)$ y $\theta(B)$, respectivamente y ε_t es un proceso de ruido blanco,

2.3.10. Procesos ARIMA Estacionales (SARIMA)

Continuamos el estudio de los procesos no estacionarios analizando un tipo de falta de estacionariedad en la media que es muy frecuente en la práctica que es el comportamiento estacional. La estacionalidad hace que la media de las observaciones no sea constante, pero evoluciona de forma previsible de acuerdo a un patrón cíclico.

Una serie es estacional cuando su valor esperado no es constante, pero varía con una pauta cíclica. En concreto si:

$$E(y_t) = E(y_{t+s})$$

entonces se dice que la serie tiene una estacionalidad de periodo s .

Se vio que es posible convertir series no estacionarias en series estacionarias tomando diferencias regulares, es decir, entre periodos consecutivos. Así mismo también es posible convertir una serie estacional en serie estacionaria aplicando una diferencia estacional, como sigue:

$$\nabla_s y_t = (1 - B^s)y_t = y_t - y_{t-s}$$

Uniendo estos resultados de convertir series no estacionarias y estacionales en estacionarias, concluimos que podemos convertir una serie con estacionalidad en estacionaria mediante la transformación:

$$w_t = \nabla_s^D \nabla^d y_t$$

Donde D es el número de diferencias estacionales y d el número de diferencias regulares. Se obtiene así el modelo ARIMA estacional multiplicativo o $SARIMA(p, d, q)(P, D, Q)$, que tiene la forma

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

Donde:

$\Phi_P(B^s) = (1 - \Phi_1 B^s - \dots - \Phi_P B^{sP})$ es el operador AR estacional de orden P ,

$\phi_p(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$ es el operador AR regular de orden p ,

$\nabla_s^D = (1 - B^s)^D$ representa las diferencias estacionales,

$\nabla^d = (1 - B)^d$ representa las diferencias regulares,

$\Theta_Q(B^s) = (1 - \Theta_1 B^s - \dots - \Theta_Q B^{sQ})$ es el operador media móvil estacional de orden Q ,

$\theta_q(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$ es el operador media móvil regular de orden q , y ε_t es un proceso de ruido blanco.

2.4. Metodología de Box y Jenkins

La metodología de Box y Jenkins constituye un procedimiento estadístico muy potente para el tratamiento y modelización de variables dinámicas (Alpuim, 1998). Su desarrollo en el campo práctico ha condicionado a la evolución de los paquetes estadísticos de ordenador, dada la complejidad de cálculos que requiere y el volumen de datos que se suele manejar.

Las principales características de esta metodología son su generalidad (puede manejar prácticamente cualquier serie de tiempo) y su sólida base teórica.

Para construir un modelo ARIMA que recoja suficientemente bien las características de la serie se hace uso de la metodología de Box y Jenkins. La metodología descrita por Box y Jenkins, consiste en determinados procedimientos para poder obtener un proceso adecuado para la serie temporal objeto de estudio y puede ser estructurada en cinco etapas:

1. Análisis Exploratorio de la Serie

Consiste en el análisis del histograma de frecuencias, las principales medidas des-

criptivas, análisis de estacionariedad mediante gráficos (gráficos de las funciones de autocorrealción y autocorrelación parcial) y la prueba estadística de existencia de raíces unitarias de Dickey y Fuller (DF).

Prueba de raíces unitarias de Dickey Fuller (DF)

Una estrategia eficiente para probar la existencia de una raíz unitaria (no estacionariedad) en una serie, es probar la hipótesis nula de un proceso con raíz unitaria y una constante versus un proceso estacionario alrededor de una tendencia.

La prueba de Dickey y Fuller se basa en asumir que la serie se puede aproximar por un proceso AR(1). Inicialmente se asume que y_t sigue un modelo AR(1) y se procede a transformar el modelo de la siguiente manera:

$$\nabla y_t = \rho y_{t-1} + \varepsilon_t$$

Donde la existencia de raíz unitaria (no estacionariedad) equivale a $\rho = 0$. Por tanto se tiene las hipótesis:

$$H_0 : \rho = 0$$

$$H_1 : \rho < 0$$

El estadístico de prueba es denotado por τ y su distribución bajo H_0 permite calcular los valores críticos, de tal forma que el criterio de rechazo es $\hat{\tau} < \tau_\alpha$ donde $\hat{\tau}$ es el valor calculado del estadístico.

2. Identificación del Modelo

En esta etapa se debe sugerir un conjunto reducido de posibles modelos.

- Selección del conjunto de estimación: Conjunto de datos que se usará para la estimación y adecuación del modelo; y del conjunto de predicción: conjunto de datos que se guardará para evaluar las predicciones.
- Determinación de los correlogramas o funciones de autocorrelacion (FAC) y funciones de autocorrelación parcial (FACP) para establecer, conjuntamente con lo observado en la primera etapa, el número de diferencias que se aplicarán y que convertirán el proceso en estacionario. En caso de presentar estacionalidad con periodo s , se aplica una diferencia estacional $(1 - B)^s$ para convertir la serie en estacionaria.

- Determinación de los órdenes del componente autorregresivo (p) y promedio móvil (q) del modelo $ARMA(p, q)$, haciendo uso de los patrones que se observan en los correlogramas, FAC y FACP:
- Especificación del modelo $ARIMA$ identificado y se sugieren otros modelos posibles.

3. Estimación de parámetros y selección del Modelo

Una vez identificado el modelo, los valores de los parámetros ϕ y θ se estiman mediante la minimización de la suma de cuadrado de los errores ε_t . Estas estimaciones se hacen con la ayuda de herramientas computacionales como Minitab, R y EViews, SPSS, etc.

La selección del mejor modelo para la serie analizada se realiza mediante los criterios de Akaike y Schwarz.

Criterio de información de Akaike (AIC)

EL AIC es uno de los criterios comunmente usados para la selección de modelos, este es una medida estadística de verosimilitud del conjunto de modelos a incluir, penalizado por el número de parámetros de los modelos. (Peña, 2010). El criterio de información de Akaike es definido como:

$$AIC = -2\log(L) + 2p$$

Donde L es la función de máxima verosimilitud del modelo ajustado y p es el número de parámetros del modelo. El menor valor del AIC, significa el mejor modelo ajustado.

Criterio de Schwarz (BIC)

Peña, Tiao, y Tsay (2001) mencionan que el término de penalización del AIC, $2p$, no depende del tamaño muestral “ n ” de la población considerada. Esto conduce al hecho de que un mismo número de parámetros comunes es seleccionado mediante el AIC, tanto para muestras pequeñas como para muestras grandes. Es por ello que el AIC no es un estimador consistente del número adecuado de parámetros. Schwarz (1979) sugirió que el AIC podría no ser asintoticamente justificable, y presentó un criterio de información alternativo a partir de un enfoque bayesiano, el

BIC (Bayesian Information Criterion). Con este criterio, se penaliza el número de parámetros con $\ln(n)$, en lugar de 2. Así,

$$BIC = -2\log(L) + \log(n)p$$

Donde L es la función de máxima verosimilitud del modelo ajustado, p es el número de parámetros del modelo y “ n ” es el tamaño muestral. De acuerdo con este criterio, el modelo que indique el menor valor de BIC será el mejor modelo ajustado.

4. Adecuación del Modelo

Una idea, a priori, de la adecuación del modelo se logra graficando la serie original y la ajustada por el modelo ARIMA contra el tiempo, de manera que se puedan observar sus similitudes y diferencias. Un modelo ARIMA es adecuado para representar el comportamiento de una serie si se cumple lo siguiente:

- Los residuos, diferencia entre el valor original de la serie y el valor estimado por el modelo, se aproximan el comportamiento de un ruido blanco. Al observar los correlogramas no deberán observarse valores significativamente diferentes de cero, como indicativo de ausencia de correlación serial, así como tampoco patrones (tendencia, ciclos) que indicarán que el modelo no extrajo toda la información posible. Así también se realiza la prueba de autocorrelaciones de Ljung Box (Q) para los residuales del model.

Prueba de Ljung-Box (Q)

La prueba de Ljung y Box es una prueba para las autocorrelaciones de series de tiempo. Que a pesar de no detectar caídas específicas en el comportamiento de ruido blanco, puede indicar si esos valores son muy altos (Morettin y Toloi, 2011).

$$H_0 : \rho(1) = \rho(2) = \dots = \rho(k) = 0$$

$$H_1 : \text{por lo menos un } \rho(k) \neq 0.$$

Con el estadístico de prueba:

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

Donde n es el tamaño de la serie y $\hat{\rho}_j$ es el estimador de la j -ésima autocorrelación. La estadística Q tendrá aproximadamente una distribución χ^2 con k grados de libertad, se rechaza la hipótesis nula si $Q(k) \geq \chi_k^2$.

- Los parámetros del modelo ARIMA seleccionado son significativamente diferentes de cero.
- El grado de ajuste es elevado en comparación al de otros modelos alternativos. La bondad del ajuste es evaluado con el Criterio de Información de Akaike (AIC) y de información de Schwarz (BIC).

5. Predicción

Una vez identificado, estimado y validada la consistencia del modelo generado por la serie y_t , el paso siguiente consiste en efectuar predicciones de valores futuros para la referida serie temporal. Las medidas de validación indican si la metodología de predicción es apropiada (errores distribuidos aleatoriamente). De esta manera, Marques (2013) menciona que en la práctica, debe ser seleccionado el método que produce menores errores de predicción sobre las observaciones conocidas de la serie observada. Las medidas para la validación del desempeño de los modelos son el principal criterio de selección de un método de predicción.

En la literatura existen varias medidas de error como: El error medio absoluto (EMA), Error cuadrático medio (ECM), error porcentual medio absoluto (EPMA), entre otros.

Sea y_t , el valor observado de la serie en el momento t , y $\hat{y}_t$ la predicción a un paso al frente para el momento t , es decir, la predicción con origen en el momento $t - 1$ y horizonte temporal de un período, los criterios de los errores de predicción utilizados se basan en las funciones siguientes:

■ Error Cuadrático Medio (ECM)

Valor medio de los desvíos al cuadrado entre los valores observados y las predicciones para los instantes $1, 2, \dots, m$

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

■ Error Medio Absoluto (EMA)

Valor absoluto medio de los desvíos entre los valores observados y las predicciones para los instantes $1, 2, \dots, m$

$$EMA = \frac{1}{m} \sum_{t=1}^m |Y_t - \hat{Y}_t|$$

■ **Error Porcentual Medio Absoluto (EPMA)**

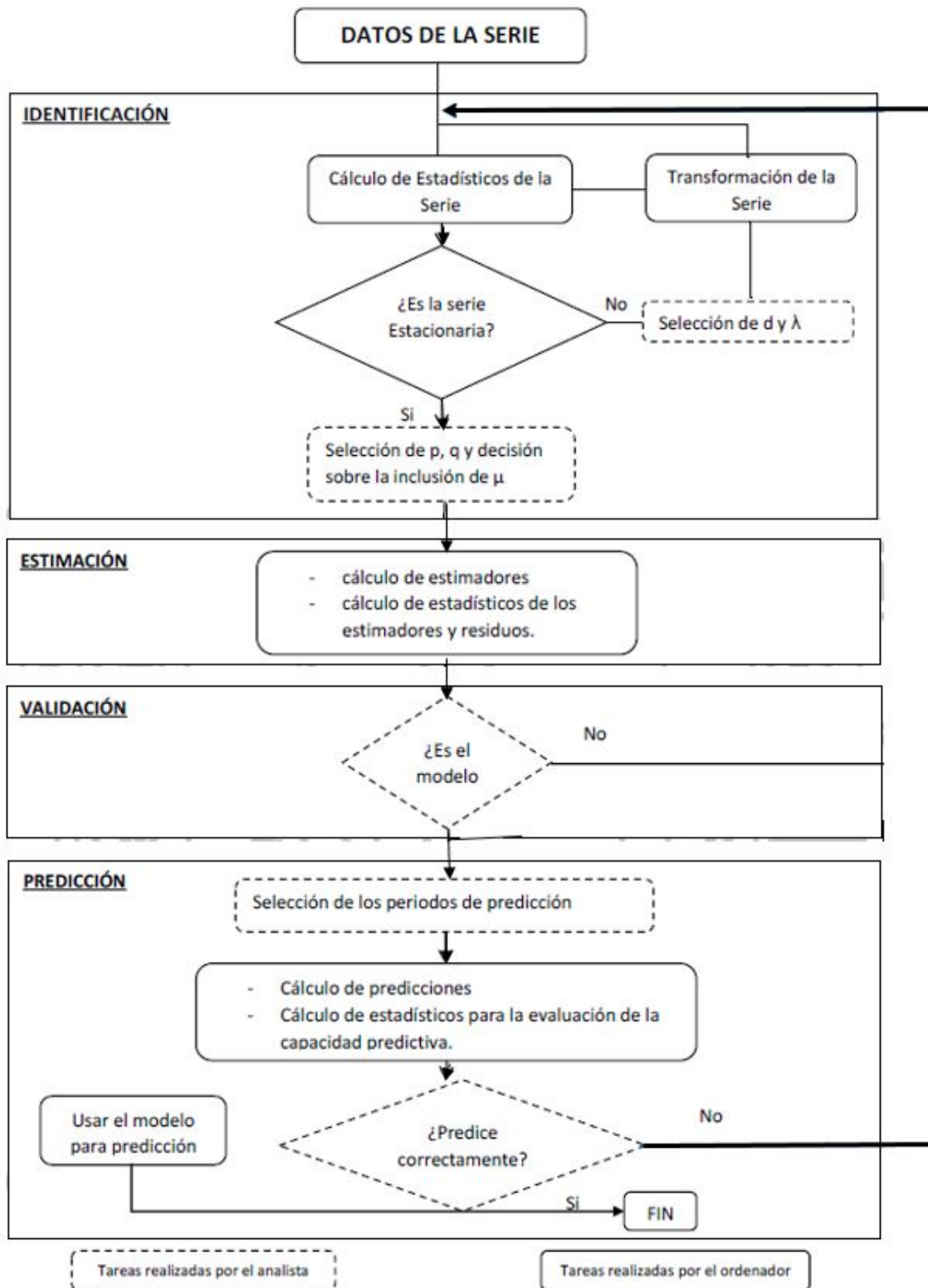
Valor porcentual medio absoluto entre los valores observados y las predicciones para los instantes $1, 2, \dots, m$

$$EPMA = \frac{1}{m} \sum_{t=1}^m \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100 \%$$

La selección del modelo más adecuado implica un análisis de las medidas de validación presentadas.

