



**UNIVERSIDAD NACIONAL DEL ALTIPLANO**  
**FACULTAD DE INGENIERÍA ESTADÍSTICA E INFORMÁTICA**  
**ESCUELA PROFESIONAL DE INGENIERÍA ESTADÍSTICA E**  
**INFORMÁTICA**



**“MODELO DE PREDICCIÓN DE LOS INGRESOS TRIBUTARIOS  
RECAUDADOS POR LA SUNAT, REFERIDO A TRIBUTOS  
INTERNOS EN LA REGIÓN DE PUNO 2004 -2017”**

**TESIS**

**PRESENTADA POR:**

**Bach. MIGUEL ANGEL CATARI RIVAS**

**PARA OPTAR EL TÍTULO PROFESIONAL DE:**

**INGENIERO ESTADÍSTICO E INFORMÁTICO**

**PUNO – PERÚ**

**2019**



## DEDICATORIA

*A mis padres: Rita Rivas y Miguel Catari, quienes me ofrecieron su constante apoyo y aliento para seguir adelante en todo aquello que emprenda.*

*A mi hermana Lizbeth Rocio, por el respaldo moral que me brindo a cada instante con inmenso cariño incondicional.*

*A mis Tias: Sonia Rivas, Lucia Rivas y Soledad Cutipa por todo el apoyo incondicional brindado.*

*A mis Tio: Walter Rivas quien me impulso a ser mejor a través de sus consejos y orientaciones, y que ahora desde el cielo me cuida y me guía.*

*A mi asesor de tesis, docentes, amigos y familiares, quienes sin su ayuda nunca hubiera podido concluir esta tesis. A todos ellos se los agradezco desde el fondo de mi corazón.*

***Miguel Angel Catari Rivas***



## AGRADECIMIENTOS

A: Mi madre Rita Rivas por ser quien me ha orientado a la búsqueda de una superación personal constante y brindarme siempre más de lo necesario.

A: Mi Asesor Dr. Samuel Donato Pérez Quispe, por su orientación y apoyo constante para la materialización de la presente investigación.

AL: Dr. Santos Octavio Morillos Valderrama, Dr. Edgar Eloy Carpio Vargas, y al Dr. José Pánfilo Tito Lipa que con su gran voluntad me han encaminado en un sendero adecuado para la presente investigación.

A LA: Superintendencia Nacional de Aduanas y de Administración Tributaria – Oficina Zonal de Juliaca, institución pública la cual me permitió realizar la presente investigación.

A: Todos mis amigos y familiares que de alguna manera han colaborado con sus ideas, comentarios y consejos durante todo el transcurso de mi carrera y especialmente en el desarrollo de esta investigación.

**Miguel Angel Catari Rivas**



## INDICE GENERAL

|  | Pág.      |
|--|-----------|
| <b>DEDICATORIA</b>                               |           |
| <b>AGRADECIMIENTOS</b>                           |           |
| <b>INDICE GENERAL</b>                            |           |
| <b>ÍNDICE DE FIGURAS</b>                         |           |
| <b>ÍNDICE DE TABLAS</b>                          |           |
| <b>ÍNDICE DE ACRÓNIMOS</b>                       |           |
| <b>RESUMEN .....</b>                             | <b>11</b> |
| <b>ABSTRACT.....</b>                             | <b>12</b> |
| <b>CAPÍTULO I</b>                                |           |
| <b>INTRODUCCIÓN</b>                              |           |
| <b>1.1 PLANTEAMIENTO DEL PROBLEMA .....</b>      | <b>15</b> |
| <b>1.2 FORMULACIÓN DEL PROBLEMA .....</b>        | <b>16</b> |
| <b>1.3 OBJETIVOS DE LA INVESTIGACION.....</b>    | <b>16</b> |
| 1.3.1. Objetivo General .....                    | 16        |
| 1.3.2. Objetivos específicos.....                | 16        |
| <b>1.4 HIPÓTESIS DE LA INVESTIGACIÓN:.....</b>   | <b>17</b> |
| 1.4.1 Hipótesis General.....                     | 17        |
| 1.4.2 Hipótesis específica .....                 | 17        |
| <b>CAPÍTULO II</b>                               |           |
| <b>REVISIÓN DE LITERATURA</b>                    |           |
| <b>2.1 ANTECEDENTES DE LA INVESTIGACIÓN.....</b> | <b>18</b> |
| 2.2.1. A nivel Internacional.....                | 18        |
| 2.2.2. A nivel Nacional.....                     | 19        |