En la figura 01 se presenta el algoritmo de la metodología presentado por Box y Jenkins.

Figura 1: Algoritmo de la Metodología de Box y Jenkins.



Fuente: Murteira y Muller (1990)

2.5. Metodología Bootstrap

2.5.1. Introducción

El primero en considerar la técnica del bootstrap de una manera sistemática fue Efron y Tibshirani (1994). El nombre "bootstrapping" fue usado por Efron para seguir las ideas de Tukey (1958) quien usó el nombre de "Jackknifing" para describir su método de remuestreo. Desde entonces, se ha convertido rápidamente en una popular y poderosa herramienta estadística usada para problemas difíciles en el análisis estadístico.

El bootstrap constituye la técnica más versátil y conocida dentro del método de remuestreo, su importancia radica en que es un método para estimar la distribución de una estadística con muestras finitas. Fernández (1999) menciona que debido a que el sustento teórico matemático - estadístico del bootstrap es bastante complejo, hasta finales de la década del 80', la eficiencia del método era probada de manera empírica, es decir, en el terreno de la práctica. Este método es computacionalmente intenso, sin embargo, las modernas computadoras son más que suficientes para los requerimientos computacionales requeridos para éste método.

Los métodos bootstrap son robustos debido a que no asumen ninguna distribución particular de la población y han sido aplicados a diversos problemas de muestreo, sustituyendo la complejidad del análisis teórico por el poder de cálculo de los ordenadores. (Efron y Tibshirani, 1994).

La esencia del "bootstrapping" es la idea que, en la ausencia de cualquier otro conocimiento sobre la distribución de una población, la distribución de valores encontrada en una muestra aleatoria de tamaño n de la población es la mejor información para aproximar su distribución de probabilidades. (Alvarez y Rubio, 2000).

2.5.2. Principio del Bootstrap

En la formulación del Bootstrap inicialmente propuesto por Efron (1979), se considera $X_1, X_2, \dots, X_n$ una secuencia de variables aleatorias i.i.d con distribución común F desconocida y de valor medio μ y de varianza σ^2 finitos, pero desconocidos.

Sea $(x_1, x_2, \dots, x_n)$ una realización de $(X_1, X_2, \dots, X_n)$. Una muestra $(x_1, x_2, \dots, x_n)$ que va a desempeñar el papel de la "población Bootstrap", de esta vamos a extraer con reposición,

un grande número, B , de muestras, que designamos de “muestras Bootstrap” o “remuestras Bootstrap”. El objetivo es estimar la función de distribución de la población F o algún aspecto particular de esta.

Así se comienza por construir la función de distribución empírica $\hat{F}_n$, estimador de F , definida por:

$$\hat{F}_n(x) = \frac{\#\{i : X_i \leq x\}}{n}$$

Esta función se obtiene a partir de los valores observados, siendo atribuido el peso $\frac{1}{n}$ a cada una de las observaciones. Con $\hat{F}_n$ fijo, definimos una muestra Bootstrap $(x_1^*, x_2^*, \dots, x_n^*)$, donde el símbolo $*$ indica que hubo remuestreo con reposición de la muestra inicial.

La metodología Bootstrap se basa exactamente en la idea de producir un grande número de muestras a partir de la muestra original (población Bootstrap) y calcular el valor de la estadística de interés para cada una de ellas, y así estimar la distribución muestral del estadístico. Los resultados pueden generalizarse para la población. La ventaja del bootstrap es que no necesitamos conocer el proceso generador de los datos.

La aproximación bootstrap es válida para la mayoría de las estadísticas las cuales son consistentes. Si $B \rightarrow \infty$ entonces las estimaciones del error estándar, del sesgo y del error cuadrático medio se igualan a las estimaciones de máxima verosimilitud. (Efron y Tibshirani, 1994).

Para el cálculo de las estimaciones bootstrap es necesario un valor de B muy grande, generalmente a partir de $B=1000$ genera buenos resultados. (Efron y Tibshirani, 1994).

2.5.3. Aplicaciones del Bootstrap

Como el resultado del proceso descrito en líneas arriba, se pueden derivar al menos 04 importantes y más usadas aplicaciones prácticas, detalladas a continuación:

a. Estimación Bootstrap del Error Típico

El bootstrap fue introducido por Efron y Tibshirani (1994) como un método basado en cálculos intensivos mediante un ordenador para estimar el error muestral de un estadístico. Explícitamente, la estimación bootstrap del error de un estadístico es como sigue:

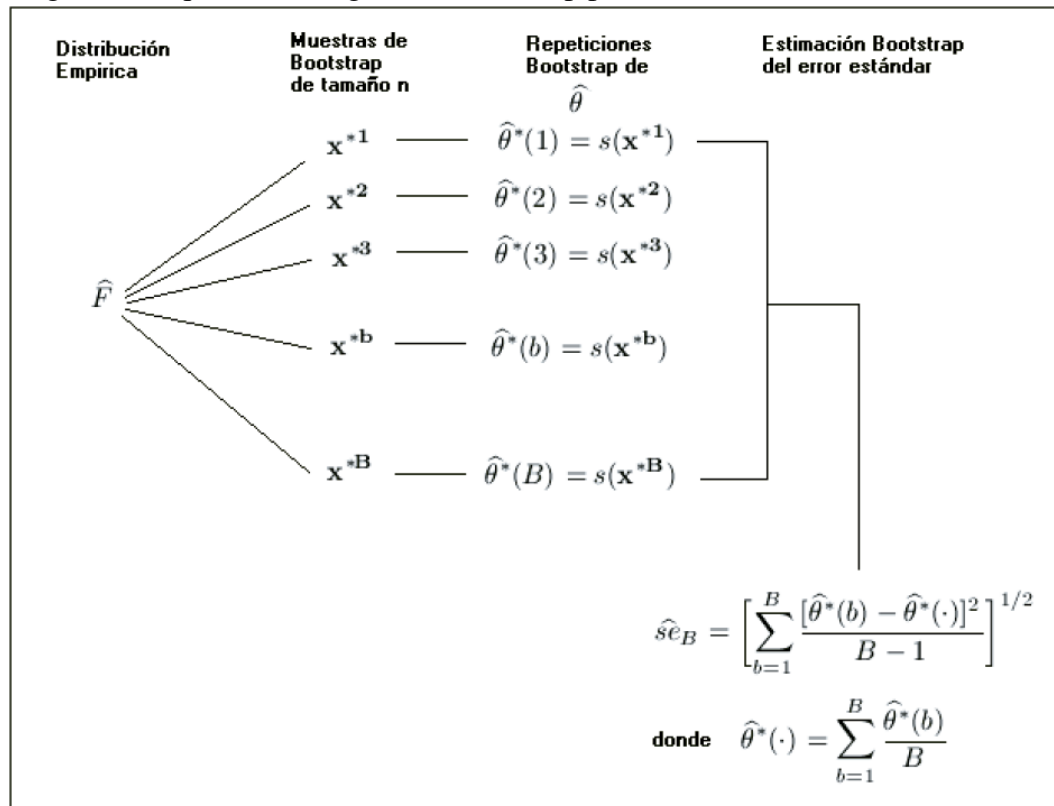
- Se extrae B muestras bootstrap independientes de la función de distribución empírica $\hat{F}$.
- Se computa el estadístico de interés en cada una de las B muestras, obteniendo $\hat{\theta}_b^*$.
- Se estima el error muestral de $\hat{\theta}$ mediante la desviación estándar de la función de distribución obtenida de los B $\hat{\theta}_i^*$, es decir a través de $\hat{F}^*(\hat{\theta}^*)$.

$$\hat{se}_B = \left\{ \sum_{b=1}^B \frac{[\hat{\theta}_b^* - \hat{\theta}^*(\cdot)]^2}{(B-1)} \right\}^{\frac{1}{2}}$$

Donde: $\hat{\theta}^*(\cdot) = \sum_{b=1}^B \frac{\hat{\theta}_b^*}{B}$

Este procedimiento se puede observar claramente en la figura 2, donde se muestra el algoritmo bootstrap para la estimación del error estándar de un estadístico $\hat{\theta} = s(x)$; cada muestra bootstrap es una variable aleatoria independiente de tamaño n de $\hat{F}$, y B es el número de replicaciones bootstrap.

Figura 2: Esquema del Algoritmo Bootstrap para la estimación del error estándar



Fuente: Efron y Tibshirani (1994).

b. Intervalo de Confianza Bootstrap

Miranda (2003), describe que existen 03 métodos frecuentemente utilizados a través de los cuales se pueden construir intervalos de confianza bootstrap y son:

- Método de aproximación normal
- Método de los percentiles
- Método del Bootstrapping estudentizado

El método de aproximación normal, utiliza la misma estructura de los procedimientos paramétricos de la construcción de intervalos de confianza. Si es posible asumir que el estadístico se distribuye según la curva normal pero el cálculo del error típico resulta analíticamente difícil o no existe fórmula para su cálculo, entonces podemos emplear la distribución muestral bootstrap para estimar el error típico o insertarlo en la correspondiente expresión del intervalo de confianza paramétrico. Así el intervalo de confianza bootstrap será de la forma

$$[\hat{\theta} - \hat{se}_R z_{1-\frac{\alpha}{2}}; \hat{\theta} + \hat{se}_R z_{1-\frac{\alpha}{2}}]$$

Donde: $\hat{se}_R = \sqrt{\frac{\sum_{b=1}^B [\hat{\theta}^*(b) - \hat{\theta}^*(\cdot)]^2}{B-1}}$, es el error estándar obtenido y $\hat{\theta}^*(\cdot) = \frac{\sum_{b=1}^B \hat{\theta}^*(b)}{B}$.

El método del percentil, hace uso literal de la idea básica del bootstrap, es decir $\hat{F}^*(\hat{\theta}^*)$ se aproxima a $F(\hat{\theta})$. La idea es muy simple: un intervalo con un nivel de confianza $1 - \alpha$ incluye todos los valores de $\hat{\theta}^*$ entre los percentiles $\frac{\alpha}{2}$ y $(1 - \frac{\alpha}{2})$ de la distribución de $\hat{F}^*(\hat{\theta}^*)$.

En este caso, el intervalo está dado de la forma:

$$[\hat{F}^{-1}(\alpha/2); \hat{F}^{-1}(1 - \alpha/2)]$$

Donde $\hat{F}^{-1}(\alpha/2)$ representa el percentil de $\alpha/2$ de la distribución del estimador $\hat{\theta}^*$, es decir un valor tal que la probabilidad acumulada hasta dicho valor sea $\alpha/2$, y $\hat{F}^{-1}(1 - \alpha/2)$ representa el percentil del $1 - \alpha/2$.

El método del percentil conserva la esencia no-paramétrica del enfoque bootstrap y libera al usuario de las asunciones de la estadística paramétrica.

El método del Bootstrapping estudentizado, La idea aquí es sustituir los percentiles de la distribución t por los percentiles de la distribución de los valores estudentizados

del estadístico $\hat{\theta}$ en la muestras bootstrap. Es decir de los valores definidos por:

$$t^*(B) = \frac{\hat{\theta}^*(B) - \hat{\theta}}{\hat{se}(\hat{\theta}^*(B))}$$

Donde $\hat{\theta}^*(B)$ y $\hat{se}(\hat{\theta}^*(B))$ son los valores del estimador y valor estimado del error del estimador en la B-ésima muestra bootstrap respectivamente; $\hat{\theta}$ es el estimado en la muestra original.

Luego, el intervalo de confianza del $100(1 - \alpha)\%$ por bootstrapping estudentizado estará dado por:

$$(\hat{\theta} + \hat{F}_t^{-1}(\frac{\alpha}{2})\hat{se}; \hat{\theta} + \hat{F}_t^{-1}(1 - \frac{\alpha}{2})\hat{se})$$

Donde $\hat{F}_t^{-1}(\frac{\alpha}{2})$ representa el percentil de $\frac{\alpha}{2}$ de la distribución de t^* , es decir un valor tal que la probabilidad acumulada hasta dicho valor sea $\frac{\alpha}{2}$ y $\hat{F}_t^{-1}(1 - \frac{\alpha}{2})$ representa el percentil del $1 - \frac{\alpha}{2}$.

c. Prueba de Hipótesis Bootstrap

Sea B el número de remuestras independientes de tamaño n de $\hat{F}$. Para cada remuestra $b, b = 1, \dots, B$ es calculado el valor de la estadística de prueba t_b^* rechazando H_0 cuando $|t_b^*| \geq c$ entonces el p-valor bootstrap utilizado puede ser aproximado por:

$$p_{boot} = \frac{1 + \#\{|t_b^*| \geq |t_{obs}|\}}{B + 1}$$

Utilizando este procedimiento para obtener el p-valor bootstrap, diferentes pruebas de hipótesis pueden ser realizados, como por ejemplo probar la hipótesis de diferencias de medias, probar la hipótesis de distribuciones. Entonces para probar esta hipótesis utilizando el procedimiento bootstrap descrito encima, tenemos que seguir los siguientes pasos:

1. Obtener B remuestras bootstrap de tamaño $n + m$, las primeras n observaciones son llamadas de z^* y las siguientes de y^*
2. Calcular el estadístico de prueba para cada una de las B remuestras

$$t_b^* = \bar{z}^* + \bar{y}^*, b = 1, \dots, B.$$

3. Obtener el p-valor bootstrap.

$$p_{boot} = \frac{1 + \#\{|t_b^*| \geq |t_{obs}|\}}{B + 1}$$

donde t_{obs} es el estadístico de prueba obtenida de la muestra original.

d. Modelo de Regresión Bootstrap

El análisis de modelos de regresión ocupa un papel muy importante en el análisis estadístico. Se estudian los efectos de variables explicativas o covariables sobre una variable respuesta. El modelo de regresión lineal es descrito en notación vectorial de la siguiente manera:

$$y_i = x_i^t \beta + \varepsilon_i, \quad i = 1, 2, \dots, n$$

con $x_i^t = (1, x_{i1}, x_{i2}, \dots, x_{i(p-1)})$ y $\beta^t = (\beta_0, \beta_1, \dots, \beta_{p-1})$

Los pasos a seguir para estimar el coeficientes de regresión β bootstrap son:

1. Calcular los estimadores MCO de β , $\hat{\beta}$, y estimar los residuales del modelo $\hat{\varepsilon}_i = y_i - x_i^t \hat{\beta}$.
2. Muestrear con reemplazo n errores $\hat{\varepsilon}_i$ desde $\hat{\varepsilon}_i$ y formar las n observaciones muestrales bootstrap desde $y_i^* = x_i^t \hat{\beta} + \hat{\varepsilon}_i^*$, para $i = 1, \dots, n$.
3. A partir de y_i^* y obtenidas desde la muestra (y^*, x) , calcular y guardar los estimadores MCO, $\hat{\beta}_{(b)}^*$.
4. Repetir los pasos 2 y 3, B veces, obteniendo $\hat{\beta}_{(1)}^*, \hat{\beta}_{(2)}^*, \dots, \hat{\beta}_{(B)}^*$

De tal forma que ahora ya se puede proseguir con la estimación de las estadísticas y estadísticas de prueba que se deseen.

2.5.4. Bootstrap en Series de Tiempo y Trabajos Relacionados

Stine (1987) propuso el primer método bootstrap para estimar el error cuadrático medio del predictor lineal clásico cuando el proceso es un modelo autorregresivo de orden p (AR(p)). El método se basó en la obtención de la función de distribución empírica de los errores centrados. A partir de ésta extraía muestras para reproducir un proceso AR(p) con los parámetros estimados inicialmente y fijando los primeros p valores de la serie. Dada la serie bootstrap se estiman los parámetros y se obtiene el predictor clásico para evaluar el error cuadrático medio de dicha predicción. Repitiendo un número elevado de veces este proceso podemos aproximar el error cuadrático medio de la predicción puntual.

Thombs y Schucany (1990) propusieron la construcción de intervalos de predicción en modelos AR(p), propusieron realizar una representación hacia atrás del proceso para pos-

teriormente utilizar la representación habitual para obtener los valores futuros. El esquema de dicho método es el siguiente:

1. Construir los residuales de la representación backward y calcular la función de distribución empírica de dichos errores centrados y reescalados según propuso Stine (1987).

2. Extraer los errores bootstrap, ε_i^* a partir de dicha distribución.

3. Se define la serie backward de la serie fijando los últimos p valores observados

$$Y_i^* = \hat{\phi}_1 Y_{i+1}^* + \dots + \hat{\phi}_p Y_{i+p}^* + \varepsilon_i^*, i = t-p, t-p-1, \dots, 1$$

$$Y_i^* = Y_i, i = t-p+1, \dots, t$$

4. Calcular los estimadores de la serie bootstrap.

5. Construir los residuales hacia adelante de la serie y obtener la función de distribución empírica de dichos errores centrados y reescalados.

6. Extraer los errores bootstrap, a_i^* , a partir de dicha distribución.

7. Definir las réplicas bootstrap de los valores futuros

$$Y_{t+i}^* = \hat{\phi}_1 Y_{t+i-1}^* + \dots + \hat{\phi}_p Y_{t+i-p}^* + a_i^*, i = 1, \dots, s$$

$$Y_{t+i}^* = Y_{t+i}, \text{ si } i \leq 0$$

8. Repetir un número B elevado de veces los pasos 2-7 para aproximar la distribución de Y_{t+s}^* . Los cuantiles correspondientes a esa distribución serán los límites del intervalo de predicción.

Estos autores demostraron la validez asintótica de este método bootstrap y realizaron un estudio de simulación donde comparan con los intervalos clásicos para diferentes distribuciones y coberturas.

McCullough (1994) sugirió que el método percentil de obtener los intervalos de predicción no era el más adecuado cuando existe asimetría en el error de predicción y, por lo tanto, propuso un método de corrección de sesgo. Además, advirtió que la representación backward utilizada por Thombs y Schucany (1990) no asegura que los errores sean independientes e idénticamente distribuidas y utilizó un procedimiento más sofisticado para obtener dichos errores hacia atrás.

La validez asintótica del método de Thombs y Schucany (TS) para modelos AR fue demostrada por Jurado y cols. (1995). La idea general se basa en obtener la serie diferenciada y aplicar el método TS para construir la réplica bootstrap de la serie que tiene la estructura autoregresiva. Posteriormente, se integra para tener réplicas bootstrap de la serie original.

Un inconveniente importante del método TS consiste en la representación backward de la serie, no sólo por el gasto computacional sino porque no es fácil cuando los procesos no tienen distribución normal. Cao y cols. (1997) proponen un método alternativo, que es consistente, para construir intervalos de predicción sin tener que realizar réplicas hacia atrás de la serie. El método se puede describir en los siguientes pasos

1. Obtener los residuos forward centrados y reescalados para construir la función de distribución empírica.
2. Extraer una muestra aleatoria, a_i^* , a partir de dicha distribución.
3. Construir los valores futuros bootstrap utilizando como errores la muestra anterior y las estimaciones iniciales de los parámetros.

$$Y_{t+i}^* = \hat{\phi}_1 Y_{t+i-1}^* + \dots + \hat{\phi}_p Y_{t+i-p}^* + a_i^*, i = 1, \dots, s$$

$$Y_{t+i}^* = Y_{t+i}, \text{ si } i \leq 0$$

4. Repetir las etapas anteriores un número elevado de veces para aproximar la distribución bootstrap del valor Y_{t+s}^* .

Cao y cols. (1997) demostraron la consistencia de este método tanto en el caso de utilizar la función de distribución empírica de los residuos directamente como en el de utilizar un bootstrap suavizado, es decir, cuando en la etapa 2 se extrae la muestra a partir de una suavización no paramétrica de la distribución.

Finalmente, cabe destacar que Pascual y cols. (1999) propusieron un método que capta el error de predicción debido a la incertidumbre en la estimación pero que no es necesaria realizar una representación backward de la serie. El esquema de remuestreo para un modelo AR(p) se describe bajo las siguientes etapas

1. Obtener los residuos forward centrados y reescalados para construir la función de distribución empírica.
2. Extraer una muestra aleatoria, a_i^* , a partir de dicha distribución.

3. Generar una serie con la misma estructura AR(p) utilizando los errores bootstrap y estimar los parámetros bootstrap de esta serie.
4. Construir los valores bootstrap futuros utilizando como errores una muestra obtenida como en la etapa 2 y las estimaciones bootstrap de los parámetros. En esta etapa se fijan los últimos p valores de la serie original para construir los valores bootstrap futuros.
5. Repetir las etapas anteriores un número elevado de veces para aproximar la distribución bootstrap del valor Y_{t+s}^* .

2.5.5. Bootstrap en Modelos ARIMA

A continuación se muestra el procedimiento de Thombs y Schucany (1990) y Pascual, Romo, y Ruiz (2006) para utilizar la verdadera distribución de los residuales y tener en cuenta la incertidumbre por la estimación de los parámetros para un modelo ARIMA.

- Estimar el modelo y obtener los parámetros, $\hat{\phi}$ y $\hat{\theta}$ y los residuos.
- Obtener una muestra con reemplazamiento de t residuos de la distribución de los residuos estimados. Sean $\varepsilon_1^*, \varepsilon_2^*, \dots, \varepsilon_T^*$ la muestra obtenida.
- Reconstruir una nueva serie con los parámetros estimados y estos residuos mediante:

$$y_t^* = \hat{\phi}_1 y_{t-1}^* + \dots + \hat{\phi}_h y_{t-h}^* + \varepsilon_t^* - \hat{\theta}_1^* \varepsilon_{t-1}^* - \dots - \hat{\theta}_q^* \varepsilon_{t-q}^*$$

tomando como valores iniciales $z_1 = \dots = z_h$ valores originales y $\varepsilon_{-q} = \dots = \varepsilon_0$ residuales del modelo estimado.

- Estimar los parámetros en la serie anterior. Sean $\hat{\phi}^*$ y $\hat{\theta}^*$ los nuevos parámetros estimados.
- Calcular las predicciones de la serie, utilizando los parámetros estimados en la etapa anterior, $\hat{\phi}^*$ y $\hat{\theta}^*$, pero con valores de y iguales a los últimos verdaderos valores observados y residuales iguales a las reales.
- Calcular el error porcentual medio absoluto $EPMA_b^*$
- Repetir los pasos 2 a 5, B veces.

- Estimar el error porcentual medio absoluto bootstrap mediante:

$$EPMA_B = \sum_{b=1}^B \frac{EPMA_b^*}{B}$$

2.6. Hipótesis de la investigación

2.6.1. Hipótesis General

El comportamiento de los ingresos se adapta a un modelo serial estacionario, Hotel El Centro, Huamanga - Ayacucho en el periodo 2011 al 2014.

2.6.2. Hipótesis Específicas

- Las predicciones de los ingresos mensuales del Hotel El Centro Huamanga - Ayacucho en el periodo 2011 - 2014, mediante la metodología de Box y Jenkins presentan una adecuada desviación absoluta media porcentual.
- Las predicciones de los ingresos mensuales del Hotel El Centro Huamanga - Ayacucho en el periodo 2011 - 2014, mediante la metodología Bootstrap presentan una adecuada desviación absoluta media porcentual.
- El comportamiento serial de los ingresos generados por el modelo de predicción Bootstrap posee una menor desviación absoluta media porcentual, que el modelo de predicción de Box y Jenkins, en el Hotel El Centro, Huamanga - Ayacucho en el periodo 2011 - 2014.

3. METODOLOGÍA DE INVESTIGACIÓN

3.1. Metodología

3.1.1. Tipo de investigación

El presente estudio reúne las condiciones metodológicas de una investigación aplicada, pues se utilizaron técnicas estadísticas a fin de aplicarlas en la predicción de los ingresos mensuales del Hotel el Centro.

3.1.2. Nivel de investigación

La presente investigación es de nivel descriptivo inferencial y explicativo de corte longitudinal. Descriptivo inferencial de corte longitudinal consistió en observar el comportamiento de la variable en estudio, y realizar una estimación de los parámetros del modelo de serie temporal, y explicativo, pues se planteó obtener un modelo de pronóstico adecuado para la variable en estudio.

3.1.3. Diseño de investigación

El diseño del presente estudio es de tipo no experimental, pues se efectuó la recolección de los datos en períodos especificados y con una sola observación en el tiempo.

3.2. Universo

Conformado por el total de los meses desde la apertura del Hotel el Centro, Huamanga, Ayacucho, hasta diciembre del 2014.

3.3. Población

Conformado por los ingresos mensuales en Nuevos Soles desde su apertura hasta diciembre del 2014, del Hotel el Centro, Huamanga, Ayacucho.

3.4. Muestra

3.4.1. Muestra de entrenamiento

Conformado por ingresos mensuales en Nuevos Soles de enero 2011 a diciembre 2014 del Hotel el Centro, Huamanga, Ayacucho. Con un total de 48 observaciones.

3.4.2. Muestra de prueba

Conformado por los ingresos mensuales en Nuevos Soles del año 2015 del Hotel el Centro, Huamanga, Ayacucho. Con un total de 12 observaciones.

3.5. Unidad de análisis

Conformado por cada mes en observación en el Hotel el Centro, Huamanga, Ayacucho de Enero 2011 a diciembre 2014.

3.6. Definición de las variables de estudio

La variable utilizada en esta investigación es dado por los ingreso mensual en Nuevos Soles, Hotel el Centro, Huamanga, Ayacucho periodo 2011-2014. Representado por y_t para $t = 1, 2, \dots, 48$.