|  |           |
|--|-----------|
| 2.2.3. A nivel Local .....                                   | 22        |
| <b>2.2 MARCO TEÓRICO .....</b>                               | <b>23</b> |
| <b>2.3 MARCO CONCEPTUAL .....</b>                            | <b>55</b> |
| 2.3.1. Muestra .....   | 55        |
| 2.3.2. Estacionalidad .....                                  | 56        |
| 2.3.3. Estacionariedad .....                                 | 56        |
| 2.3.4. Población.....  | 56        |
| 2.3.5. Unidad Muestral.....                                  | 56        |
| 2.3.6. Serie .....   | 56        |
| 2.3.7. Ruido Blanco .....                                    | 56        |
| 2.3.8. Pronostico .....                                      | 56        |
| 2.3.9. Variable Dependiente.....                             | 57        |
| 2.3.10. Variable Independiente .....                         | 57        |
| 2.3.11. Modelo .....   | 57        |
| 2.3.12. Modelo Box-Jenkins .....                             | 57        |
| 2.3.13. Modelo Univariante de Box-Jenkins .....              | 57        |
| 2.3.14. Modelo Univariante de Box-Jenkins Integrado .....    | 57        |
| 2.3.15. Modelo Univariante De Box-Jenkins No Integrado ..... | 58        |

### **CAPÍTULO III**

#### **MATERIALES Y MÉTODOS**

|  |           |
|--|-----------|
| <b>3.1. UBICACIÓN GEOGRÁFICA DEL ESTUDIO .....</b> | <b>59</b> |
| <b>3.2. MÉTODO DE ANÁLISIS DE DATOS .....</b>      | <b>59</b> |
| <b>3.3. POBLACIÓN Y MUESTRA .....</b>              | <b>62</b> |
| 3.3.1. Población .....                             | 62        |
| 3.3.2. Muestra .....                               | 62        |



|   |           |
|---|-----------|
| 3.3.3. Unidad Muestra .....                               | 62        |
| <b>3.4. MÉTODO DE RECOLECCIÓN DE DATOS .....</b>          | <b>62</b> |
| <b>3.5. OPERACIONALIZACIÓN DE VARIABLES .....</b>         | <b>63</b> |
| <b>CAPÍTULO IV</b>  |           |
| <b>RESULTADOS Y DISCUSIÓN</b>                             |           |
| <b>4.1. IDENTIFICACIÓN DEL MODELO.....</b>                | <b>65</b> |
| <b>4.2. ANÁLISIS DE ESTACIONARIEDAD DE LA SERIE .....</b> | <b>68</b> |
| <b>4.3. DIFERENCIA NO ESTACIONAL.....</b>                 | <b>69</b> |
| <b>4.4. DIFERENCIA ESTACIONAL.....</b>                    | <b>71</b> |
| <b>V. CONCLUSIONES.....</b>                               | <b>93</b> |
| <b>VI. RECOMENDACIONES .....</b>                          | <b>95</b> |
| <b>VII. REFERENCIAS BIBLIOGRÁFICAS.....</b>               | <b>96</b> |