3.7. Métodos

3.7.1. Técnica de recolección datos y procesamiento de la información

Instrumento indirecto de recolección de datos de enfoque cuantitativo, técnica del análisis de contenido, los datos fueron obtenidos del libro físico de registro de huéspedes del Hotel El Centro, Huamanga, Ayacucho periodo 2011 al 2014. ver Anexo B.

El procesamiento de la información se realizó mediante los paquetes: IBM SPSS 20, Software RStudio versión 0.98.873, Eviews 8.0 y Microsoft Excel 2013.

3.7.2. Metodos de análisis de datos

Para el presente trabajo de investigación primeramente se hizo uso de la metodología de Box y Jenkins en las series de tiempo de los denominados modelos ARIMA, y enseguida el análisis serial mediante la metodología Bootstrap. Los pasos a seguir en la obtención del modelo para los ingresos mensuales del Hotel el Centro periodo 2011 - 2014 mediante la metodología de Box y Jenkins fueron:

1. Análisis exploratorio e identificación de la serie

En la que se realizaron los siguientes pasos:

- Representación gráfica de la serie.
- Histograma de frecuencias.
- Test de estacionariedad de Dickey y Fuller.

2. Estimación y selección del modelo

- Se realizó la estimación de los parámetros mediante el método de los mínimos cuadrados.
- El mejor modelo de la serie estudiada fue obtenida mediante la comparación del criterio de Akaike (AIC) y criterio de Schwarz (BIC).

3. Validación

Para la adecuación entre el modelo y los datos se analizaron los residuales de cada modelo sugerido, se observaron en que medida los residuos del modelo estimado se aproximan al comportamiento de un ruido blanco. Para esto se procedió a analizar lo siguiente:

- Funciones de autocorrelación (FAC) y funciones de autocorrelación parcial (FACP).
- Test de Ljung-Box (Q).
- Histograma de frecuencia de los residuales.
- Test de normalidad de Jarque y Bera.

4. Predicción

Se consideró como conjunto de datos de prueba a los ingresos mensuales para el

periodo 2015. Así se obtuvo la predicción para el año 2015 y el error porcentual medio absoluto (EPMA).

Las predicciones para los ingresos mensuales del Hotel El Centro mediante la metodología Bootstrap fueron basados en el algoritmo bootstrap presentados por Thombs y Schucany (1990) y Pascual y cols. (2006) para series de tiempo, detalladas en el capítulo 2.5. y mostradas en resumen a continuación:

1. Los parámetros estimados fueron obtenidos del modelo adecuado para la serie y_t , mediante la metodología de Box y Jenkins.
2. De igual manera los residuos estimados fueron obtenidos del modelo de Box y Jenkins, de éstos se obtuvieron muestras con reemplazamiento (muestras bootstrap) de tamaño n .
3. Se reconstruyó una nueva serie y_t^* para los periodos 2011 al 2014, con los parámetros estimados del paso 1 y los residuos obtenidos del paso 2.
4. Se realizó un modelamiento a la serie y_t^* con las mismas características que se obtuvieron con la metodología de Box y Jenkins inicial. Se estimaron los parámetros de la nueva serie y_t^* .
5. Considerando el conjunto de datos de prueba a los ingresos mensuales para el periodo 2015, se calculó el error porcentual absoluto medio (EPMA).
6. Se repitieron los pasos del 2 al 5 un número de B veces (muestras bootstrap).
7. Se calculó el valor bootstrap del error porcentual medio absoluto de la siguiente manera:

$$EPMA_B = \sum_{i=1}^B \frac{EPMA_i}{B}$$

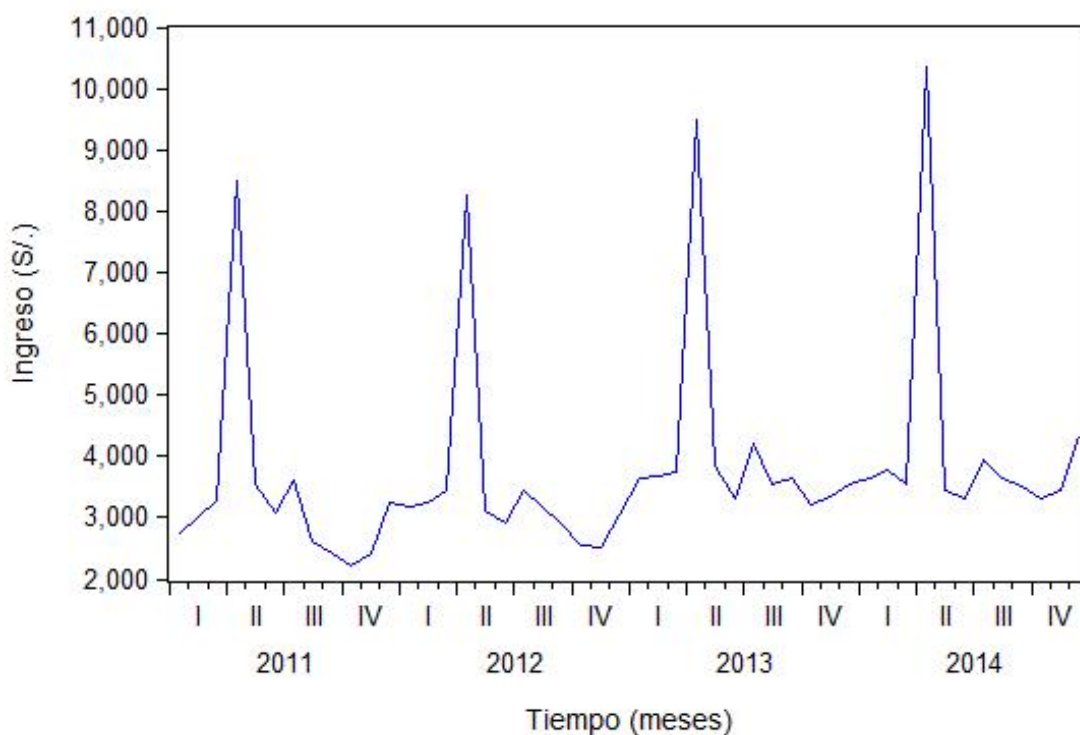
Finalmente se realizó una comparación entre ambas metodologías mediante los errores porcentuales medios absolutos $EPMA$ y $EPMA_B$, y con el mejor modelo obtenido de ambas metodologías se procedió a obtener los valores hacia adelante, para el caso de este estudio se obtuvieron las predicciones de los ingresos mensuales de enero a diciembre del año 2016.

4. RESULTADOS Y ANÁLISIS

4.1. Análisis exploratorio

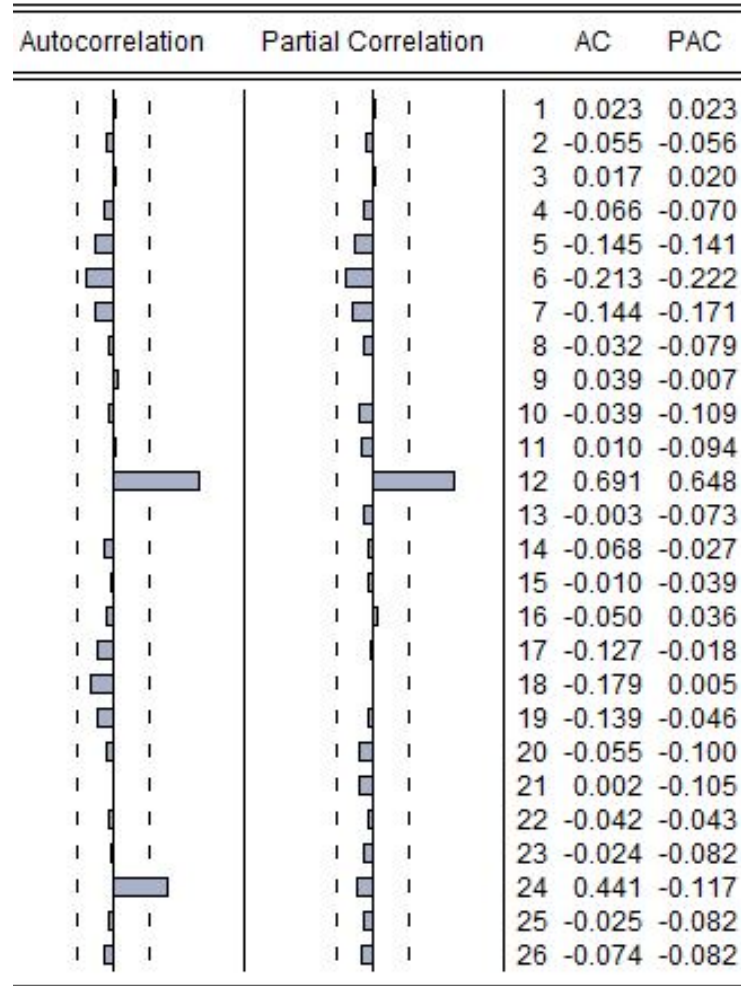
En la Figura 03 la evolución de la serie de los ingresos mensuales representado por y_t para $t = 1, 2, \dots, 48$, se observó que la serie presenta una ligera tendencia lineal creciente, es decir, el incremento de los ingresos mensuales cada año, y también la posible presencia de estacionalidad, debido a la existencia de patrones repetidos a lo largo de toda la serie.

Figura 3: Evolución del ingreso mensual en Nuevos Soles (S/.) del hotel El Centro en el periodo 2011 al 2014 (y_t).



Fuente: Elaboración propia

Figura 4: Correlogramas y valores de la Función de Autocorrelación (Autocorrelation) (AC) y Función de Autocorrelación Parcial (Partial Correlation) (PAC) del ingreso mensual del hotel El Centro periodo 2011 al 2014 (y_t).



Fuente: Elaboración propia

En la Figura 04 se observó, que los valores de las autocorrelaciones (AC) y autocorrelación parcial (PAC) para y_t mostraron valores altos en los rezagos 12 y 24, estos observados también en los autocorrelogramas (Autocorrelation y Partial Correlation), mostrando la presencia de estacionalidad en la serie de periodo $s=12$.

La presencia de no estacionariedad en y_t fue validada por la prueba de raíz unitaria de Dickey y Fuller, como sigue:

H_0 : Los ingresos mensuales (y_t) tiene raíz unitaria

H_1 : Los ingresos mensuales (y_t) no tiene raíz unitaria

Del Cuadro 1 se observó un valor del estadístico de Dickey y Fuller de $\tau = 1,581$, siendo este un mayor valor para valores críticos a niveles de significación de 0.01, 0.05 y 0.1, y con valor de probabilidad de 0.970, así no rechazando la hipótesis nula de presencia de raíz unitaria, lo cual quiere decir la serie y_t es no estacionaria.

Cuadro 1: Prueba de raíz unitaria de Dickey y Fuller para la serie ingreso mensual del hotel El Centro periodo 2011 al 2014 (y_t).

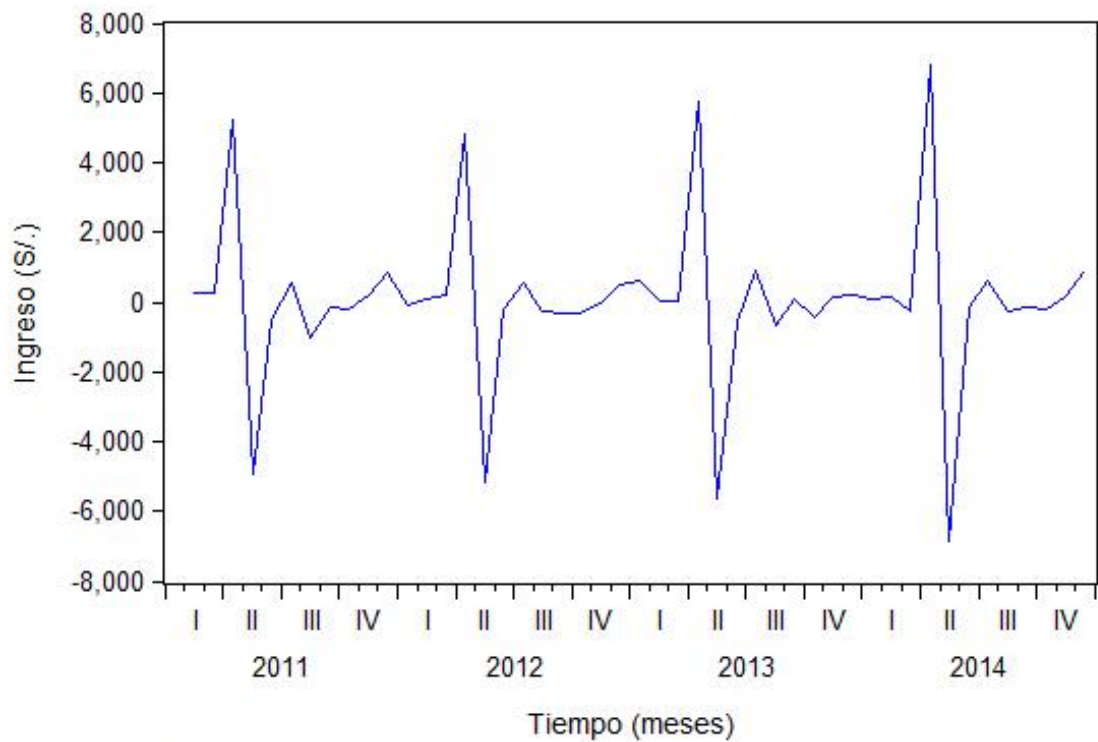
	τ	Prob.*
Estadístico de prueba de Dickey y Fuller	1.581	0.9696
Valores críticos: 1 % nivel	-2.633	
5 % nivel	-1.951	
10 % nivel	-1.611	

Fuente: Elaboración propia

Nota: resultado obtenido para el rezago 12, obtenido automaticamente mediante el Software Eviews 8 para un número máximo de rezagos de 16 y seleccionado mediante el criterio de información de Schwarz.

Por lo tanto, de las Figuras 03 y 04 y Cuadro 1 se determinó mediante un análisis gráfico, análisis de las autocorrelaciones y la prueba de estacionariedad de Dickey y Fuller para un nivel de confianza del 95 %, que la serie de los ingresos mensuales del Hotel el Centro durante el periodo 2011-2014 es no estacionaria, presentando ligera tendencia y estacionalidad en su estructura, dichos patrones de estacionalidad son muy pronunciados en el cuarto mes de cada año, este patrón es fácilmente deducido, pues los meses de abril el turismo en la ciudad de Ayacucho se incrementa debido a la fiesta de la semana santa. Dicho patrón se repite cada año, por tanto se consideró la serie de patrón estacional de 12 meses, es decir $s = 12$.

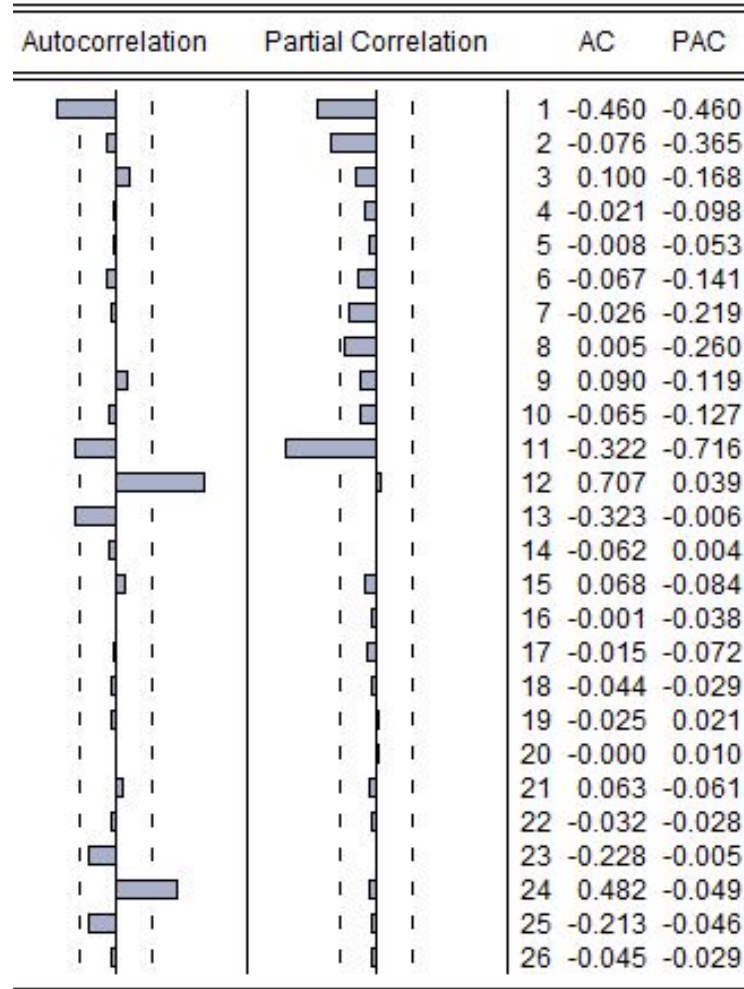
Figura 5: Evolución de la primera diferencia regular del ingreso mensual en Nuevos Soles (S/.) del hotel El Centro periodo 2011 al 2014 (z_t).



Fuente: Elaboración propia

Con el fin de transformar la serie y_t a una serie estacionaria se realizó una primera diferencia regular z_t de los ingresos mensuales y_t . En la Figura 05 se muestra la evolución de la primera diferencia de y_t denotada como $z_t = \nabla y_t = y_t - y_{t-1}$, para $t = 1, 2, \dots, 47$, donde se observó la ausencia de tendencia y presencia de picos altos en el cuarto mes de cada año, evidenciando una posible presencia de estacionalidad, por tanto la no estacionariedad de z_t .

Figura 6: Correlogramas y valores de la Función de Autocorrelación (Autocorrelation) (AC) y Función de Autocorrelación Parcial (Partial Correlation) (PAC); del hotel El Centro periodo 2011 al 2014 (z_t).



Fuente: Elaboración propia

Para comprobar si la serie transformada z_t es estacionaria se analizaron los correlogramas, las funciones de autocorrelación y la prueba de Dickey y Fuller, como sigue:

En la Figura 06 para la serie diferenciada z_t , se observó que en los valores de las autocorrelaciones (AC) mostraron valores altos en valor absoluto en los rezagos 1, 12 y 24, y en las autocorrelaciones parciales (PAC) en los rezagos 1, 2, 11. Observados también en los autocorrelogramas (Autocorrelation y Partial Correlation), evidenciando la no estacionariedad de z_t

La presencia de no estacionariedad en la serie fue validada por la prueba de raíz unitaria

de Dickey y Fuller, como sigue:

H_0 : Los ingresos mensuales (z_t) tiene raiz unitaria

H_1 : Los ingresos mensuales (z_t) no tiene raiz unitaria

del Cuadro 2 se observó un valor del estadístico de Dickey y Fuller de $\tau = -1,288$, siendo el valor, mayor para valores críticos a niveles de significación de 0.01, 0.05 y 0.1, y con valor de probabilidad de 0.178, así no rechazando la hipótesis nula de presencia de raiz unitaria.

Cuadro 2: Prueba de raiz unitaria de Dickey y Fuller para la serie con la primera diferencia regular del ingreso mensual del hotel El Centro periodo 2011 al 2014 (z_t).

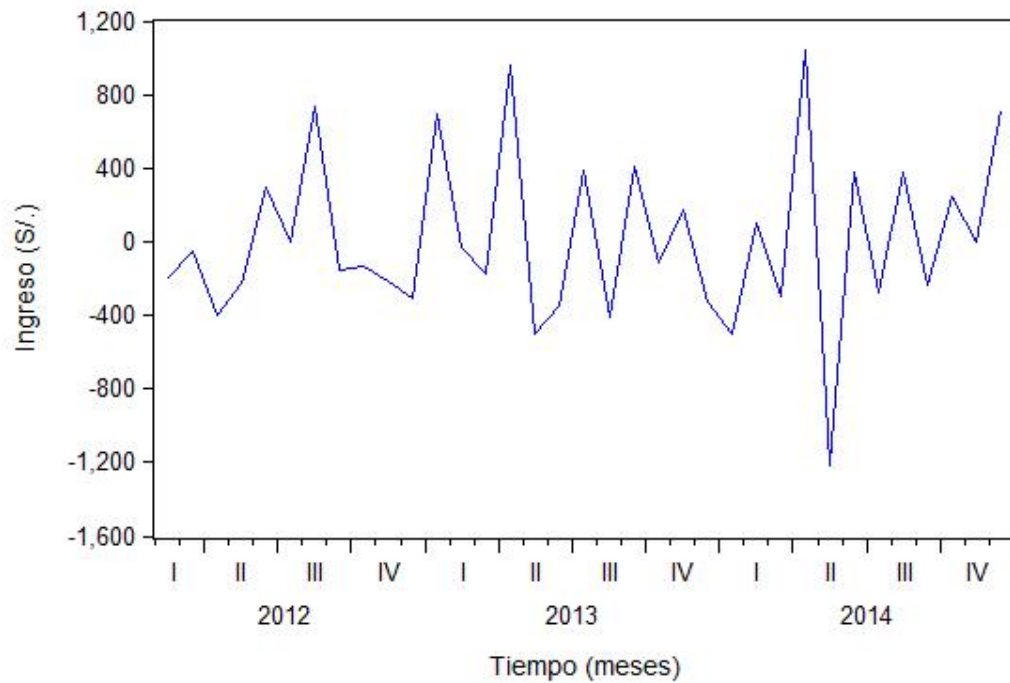
	τ	Prob.
Estadístico de prueba de Dickey y Fuller	-1.288	0.1784
Valores críticos: 1 % nivel	-2.635	
5 % nivel	-1.951	
10 % nivel	-1.611	

Fuente: Elaboración propia

Nota: resultado obtenido para el rezago 12, obtenido automaticamente mediante el Software Eviews 8 para un número máximo de rezagos de 16 y seleccionado mediante el criterio de información de Schwarz.

Por tanto, se realizó una diferencia regular a y_t con el propósito de tranformarlo a una serie estacionaria, así se determinó que la serie $z_t = \nabla y_t = y_t - y_{t-1}$ mediante un análisis gráfico, análisis de las autocorrelaciones y la prueba de estacionariedad de Dickey y Fuller para un nivel de confianza del 95 %, que la serie diferenciada z_t es no estacionaria, observados en las Figuras 05 y 06 y Cuadro 2, se mostró que la serie diferenciada elimina el problema de tendencia, sin embargo permanece el problema de estacionalidad en la serie.

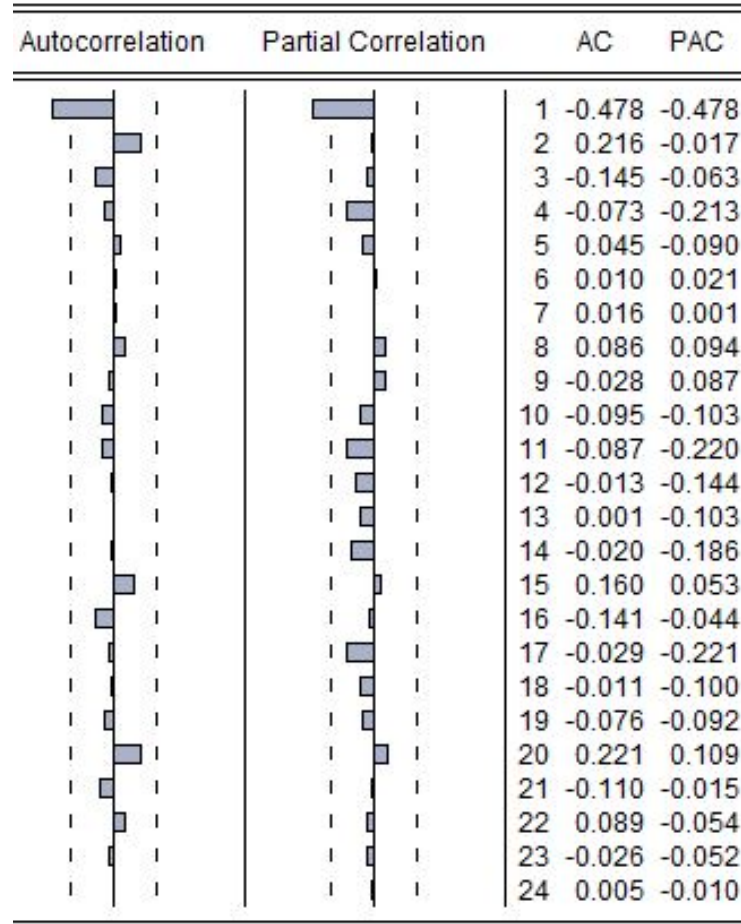
Figura 7: Evolución de la primera diferencia regular y diferencia estacional del ingreso mensual en Nuevos Soles (S/.) del hotel El Centro durante los años 2011 al 2014 (w_t).



Fuente: Elaboración propia

Se realizó otra diferencia adicional ahora en la parte estacional de la serie y_t , así, sea w_t una transformación de la serie y_t , en su parte estacional. Así en la Figura 07 en la evolución de la primera diferencia regular y diferencia estacional de y_t denotada como $w_t = \nabla \nabla_{12} y_t = y_t - y_{t-1} - y_{t-12} + y_{t-13}$, para $t = 1, 2, \dots, 35$, no se observó algún tipo de tendencia en la serie, y no mostró picos altos durante meses repetidos en cada año, es decir, existe evidencia de la no presencia de estacionalidad en la serie w_t .

Figura 8: Correlogramas y valores de la Función de Autocorrelación (Autocorrelation) (AC) y Función de Autocorrelación Parcial (Partial Correlation) (PAC) del ingreso mensual del hotel El Centro periodo 2011 al 2014 (w_t).



Fuente: Elaboración propia

Se corroboró lo mostrado en la Figura 07 mediante el análisis de los correlogramas, las funciones de autocorrelación y la prueba de Dickey y Fuller para w_t .

En la Figura 08 para la serie diferenciada w_t , se observó que en los valores de las autocorrelaciones (AC) y en las autocorrelaciones parciales (PAC) mostraron valores altos en valor absoluto en el rezago 1. observados también en los autocorrelogramas (autocorrelation y Partial Correlation).