**Área** : Estadística

**Tema** : Series de Tiempo

**Fecha de Sustentación:** 30 de octubre del 2019



## ÍNDICE DE FIGURAS

|   | <b>Pág.</b> |
|---|-------------|
| <b>Figura 1</b> Coeficiente de autocorrelación y de autocorrelación parcial .....   | 43          |
| <b>Figura 2</b> Funciones de autocorrelación.....   | 48          |
| <b>Figura 3</b> Función de autocorrelación y de autocorrelación parcial.....  | 50          |
| <b>Figura 4</b> Procedimiento del método.....   | 61          |
| <b>Figura 5</b> Serie original de recaudaciones tributarias mensuales, con respecto a tributos<br>internos de los años 2004 – 2017 .....  | 65          |
| <b>Figura 6</b> Función de autocorrelación de la recaudación de tributos internos de la región<br>de puno .....   | 66          |
| <b>Figura 7</b> Función de autocorrelación parciales de recaudación de tributos internos en la<br>región de puno .....  | 67          |
| <b>Figura 8</b> Serie transformada (transf. Box - cox) de la recaudación de tributos internos en<br>la región de puno. ....   | 68          |
| <b>Figura 9</b> Función de autocorrelación de la primera diferencia no estacional de la serie<br>transformada para la recaudación de tributos internos en la región de puno. ....             | 69          |
| <b>Figura 10</b> Función de autocorrelación parcial de la primera diferencia no estacional de<br>la serie transformada para la recaudación de tributos internos en la región de<br>puno. .... | 70          |
| <b>Figura 11</b> Función de autocorrelación parcial de la primera diferencia estacional de la<br>serie transformada para la recaudación de tributos internos en la región de<br>puno. ....    | 71          |
| <b>Figura 12</b> Función de autocorrelación parcial de la primera diferencia no estacional de<br>la serie transformada para la recaudación de tributos internos en la región de<br>puno. .... | 72          |



|  |    |
|--|----|
| <b>Figura 13</b> Función de autocorrelación de los residuos del modelo estimado con límites de significancia del 5% para las autocorrelaciones. ....                         | 76 |
| <b>Figura 14</b> Función de autocorrelación parcial de los residuos del modelo estimado con límites de significancia del 5% para las autocorrelaciones. ....                 | 77 |
| <b>Figura 15</b> Distribución de normalidad de los residuos del modelo .....   | 78 |
| <b>Figura 16</b> Ploteo de la normalidad de los residuos del modelo estimado.....  | 78 |
| <b>Figura 17</b> Transformación de box-cox .....   | 80 |
| <b>Figura 18</b> Valores de la transformación box-cox de recaudaciones tributarias mensuales, con respecto a tributos internos de los años 2004 – 2017. ....                 | 81 |
| <b>Figura 19</b> Valores reales y ajustados de la serie transformada. ....   | 83 |
| <b>Figura 20</b> Función de autocorrelación de los residuos del modelo estimado con límites de significancia del 5% para las autocorrelaciones. ....                         | 86 |
| <b>Figura 21</b> Función de autocorrelación parcial de los residuos del modelo de la serie transformada con límites de significancia del 5% para las autocorrelaciones. .... | 87 |
| <b>Figura 22</b> Distribución de normalidad de los residuos de la serie transformada.....  | 88 |
| <b>Figura 23</b> Normalidad de los residuos nuevo modelo de los datos con transformación box-cox.....  | 89 |
| <b>Figura 24</b> Pronostico de la serie transformada .....   | 90 |



## ÍNDICE DE TABLAS

|   | <b>Pág.</b> |
|---|-------------|
| <b>Tabla 1</b> Comportamientos de las $f_{ac}$ y $f_{acp}$ .....  | 51          |
| <b>Tabla 2</b> Operacionalización de variables .....  | 63          |
| <b>Tabla 3</b> Serie histórica mensual en miles de soles de la recaudación tributaria, con respecto a tributos internos de la región de puno, periodo 2004 – 2017. .... | 64          |
| <b>Tabla 4</b> Resumen de los parámetros del modelo arima (0, 1,1) (0,1,1) .....  | 73          |
| <b>Tabla 5</b> Estadística chi-cuadrada modificada de box-pierce (ljung-box) para el modelo de la serie transformada .....  | 76          |
| <b>Tabla 6</b> Resumen de los parámetros del modelo de la serie transformada .....  | 82          |
| <b>Tabla 7</b> Bondad de ajuste del modelo de la serie transformada.....  | 83          |
| <b>Tabla 8</b> Estadística chi-cuadrada modificada de box-pierce (ljung-box) para el modelo de la serie transformada .....  | 85          |
| <b>Tabla 9</b> Pronostico del modelo de la serie transformada de la recaudación de tributos internos en la región de puno para el año 2018 .....                        | 91          |



## ÍNDICE DE ACRÓNIMOS

|        |  |
|--------|--|
| GEM    | Gravamen Especial a la Minería.  |
| IGV    | Impuesto General a las Ventas  |
| IEPDTN | Impuesto Extraordinario para la Promoción y Desarrollo Turístico Nacional. |
| IEMCT  | Impuesto Especial a la Minería, Casinos y Tragamonedas.                    |
| ITAN   | Impuesto Temporal a los Activos Netos.                                     |
| ITF    | Impuesto a las Transacciones Financieras.                                  |
| IR     | Impuesto a la Renta  |
| ISC    | Impuesto Selectivo al Consumo  |
| MEF    | Ministerio de Economía y Finanzas  |
| NRUS   | Nuevo Régimen Único Simplificado   |
| SUNAT  | Superintendencia Nacional de Aduanas y de Administración Tributaria        |
| UIT    | Unidad Impositiva Tributaria   |
| RER    | Régimen Especial del Impuesto a la Renta                                   |
| REG    | Régimen General del Impuesto a la Renta                                    |
| RM     | Regalías Mineras.  |



## RESUMEN

El presente trabajo de investigación tiene como principal finalidad determinar el modelo de predicción de los ingresos tributarios recaudados por la SUNAT, referido a tributos internos en la Región de Puno 2004 – 2017. La población está constituida por el total de ingresos por tributos internos registrados mensualmente, en la base de datos de la Administración Tributaria, respecto a la región de Puno que se encuentren registrados y publicados por SUNAT, para el desarrollo de la presente investigación se realizó el uso de la metodología de Box Jenkins, la muestra del presente estudio fueron de 168 datos mensuales de recaudación de tributos internos correspondientes a la Región de Puno, comprendidos en el periodo de 2004 a 2017, planteándose como hipótesis que el modelo de predicción ARIMA proporcionara un ajuste aceptable para describir y predecir la recaudación por la SUNAT, referido a tributos internos en la Región de Puno 2014 – 2017. Llegando a la siguiente conclusión, el modelo estimado ARIMA (0, 1, 1) (0, 1, 1)<sub>12</sub> con transformación Box-Cox Siendo  $\lambda = 0$ , la nueva transformación es  $\ln y_t$  pudo proyectar numéricamente los pronósticos, permitiendo obtener resultados a futuro. La información proyectada es de 12 meses, teniendo como modelo obtenido  $Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} + 0.6261\varepsilon_{t-1} - 0.8974\varepsilon_{t-12} + 0.5619\varepsilon_{t-13}$ , siendo los pronósticos de alta precisión con un MAPE de 0.99%, a partir de la información existente de la cantidad en miles de soles recaudados por tributos internos de la SUNAT.