Por la prueba de raíz unitaria para la serie $w_t = \nabla \nabla_{12} y_t = y_t - y_{t-1} - y_{t-12} + y_{t-13}$ con:

H_0 : Los ingresos mensuales (w_t) tiene raíz unitaria

H_1 : Los ingresos mensuales (w_t) no tiene raíz unitaria

del Cuadro 3 se observó un valor del estadístico de Dickey y Fuller de $\tau = -9,688$, siendo el valor menor para valores críticos para niveles de significación de 0.01, 0.05 y 0.1, y con valor de probabilidad de 0.000, así rechazando la hipótesis nula de presencia de raíz unitaria, es decir la presencia de estacionariedad de la serie transformada w_t .

Cuadro 3: Prueba de raíz unitaria de Dickey y Fuller para la serie diferenciada regularmente y estacionalmente del ingreso mensual (Dif(ingresos,1,12)) del hotel El Centro durante los años 2011 al 2014 (w_t).

	τ	Prob.
Estadístico de prueba de Dickey y Fuller	-9.688	0.000
Valores críticos: 1 % nivel	-2.635	
5 % nivel	-1.951	
10 % nivel	-1.611	

Fuente: Elaboración propia

Nota: resultado obtenido para el rezago 12, obtenido automaticamente mediante el Software Eviews 8 para un número máximo de rezagos de 16 y seleccionado mediante el criterio de información de Schwarz.

De esta manera se realizó otra diferencia en su parte estacional, esta de patrón estacional de 12 meses ($s=12$), es decir, una diferencia regular y una diferencial estacional, así se determinó que la serie $w_t = \nabla \nabla_{12} y_t = y_t - y_{t-1} - y_{t-12} + y_{t-13}$ mediante un análisis gráfico, análisis de las autocorrelaciones y la prueba de estacionariedad de Dickey y Fuller para un nivel de confianza del 95 %, que la serie diferenciada w_t es estacionaria, esto se observó en las Figuras 07 y 08 y Cuadro 3, donde no se observó algún tipo de tendencia y estacionalidad presente en la estructura de la serie.

4.2. Identificación de los modelos

Siendo que la serie y_t requiere una diferencia regular y una diferencia estacional para que esta sea estacionaria, se planteó una estructura ARIMA estacional, es decir un modelo SARIMA con patrón estacional $s=12$, así se observaron los correlogramas de la Figura 08, donde mostró desfases en su primer rezago en las autocorrelaciones y autocorrelaciones parciales, dando indicios de ser un procesos autorregresivo (AR) o de medias móviles (MA), ambos de orden 1, tal como Alpuim (1998) muestra los posibles modelos para una serie a partir de sus correlogramas. Por tanto se propuso 2 modelos para su ajuste, de acuerdo a la forma que presentan en sus respectivos correlogramas de autocorrelacion y autocorrelacion parcial mostradas en la Figura 08. Los modelos propuestos fueron el modelo $SARIMA(1, 1, 0)(0, 1, 0)_{12}$ y el modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$, detalladas a continuación:

1. Modelo $SARIMA(1, 1, 0)(0, 1, 0)_{12}$ cuya ecuación viene dada por la forma:

$$(1 - \phi B)(\nabla_{12} \nabla y_t) = \varepsilon_t$$

$$y_t = (1 + \phi)y_{t-1} - \phi y_{t-2} + y_{t-12} - (1 + \phi)y_{t-13} - \phi y_{t-14} + \varepsilon_t$$

2. Modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$ cuya ecuación viene dada por la forma:

$$\nabla_{12} \nabla y_t = (1 - \theta B)\varepsilon_t$$

$$y_t = y_{t-1} + y_{t-12} - y_{t-13} + \varepsilon_t - \theta \varepsilon_{t-1}$$

4.3. Estimación y selección del modelo

La estimación de parámetros para cada modelo fue realizado mediante el método de los mínimos cuadrados, obteniendo las estimaciones de los coeficientes de los modelos, mostrados a continuación:

a. Estimación del modelo $SARIMA(1, 1, 0)(0, 1, 0)_{12}$

Los valores estimados de los coeficientes del modelo $SARIMA(1, 1, 0)(0, 1, 0)_{12}$ para los ingresos mensuales y_t se presentan en el Cuadro 4, siendo significativo a un nivel de confianza del 95 % con valor del estadístico $t = -3,284$ y valor de probabilidad $p=0.0024$, así la ecuación es dada por:

$$(1 - \phi B)(\nabla_{12} \nabla y_t) = \varepsilon_t$$

$$(1 - \phi B)(y_t - y_{t-1} - y_{t-12} + y_{t-13}) = \varepsilon_t$$

$$y_t = (1 + \phi)y_{t-1} - \phi y_{t-2} + y_{t-12} - (1 + \phi)y_{t-13} - \phi y_{t+14} + \varepsilon_t$$

Siendo el valor de $\phi = -0,513$ se tiene:

$$y_t = 0,487y_{t-1} + 0,513y_{t-2} + y_{t-12} - 0,478y_{t-13} + 0,513y_{t+14} + \varepsilon_t$$

Donde ε_t es ruido blanco con $\varepsilon_t \sim N(0, 1)$.

Indicando que el modelo está impuesto una dependencia de tipo lineal entre el valor del proceso en determinados instantes y_t y sus valores pasados para $t=1,2,12,13$ y 14 .

Cuadro 4: Modelo $SARIMA(1, 1, 0)(0, 1, 0)_{12}$ estimado y principales medidas estadísticas, para la serie del ingreso mensual del hotel El Centro periodo 2011 al 2014.

	Coef.	Std. Error	Est. t	Prob.
ϕ	-0.513	0.156	-3.284	0.0024
Criterio de inf. Akaike (AIC)		14.902		
Criterio de inf. Schwarz (BIC)		14.947		
Est. Durbin - Watson		1.956		

Fuente: Elaboración propia

Nota: Estimaciones obtenidas por el método de los mínimos cuadrados mediante el Software Eviews 8.

b. Estimación del modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$

Los valores estimados de los coeficientes del modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$ para los ingresos mensuales y_t se presentan en el Cuadro 5, siendo significativo a un nivel de confianza del 95 % con valor del estadístico $t = -5,152$ y valor de probabilidad $p=0.000$, así la ecuación es dada por:

$$\nabla_{12}\nabla y_t = (1 - \theta B)\varepsilon_t$$

$$y_t - y_{t-1} - y_{t-12} + y_{t-13} = \varepsilon_t - \theta\varepsilon_t$$

$$y_t = y_{t-1} + y_{t-12} - y_{t-13} + \varepsilon_t - \theta\varepsilon_{t-1}$$

Siendo el valor de $\theta = -0,673$ se tiene:

$$y_t = y_{t-1} + y_{t-12} - y_{t-13} + \varepsilon_t + 0,673\varepsilon_{t-1}$$

Donde ε_t es ruido blanco con $\varepsilon_t \sim N(0, 1)$.

Indicando que el modelo está impuesto una dependencia de tipo lineal entre el valor del proceso en determinados instantes y_t y sus valores pasados para $t=1, 12$ y 13 , y también con sus valores pasados de las innovaciones ε_t para $t=1$.

Cuadro 5: Modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$ estimado y principales medidas estadísticas, para la serie del ingreso mensual del hotel El Centro periodo 2011 al 2014.

Variable: Ingresos mensual (y_t)				
	Coef.	Std. Error	Est. t	Prob.
θ	-0.673	0.131	-5.152	0.0000
Criterio de inf. Akaike (AIC)		14.848		
Criterio de inf. Schwarz (BIC)		14.892		
Est. Durbin - Watson		1.824		

Fuente: Elaboración propia

Nota: Estimaciones obtenidas por el método de los mínimos cuadrados mediante el Software Eviews 8.

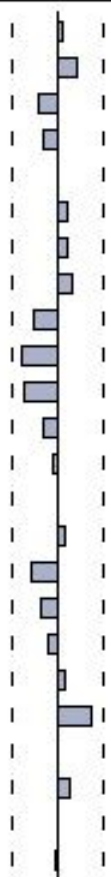
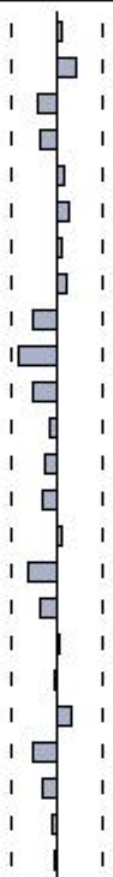
Para la selección del mejor modelo de estimación se ha computado para cada una de las series consideradas los criterios de información, se observó en el Cuadro 4 y Cuadro 5 que ambos modelos presentaron buen comportamiento en su ajuste, con valores del estadístico de Durbin y Watson de 1.956 y 1.824, ambos cercanos a 2. Comparando ambos modelos por los criterios de información de Akaike (AIC) y de Schwarz (BIC) se tiene que el modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$ con valores de AIC=14.848 y BIC=14.892 presentó menores valores con respecto al modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$ con valores de AIC=14.902 y BIC=14.947, es decir el modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$, cuya ecuación:

$$y_t = y_{t-1} + y_{t-12} - y_{t-13} + \varepsilon_t + 0,673\varepsilon_{t-1}$$

es mejor modelo con respecto a la estimación de sus parámetros.

4.4. Validación del modelo

Figura 9: Correlogramas y valores de la Función de Autocorrelación (Autocorrelation) (AC) y función de autocorrelación parcial (Partial Correlation) (PAC); y valor del estadístico de prueba de Ljung - Box (Q-Stat) para los residuos del modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$ de la serie del ingreso mensual del hotel El Centro periodo 2011 al 2014.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 0.033	0.033	0.0421	
		2 0.144	0.143	0.8586	0.354
		3 -0.141	-0.154	1.6673	0.434
		4 -0.111	-0.125	2.1788	0.536
		5 0.002	0.057	2.1789	0.703
		6 0.078	0.098	2.4499	0.784
		7 0.074	0.026	2.7031	0.845
		8 0.113	0.079	3.3144	0.854
		9 -0.173	-0.180	4.8120	0.777
		10 -0.271	-0.289	8.6298	0.472
		11 -0.250	-0.178	12.004	0.285
		12 -0.109	-0.055	12.667	0.316
		13 -0.029	-0.093	12.717	0.390
		14 0.004	-0.111	12.718	0.470
		15 0.059	0.035	12.941	0.531
		16 -0.195	-0.214	15.529	0.414
		17 -0.118	-0.134	16.538	0.416
		18 -0.080	0.023	17.025	0.453
		19 0.042	-0.010	17.170	0.511
		20 0.256	0.099	22.824	0.245
		21 -0.006	-0.186	22.827	0.297
		22 0.084	-0.102	23.523	0.317
		23 -0.004	-0.032	23.524	0.373
		24 -0.022	-0.024	23.580	0.427

Fuente: Elaboración propia

Se determinó que el modelo que mejor ajusta a la serie y_t en cuanto a su estimación es el modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$, este modelo fue validado por un análisis del comportamiento de sus residuos ε_t , la cual deberían de tener un comportamiento de ruido blanco, es decir la no autocorrelación y la normalidad de los residuales.

Se observaron en la Figura 09 los residuales del modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$, pre-

sentando los valores de las autocorrelaciones (AC) y las autocorrelaciones parciales (PAC) valores pequeños, y ningún valor fuera de las bandas en todos los rezagos, esto apreciable también en los correlogramas (Autocorrelation y Partial Correlation).

La prueba de autocorrelación en los residuales ε_k de Ljung - Box se realizó para cada rezago $k = 1, 2, \dots, 24$, como sigue:

$$H_0 : \rho(1) = \rho(2) = \dots = \rho(k) = 0$$

$$H_1 : \text{por lo menos un } \rho(k) \neq 0.$$

Se observaron que para cada rezago, los valores del estadístico de Ljung - Box (Q-Stat) presentaron valores altos y los valores de probabilidad (Prob.) fueron mayores del 0.05, así para un nivel de significación del 5 % no fue rechazado la hipótesis nula de autocorrelación en los residuales ε_k para $k = 1, 2, \dots, 24$.

Por tanto, se observaron en los correlogramas y por el contraste de Ljung y Box, para un nivel de confianza del 95 % que no existe correlación, es decir la independencia entre los residuales.

Se observó también en la Figura 10 la forma de distribución de los residuales del modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$, presentando mediante el histograma de frecuencia una aparente distribución normal, realizando la prueba de normalidad de Jarque y Bera se obtuvo:

H_0 : Los residuos se distribuyen según una distribución normal

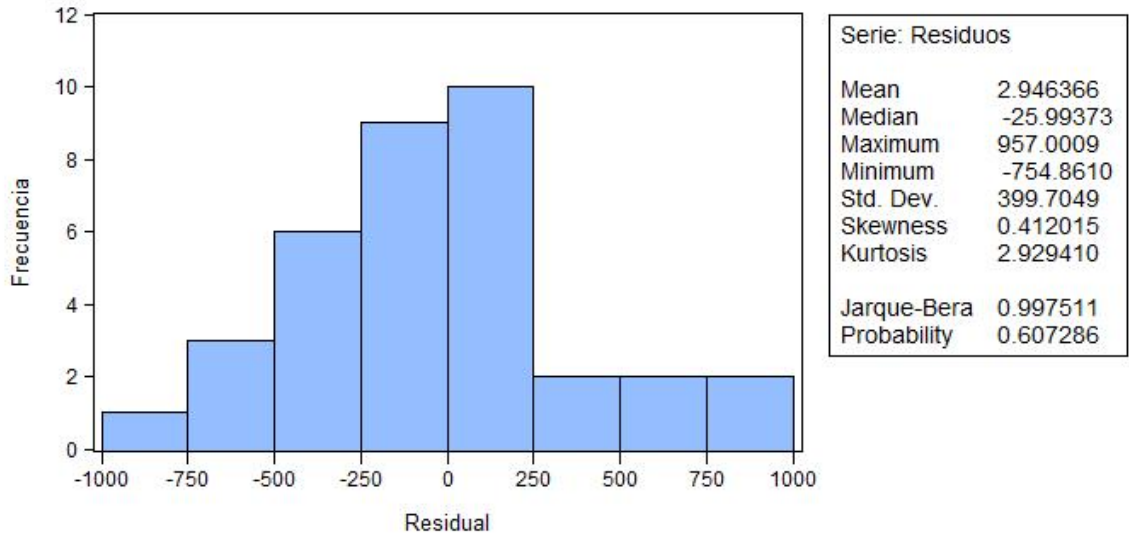
H_1 : Los residuos no se distribuyen según una distribución normal

El valor del estadístico de Jarque-Bera $JB = 0.997$, y valor de probabilidad (probability) de 0.607. Por tanto a un nivel de confianza del 95 % no se rechaza H_0 , es decir los residuos se distribuyen según una distribución normal.

Del histograma de frecuencia de los resíduos y por la prueba de Jarque y Bera para un nivel de confianza del 95 %, Figura 10, se observó que los residuos se distribuyen según una distribución normal.

De esta manera, los residuos del modelo se comportan como un ruido blanco, es decir, el modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$ es adecuado para describir la serie de los ingresos mensuales.

Figura 10: Histograma de frecuencia y principales medidas estadísticas para los residuales del modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$ de la serie del ingreso mensual del hotel El Centro periodo 2011 al 2014.



Fuente: Elaboración propia

4.5. Pronósticos

1. Pronóstico de Box y Jenkins del modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$

Los valores previstos de intervalos de predicción mediante la metodología de Box y Jenkins, a niveles de confianza del 80 % y 95 % con el modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$ para los ingresos mensuales, donde la información proyectada fue de 12 meses desde la información existente de ingresos mensuales del año 2015, se muestran en los Cuadros 6 y 7 y en las Figuras 11 y 12, donde se observó que los intervalos de predicción inferiores y superiores a niveles de confianza del 80 % y 95 % cubren todos los valores reales, sin embargo los intervalos superiores se encuentran muy próximos a los valores reales, y se observó que el valor del error porcentual medio absoluto fue de $EPMA = 7.784$.

2. Pronóstico Bootstrap del modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$

Los valores previstos de intervalos de predicción Bootstrap, para muestras bootstrap de tamaño $B=3000$ y a niveles de confianza del 80 % y 95 % con el modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$ para los ingresos mensuales, donde la información proyectada fue de 12 meses desde la información existente de ingresos mensuales del

año 2015 se muestran en los Cuadros 6 y 7 y en las Figuras 11 y 12, donde se observaron que los intervalos de predicción inferiores y superiores cubren a todos los valores reales de los ingresos mensuales, se observó también que los valores reales se encuentran en promedio en el centro de los intervalos de predicción y se obtuvo también que el valor del error medio absoluto bootstrap fue de $EPMA_B = 7.462$.

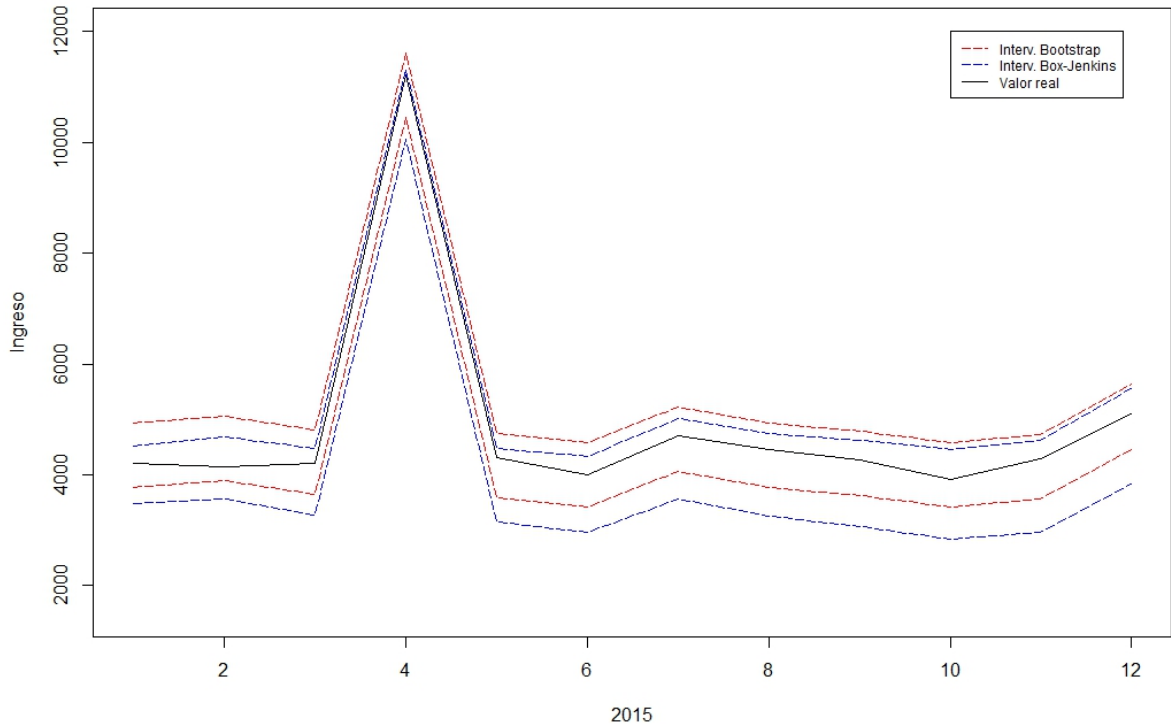
Cuadro 6: Pronósticos del modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$ de los ingresos mensuales para el año 2015 para intervalos de confianza al 80 % y valores del error porcentual medio absoluto (EPMA), mediante metodología Bootstrap y metodología de Box y Jenkins.

Año	Mes	Valor real	Intervalos de confianza		Intervalos de confianza	
			Bootstrap		Box y Jenkins	
			Inferior	Superior	Inferior	Superior
2015	Enero	4200	3764.248	4933.543	3476.905	4514.482
	Febrero	4150	3894.248	5063.543	3567.696	4683.691
	Marzo	4200	3644.248	4813.543	3281.067	4470.320
	Abril	11220	10454.248	11623.543	10056.567	11314.820
	Mayo	4300	3574.248	4743.543	3143.863	4467.525
	Junio	3990	3414.248	4583.543	2952.700	4338.687
	Julio	4710	4054.248	5223.543	3562.880	5008.507
	Agosto	4460	3764.248	4933.543	3244.242	4747.145
	Setiembre	4260	3614.248	4783.543	3066.656	4624.731
	Octubre	3910	3414.248	4583.543	2840.015	4451.373
	Noviembre	4290	3564.248	4733.543	2964.226	4627.161
	Diciembre	5100	4464.248	5633.543	3839.214	5552.174
			$EPMA_B = 7.462$		EPMA=7.784	

Fuente: Elaboración propia

Nota: Estimaciones de pronósticos Bootstrap obtenidas mediante un programa realizado con el *Software R* (Anexo A) para un número de remuestras Bootstrap de 3000.

Figura 11: Pronósticos por intervalos de confianza al 80 % para el año 2015 del modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$ de la serie del ingreso mensual, mediante la metodología de Box y Jenkins y metodología Bootstrap.



Fuente: Elaboración propia

4.6. Comparación

Se observaron los pronósticos para el año 2015 del modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$ en intervalos de confianza al 80 % y 95 % mediante las metodologías Bootstrap y Box y Jenkins, mostrando que los valores reales de enero a diciembre se encuentran dentro de los límites inferiores y superiores para ambas metodologías, respecto a sus intervalos de predicción a niveles de confianza del 80 % y 95 % de los ingresos mensuales mostrados en las Figuras 11 y 12, se observaron que los intervalos de predicción mediante la metodología Bootstrap presentan menores amplitudes en cada mes del año 2015, respecto a los intervalos de predicción mediante la metodología de Box y Jenkins.

Comparando ambos modelos mediante su error porcentual medio absoluto (EPMA), se observó que el modelo mediante la metodología Bootstrap es mejor modelo en cuan-

to a su predicción, pues presenta menor error porcentual medio absoluto con valor de $EPMA_B = 7.462$, en contraste a la metodología de Box y Jenkins que presenta un valor del error porcentual medio absoluto de $EPMA = 7.784$.

El pronóstico en intervalos de predicción a niveles de confianza del 80% y 95 % del mejor modelo para el año 2016 de los ingresos mensuales del Hotel el Centro, Ayacucho, se muestran en el Cuadro 7.

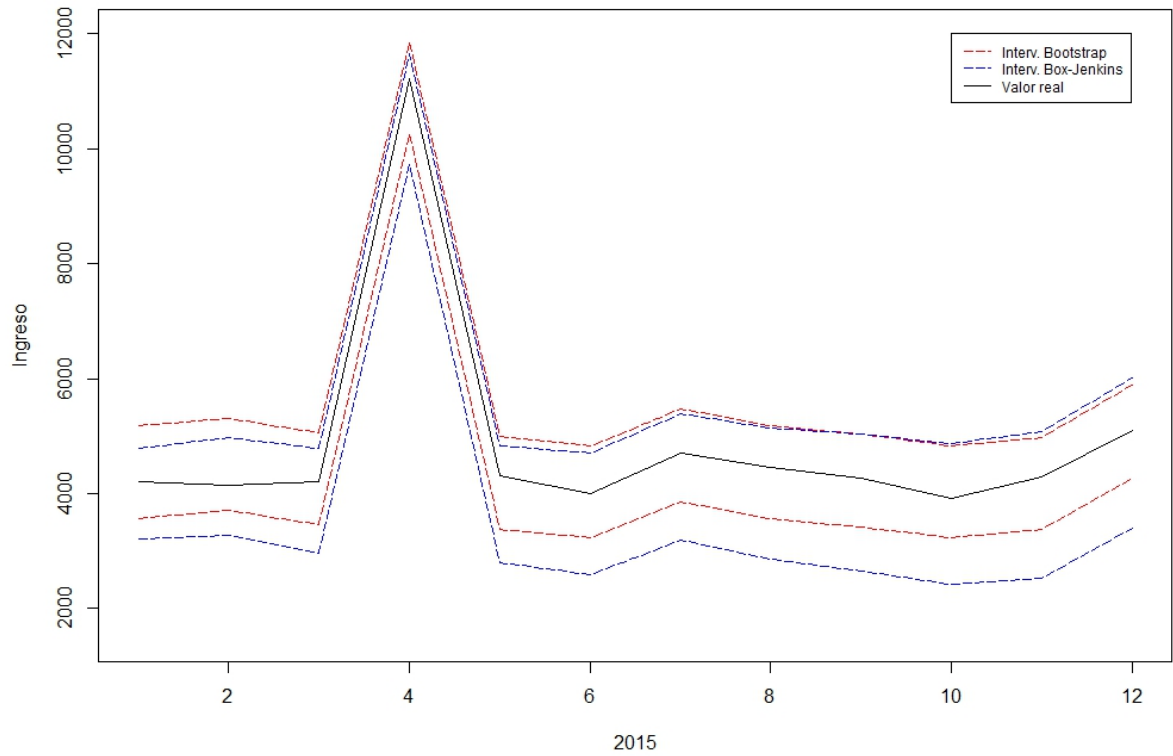
Cuadro 7: Pronósticos del modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$ de los ingresos mensuales para el año 2015 para intervalos de confianza al 95 % y valores del error porcentual medio absoluto (EPMA), mediante metodología de Box y Jenkins y metodología Bootstrap.

Año	Mes	Valor real	Intervalos de confianza		Intervalos de confianza	
			Bootstrap		Box y Jenkins	
			Inferior	Superior	Inferior	Superior
2015	Enero	4200	3569.961	5180.252	3202.275	4789.112
	Febrero	4150	3699.961	5310.252	3272.310	4979.077
	Marzo	4200	3449.961	5060.252	2966.291	4785.097
	Abril	11220	10259.961	11870.252	9723.527	11647.860
	Mayo	4300	3379.961	4990.252	2793.510	4817.877
	Junio	3990	3219.961	4830.252	2585.852	4705.536
	Julio	4710	3859.961	5470.252	3180.245	5391.142
	Agosto	4460	3569.961	5180.252	2846.448	5144.939
	Setiembre	4260	3419.961	5030.252	2654.259	5037.128
	Octubre	3910	3219.961	4830.252	2413.514	4877.873
	Noviembre	4290	3369.961	4980.252	2524.074	5067.314
	Diciembre	5100	4269.961	5880.252	3385.820	6005.567
			$EPMA_B = 7.462$		$EPMA = 7.784$	

Fuente: Elaboración propia

Nota: Estimaciones de pronósticos Bootstrap obtenidas mediante un programa realizado con el *Software R* (Anexo A) para un número de remuestras Bootstrap de 3000.