**Palabras Clave:** Recaudación, Modelo, Predicción, Series de Tiempo, Tributos Internos.



## ABSTRACT

The main purpose of this research work is to determine the prediction model of tax revenues collected by SUNAT, referring to internal taxes in the Region of Puno 2004 - 2017. The population is made up of the total internal tax revenues registered monthly in the Tax Administration's database for the region of Puno, which are registered and published by SUNAT, the sample of the present study was 168 monthly data of internal tax collection corresponding to the Region of Puno, comprised in the period from 2004 to 2017, posing as a hypothesis that the ARIMA prediction model would provide an acceptable adjustment to describe and predict the collection by SUNAT, referred to internal taxes in the Region of Puno 2014 - 2017. Reaching the following conclusion, the estimated model ARIMA (0, 1, 1) (0, 1, 1)<sub>12</sub> with Box-Cox transformation Being  $\lambda = 0$ , the new transformation is  $\ln y_t$  was able to numerically project the forecasts, allowing to obtain results in the future. The projected information is 12 months, having as a model obtained  $Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} + 0.6261\varepsilon_{t-1} - 0.8974\varepsilon_{t-12} + 0.5619\varepsilon_{t-13}$ , being the forecasts of high precision with a MAPE of 0.99%, based on the existing information of the amount in thousands of soles collected by SUNAT internal taxes.

**Keywords:** collection, model, prediction time series, and internal taxes.



# CAPÍTULO I

## INTRODUCCIÓN

La Superintendencia Nacional de Aduanas y de Administración Tributaria (SUNAT) es una institución del estado que tiene por finalidad primordial administrar los tributos del gobierno nacional y los conceptos tributarios y no tributarios que se le encarguen por ley o de acuerdo a los convenios interinstitucionales que se celebren, proporcionando los recursos requeridos para la solvencia fiscal y la estabilidad macroeconómica; asegurando la correcta aplicación de la normatividad que regula la materia y combatiendo los delitos tributarios y aduaneros conforme a sus atribuciones, asimismo cuenta con las siguientes funciones; a) Administrar todos los tributos internos con excepción de los municipales. b) Recaudar todos los tributos internos con excepción de los municipales, a través del Banco de la Nación, pudiendo este suscribir convenios con otras entidades bancarias. c) Fiscalizar el cumplimiento a las obligaciones tributarias, a efecto de combatir la evasión fiscal. d) Otorgar el aplazamiento y/o fraccionamiento para el pago de la deuda tributaria de acuerdo con la ley. e) Ejercer los actos de coerción para el cobro de la deuda tributaria por medio del Juzgado Coactivo competente.

Saber el comportamiento de la cantidad recaudada de ingresos por tributos internos en un futuro es muy importante ya que la información anticipada permite tomar decisiones para la prevención y planificación efectiva de la estabilidad macroeconómica ante de los incidentes que puedan surgir.

La metodología de Box Jenkins, es ampliamente utilizada en diversos campos de la ciencia, que ha tenido bastante éxito modelando datos que fluctúan a través del tiempo, siempre que se cumplan los supuestos teóricos, por ello en el presente trabajo de



investigación se busca utilizarlo correctamente, utilizando para el análisis de la serie temporal registros que ascienden a 168 datos mensuales de recaudación de tributos internos correspondientes a la Región de Puno, comprendidos en el periodo de 2004 a 2017.

En ese sentido, identificar las metodologías más idóneas para predecir valores futuros sobre indicadores correspondientes al sector tributario, se desarrolla aquí un breve marco teórico, así como su desarrollo sobre el indicador de tributación interna en la Región de Puno.

El presente proyecto de investigación cuenta con cuatro capítulos, donde:

En el capítulo I, se explica los fundamentos para la realización de la tesis, se describen los objetivos y las hipótesis.

En el capítulo II, se describe el marco teórico y presenta los diversos conceptos necesarios para el correcto entendimiento de la tesis.

En el capítulo III, se describe los materiales y métodos para la obtención del modelo de predicción univariante de Box – Jenkins para la recaudación de tributos internos en la Región de Puno.

En el capítulo IV, se muestra las conclusiones y las recomendaciones del modelo de predicción mensual con respecto a la recaudación de tributos internos en la Región de Puno correspondiente a los periodos 2004-2017.