Figura 12: Pronósticos por intervalos de confianza al 95 % de confianza para el año 2015 del modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$ de la serie del ingreso mensual, mediante la metodología de Box y Jenkins y metodología Bootstrap.



Fuente: Elaboración propia

Cuadro 8: Intervalos de predicción de los ingresos mensuales para el año 2016 mediante la metodología Bootstrap a niveles de confianza del 80 % y 95 %.

Año	Mes	IC Bootstrap 80 %		IC Bootstrap 95 %	
		Inferior	Superior	Inferior	Superior
2016	Enero	3878.497	6217.087	3489.923	6710.505
	Febrero	4008.497	6347.087	3619.923	6840.505
	Marzo	3758.497	6097.087	3369.923	6590.505
	Abril	10568.497	12907.087	10179.923	13400.505
	Mayo	3688.497	6027.087	3299.923	6520.505
	Junio	3528.497	5867.087	3139.923	6360.505
	Julio	4168.497	6507.087	3779.923	7000.505
	Agosto	3878.497	6217.087	3489.923	6710.505
	Setiembre	3728.497	6067.087	3339.923	6560.505
	Octubre	3528.497	5867.087	3139.923	6360.505
	Noviembre	3678.497	6017.087	3289.923	6510.505
	Diciembre	4578.497	6917.087	4189.923	7410.505

Fuente: Elaboración propia

En el Cuadro 8 se observaron las predicciones por intervalo a niveles de confianza del 80 % y 95 % mediante la metodología Bootstrap de los ingresos mensuales de enero a diciembre del 2016, se realizaron 3000 muestras Bootstrap obtenidas mediante un programa realizado con el *Software R*, mostradas en el Anexo A.

5. CONCLUSIONES Y SUGERENCIAS

5.1. Conclusiones

- La serie de los ingresos mensuales del Hotel el Centro en el periodo 2011 al 2014 presenta estacionalidad en su estructura, con patrón estacional de 12 meses, por tanto se adapta a un modelo estacionario SARIMA. El modelo de predicción propuesto por las metodologías Bootstrap y de Box y Jenkins que mejor describe a la serie del ingreso mensual del Hotel El Centro en el periodo 2011 al 2014 es el modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$, cuya ecuación es:

$$y_t = y_{t-1} + y_{t-12} - y_{t-13} + \varepsilon_t + 0,673\varepsilon_{t-1}$$

- El modelo propuesto mediante la metodología de Box y Jenkins, presenta resultados adecuados en cuanto a su predicción del ingreso mensual del Hotel el Centro, Ayacucho en el periodo 2011 al 2014, presentando su valor del error absoluto medio porcentual (7.784 %) por debajo del 10 %.
- El modelo propuesto mediante la metodología Bootstrap, presenta resultados adecuados en cuanto a su predicción del ingreso mensual del Hotel el Centro, Ayacucho en el periodo 2011 al 2014, presentando su valor del error absoluto medio porcentual (7.462 %) por debajo del 10 %.
- Mediante la metodología Bootstrap se obtienen mejores predicciones en contraste con la metodología de Box y Jenkins, pues las predicciones evaluados para la serie del ingreso mensual del Hotel el Centro, Ayacucho en el periodo 2011 al 2014, mediante la metodología Bootstrap proporcionan un menor error porcentual medio absoluto, y también presenta un mejor desempeño en la obtención de intervalos de predicción, pues se verificó que mediante la metodología Bootstrap se obtienen intervalos de predicción con menores amplitudes a niveles de confianza del 80 % y 95 %.

5.2. Sugerencias

- Para la aplicación de la metodología Bootstrap en series temporales estacionarios, se sugiere trabajar con muestras grandes, o de patrones estacionales menores a $s = 6$, para así no perder información y obtener mejores resultados.
- Analizar la serie de los ingresos mensuales mediante otras técnicas de predicción y compararlas con la metodología Bootstrap.
- Para un posterior trabajo de investigación, tener en cuenta los procesos que incluyan la heterocedasticidad condicional de una serie y analizarlas mediante la metodología Bootstrap.

6. BIBLIOGRAFÍA

Alarcón, A. (2012). *Evaluación de modelos econométricos alternativos de series de tiempo para el pronóstico de la inflación en el ecuador en el corto plazo: periodo 2000 - 2010*. Tesis Lic. Universidad de Cuenca.

Alpuim, T. (1998). *Séries temporáís*. Lisboa, 2da edição: Associação dos Estudantes da Faculdade de Ciências.

Alvarez, A. M., y Rubio, P. O. (2000). Metodología bootstrap en series heterocedásticas. una aplicación al ibex-35. *XIV Reunión ASEPELT. Anales de Economía Aplicada*(9), 84-99.

Box, G. P., y Jenkins, G. M. (1970). *Time series analysis forecasting and control*. san francisco: holden-day.

Cao, R., Febrero, M., Manteinga, G., y Sánchez, P. (1997). Saving computer time in constructing consistent bootstrap prediction intervals for autoregressive processes. *Communications in Statistics-Simulation and Computation*, 26, 961-978.

Carvalho, O. W., Henning, E., y Moro, G. (2013). Aplicação de um modelo sarima na previsão de vendas de motocicletas. *Revistas Científicas de America Latina, Caribe, España e Portugal, Exacta*, 11, 77-88.

Davison, A., y Hinkley, D. (1997). *Bootstrap methods and their application*. Cambridge: Cambridge University Press.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7, 1-26.

- Efron, B., y Tibshirani, R. (1994). *An introduction to the bootstrap*. New York: Chapman and Hall.
- Fan, J., y Yao, Q. (1994). *Nonlinear time series*. New York: Springer.
- Fernández, J. (1999). *El método jackknife para la estimación de varianzas de estimadores*. FCM: Universidad Nacional Mayor de San Marcos.
- Guevara, L. R. (2012). *Metodología para la medición de la atención en una central telefónica usando box y jenkins*. Tesis Lic. Universidad Nacional Mayor de San Marcos.
- Henrique, M. (2003). *Modelos de previsão em séries temporais. aplicação da metodologia bootstrap*. Tese Ms. Universidade nova de Lisboa, Faculdade de Ciências y Tecnologia.
- Jurado, G., Manteinga, G., y Sánchez, P. (1995). Predicting using box-jenkins, nonparametric and bootstrap techniques. *Technometrics*, 37, 303-310.
- Kim., J., Song, H., Wong, K., Athanasopoulos, G., y Liu, S. (2009). Beyond point forecasting: Evaluation of alternative prediction intervals for tourist arrivals. *Monash University. Department of Econometrics and Business Statistics*, 30, 1-19.
- Marques, J. (2013). *A metodologia bootstrap associada ao método de holt-winters na previsão de séries temporais*. Brasil: Universidade de Minho, Escola de Ciências.
- McCullough, B. (1994). Bootstrapping forecast intervals: An application to ar(p) models. *Journal of Forecasting*, 13, 51-66.
- Medina, C. M., y Vizconde, T. O. (2011). Modelo sarima para la llegada mensual de visitantes extranjeros por el aeropuerto internacional jorge Chávez. *UCV - Scientia*, 3, 59-70.
- Miranda, A. (2003). *El método de remuestreo y su aplicación en la investigación biomédica*. Ciudad de la Habana: Escuela Nacional de Salud Pública Carlos J. Finlay.
- Mirando, J., y Ramos, W. (2010). Pronóstico de la tendencia nacional y regional de las enfermedades diarreicas agudas en menores de cinco años de edad en el Perú mediante un modelo arima con el enfoque box-jenkins. *Revista Peruana de Epidemiología*, 14, 1-8.

- Morettin, P. A., y Toloi, C. M. C. (2011). *Análise de séries temporais*. São Paulo: Bloucher.
- Murteira, B., y Muller, D. (1990). *Análise de sucessão cronológicas*. Lisboa: McGraw-Hill.
- Pascual, L., Romo, J., y Ruiz, E. (1999). Bootstrap predictive inference for arima processes. *Working Papers. Statistics and Econometrics*, 40(4), 86-98.
- Pascual, L., Romo, J., y Ruiz, E. (2006). Bootstrap prediction for return and volatilities in garch models. *Computational Statistics and Data Analysis*, 50(9), 2293-2312.
- Peña, D. (2010). *Análisis de series temporales*. Madrid, ES: Alianza editorial.
- Peña, D., Tiao, G., y Tsay, R. (2001). *A course in time series analysis*. New York: J Wiley.
- Quiroz, M. (2014). *Análisis de proyección del consumo eléctrico residencial de la región piura mediante modelo univariante arima*. Tesis Maestria, Universidad Nacional Pedro Ruiz Gallo.
- Ricketts, C., y Berry, J. (1994). Teaching statistics through resampling. *Teaching Statistics*, 6, 41-44.
- Schwarz, G. (1979). Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464.
- Shimizu, K. (2009). *Bootstrapping stationary arma-garch models*. Wiesbaden: Springer Science.
- Stine, R. (1987). Estimating properties of autoregressive forecasts. *American Statistical Association*, 82, 1072-1078.
- Thombs, L., y Schucany, W. (1990). Bootstrap prediction intervals for autoregression. *American Statistical Association*, 85, 486-492.
- Tineo, F. (2005). *Método de remuestreo bootstrap*. Tesis Lic. Universidad Nacional Mayor de San Marcos.

Tukey, J. W. (1958). Bias and confidence in not-quite large sample. *Annals of Mathematical Statistics*, 29, 614.

Uriel, E., y Peiró, A. (2005). *Introducción al análisis de series temporales*. España: Alfa centauro.

Urquiza, A. O. (2011). *Análisis estocástico arima para el modelamiento y predicción de la demanda eléctrica en el sector residencial de lima sur*. Tesis Maestría, Universidad Nacional de Ingeniería.

7. ANEXOS

8. A. Programa del algoritmo Bootstrap mediante el Software *R* para el cálculo de las predicciones h pasos hacia adelante del modelo $SARIMA(0, 1, 1)(0, 1, 0)_{12}$ de la serie de ingresos mensuales.

```
boot_tesis<-function(datos,B,h,alpha){  
  #B: Número de remuestras bootstrap  
  #h: Número de predicciones hacia adelantes  
  #alpha: Nivel de confianza  
  modelo<-Arima(datos,order=c(0,1,1),seasonal=c(0,1,0),  
                method="CSS",include.mean = F)  
  datos<-as.matrix(datos)  
  datos_prueba<-c(4200,4150,4200,11220,4300,3990,4710,4460,  
                  4260,3910,4290,5100)  
  p<-length(datos)  
  resi<-matrix(modelo$res)[14:48,]  
  resi1<-matrix(modelo$res)  
  n<-length(resi)  
  ECM<-liboot<-lsboot<-liboot1<-lsboot1<-ser2<-c()  
  ajust<-matrix(0,ncol=B,nrow=p)  
  ser1<-datos[1:13,]  
  pred<-matrix(0,ncol=B,nrow=h)  
  for (i in 1:B){  
    muestra<-sample(resi,n,replace=TRUE)  
    residuos<-c(rep(0,13),muestra)  
    for (j in 1:p){  
      if(j<=14) ser2[j]<-datos[j]  
      if(j>14) ser2[j]<-datos[j-1]+datos[j-12]-datos[j-13]+  
        residuos[j]+modelo$coef[1]*residuos[j-1]  
    }  
  }
```

```

modestimado<-Arima(ser2[12:48],order=c(0,1,1),seasonal=c(0,1,0),
                    method="CSS",include.mean = F)
muestra1<-sample(resi,n,replace=TRUE)
residuos1<-c(rep(0,13),muestra1)
for(j in 1:h){
  if(j==1) pred[j,i]<-datos[p+j-1,]+datos[p+j-12,]-datos[p+j-13,]
  +modestimado$coef[1]*residuos1[p+j-1]
  if(j>=2&j<=12) pred[j,i]<-pred[j-1,i]+datos[p+j-12,]
  -datos[p+j-13,]
  if(j==13) pred[j,i]<-pred[j-1,i]+pred[j-12,i]-datos[p+j-13,]
  if(j>=14) pred[j,i]<-pred[j-1,i]+pred[j-12,i]-pred[j-13,i]
}

ECM[i]<-sum(abs((datos_prueba-pred[,i])/datos_prueba)*100)/12
}
for(k in 1:h){
  liboot[k]<-quantile(pred[k,],alpha/2)
  lsboot[k]<-quantile(pred[k,],1-alpha/2)
}
puntual<-forecast.Arima(modelo,h)$mean
ECM1=sum(abs((datos_prueba-puntual)/datos_prueba)*100)/12
li<-forecast.Arima(modelo,h)$lower[,2]
ls<-forecast.Arima(modelo,h)$upper[,2]
meanECM<-mean(ECM)
plot.ts(datos_prueba,ylim=c(1500,12000),xlab="2015",
        ylab="Ingreso")
lines(liboot,col="red",type="l",lty=5)
lines(lsboot,col="red",type="l",lty=5)
lines(li,col="blue",type="l",lty=5)
lines(ls,col="blue",type="l",lty=5)
legend(10,12000,c("Interv. Bootstrap","Interv. Box-Jenkins",
                  "Valor real"),col=c("red","blue","black"),

```

```
lty=c(5,5,1),cex=0.75)
return(list(LiBoot=liboot,LsBoot=lsboot,LI=li,LS=ls,
           EPMAboot=meanECM,EPMA=ECM1))
}
```

[illegible]