## 1.1 PLANTEAMIENTO DEL PROBLEMA

Actualmente nos encontramos inmersos en un sistema tributario que cuenta con una complicada normatividad tributaria, ya que los ingresos por tributación interna recaudados por la SUNAT son uno de los principales ejes en la actividad económica del estado la cual fluctúa con respecto a la temporada incrementa o disminuye, sin embargo en la actualidad no se cuenta con estudios que nos permitan el pronóstico adecuado de la recaudación de tributos internos en las regiones de nuestro país, ya que se cuenta con una variedad geográfica y climática las cuales dificultan el estudio de los ingresos que recaudación interna en cada una de las regiones, la Superintendencia Nacional de Aduanas y de Administración Tributaria es un órgano adscrito del Ministerio de Economía y Finanzas (MEF), por tal motivo impone metas anuales para la recaudación de tributos. Motivo por el cual existe la necesidad de realizar los pronósticos para los meses venideros y/o años en la Región de Puno, y así contar con una herramienta para realizar los pronósticos y así prever el comportamiento y variaciones que puedan presentarse en el transcurso de los meses al realizar la recaudación de tributos internos, lo que nos permite tomar decisiones informadas.

La intención con la que cuenta la presente investigación es, que utilizando la metodología Box – Jenkins obtener un modelo de predicción de los ingresos tributarios recaudados por la SUNAT, referido a tributos internos en la Región de Puno 2004 – 2017 a fin de pronosticar el monto de ingresos de los tributos internos, el cual permitirá realizar el cobro de los tributos correspondientes a los tributos internos los cuales son: el Impuesto General a las Ventas, Impuesto a la Renta, Impuesto Selectivo al Consumo, Nuevo Régimen Único Simplificado, Fraccionamiento, Impuesto Extraordinario de Solidaridad, Multas, Impuesto Temporal a los Activos Netos, Impuesto a las Transacciones



Financieras, Impuesto de Solidaridad a la Niñez Desamparada, Impuesto Extraordinario de Promoción Turística, Impuesto a las Acciones del Estado, Impuesto a los Casinos de Juego y Máquinas Tragamonedas, así programar, organiza evaluar y ejecutar el cobro de los tributos, detectando la evasión tributaria en la Región de Puno.

## **1.2 FORMULACIÓN DEL PROBLEMA**

La pregunta que se da para dar solución al problema es:

¿Cuál es el modelo de predicción para pronosticar los ingresos de tributación recaudados por la SUNAT, referido a tributos internos en la Región de Puno 2004-2017?

## **1.3 OBJETIVOS DE LA INVESTIGACIÓN**

### **1.3.1. Objetivo General**

Determinar el modelo de predicción de los ingresos tributarios recaudados por la SUNAT, referido a tributos internos en la Región de Puno 2004 – 2017.

### **1.3.2. Objetivos específicos**

- Determinar si la serie de tiempo cuenta con tendencia creciente.
- Determinar si la serie de tiempo cuenta con estacionalidad.
- Realizar los pronósticos para los 24 meses del año 2018 - 2019 mediante el modelo determinado.



## **1.4 HIPÓTESIS DE LA INVESTIGACIÓN:**

### **1.4.1 Hipótesis General**

El modelo de predicción ARIMA proporcionara un mejor ajuste para describir y predecir la recaudación por la SUNAT, referido a tributos internos en la Región de Puno 2014 – 2017.

### **1.4.2 Hipótesis específica:**

- La serie de tiempo de la presente investigación cuenta con una tendencia creciente.
- La serie de tiempo de la presente investigación cuenta con estacionalidad.



## CAPÍTULO II

### REVISIÓN DE LITERATURA

#### 2.1 ANTECEDENTES DE LA INVESTIGACIÓN

##### 2.2.1. A nivel Internacional

Mendieta & Rodríguez (2004) Concluyen que el Impuesto al Valor Agregado (I.V.A) ha sido administrado su recaudación de forma técnica y eficiente en especial desde la creación del Servicio de Rentas Internas en 1998 y gracias a la misma el estado puede en gran parte solventar los gastos públicos.

Y que el alto valor del impuesto que combinado con el pago de otros impuestos hace que los contribuyentes no declaren el IVA debido a la mala situación económica.

Los contribuyentes que evaden priorizan sus intereses; las necesidades de la comunidad como salud, educación, etc. pasan a un segundo plano. Esto da como resultado un enriquecimiento individual que no distribuye eficientemente los recursos pertenecientes a la sociedad.

Amaya (2014) Concluye que: Los impuestos constituyen una herramienta vital para alcanzar el bienestar de la sociedad. En su calidad de ingresos permanentes, los impuestos están destinados a cubrir los gastos permanentes del Estado, como la provisión de salud, educación, defensa, administración de justicia, orden interno e infraestructura, entre otros bienes y servicios públicos. Debido a la importancia de los impuestos en la economía, la hacienda pública, debe procurar recaudar todo



aquello que el presupuesto general del estado ha estimado como ingresos tributarios, es decir, una buena política tributaria deberá cumplir con el principio de suficiencia recaudatoria.

Solera (2015) Concluye que: el índice mensual de actividad económica impacta de manera significativa la recaudación solo durante el sexto mes. Las estimaciones utilizan un índice de actividad que incluye todos los sectores económicos, sin importar su tratamiento fiscal. Dado que los sectores económicos más importantes de este índice tienen un tratamiento fiscal diferenciado, al inicio la influencia de estas actividades es mínima e inexistente sobre la recaudación. No obstante el efecto derrame de estas, en relación a los encadenamientos productivos que generan, se estarían reflejando en pago de impuestos hasta un semestre posterior.

Juez & Martillo (2017) Concluye que: Al elaborar el modelo econométrico logarítmico relacionando la presión fiscal con las variaciones del PIB como medida de crecimiento de la economía se obtuvo  $\log PIB = 4.65883 + 0.63793 \log PresiónFiscal + \mu$  que se interpreta que, en promedio, cuando la presión fiscal se incrementa en un 1%, el PIB se incrementa en 0.63%, aproximadamente, comprobándose la hipótesis que el incremento de la presión impactó positivamente en el crecimiento de la economía ecuatoriana en el período estudiado, esto es, del 2007 al 2016.

### **2.2.2. A nivel Nacional**

Villalobos (2014) Afirma que la serie de ingresos tributarios recaudados de Impuesto General a las Ventas a nivel nacional, por la SUNAT, durante el periodo



enero del 2004 a diciembre del 2013; presentan la componente de tendencia creciente, estacionalidad creciente en enero de cada año; además es heterocedástica.

Ramos (2020) Sostiene que los ingresos tributarios por meses del gobierno central desde el año 2003 hasta el 2018. Podemos apreciar un comportamiento creciente a lo largo de los años, y además se observan valores “altos” a inicios de cada año, siendo menos pronunciados en los primeros años, lo que indica un comportamiento estacional de periodicidad 12 por ser datos mensuales.

También que los ingresos tributarios empiezan con un valor alto en el mes de enero, disminuyen en el mes de febrero, y luego crece hasta llegar al punto más alto en abril y luego disminuye en mayo y se mantiene casi constante los siguientes meses del año, y este patrón se repite cada año. De esto se deduce que la serie de tiempo de los ingresos tributarios mensuales tiene un comportamiento estacional. En la validación del modelo SARIMA (2,1,0)(1,1,1) se determina que los errores del modelo cumplen con los supuestos de normalidad, estacionariedad e independencia y por lo tanto el modelo es apto para realizar pronósticos

De La Roca & Hernández (2004) Concluyen que el número de asalariados formales en la ENNIV 2000 disminuye al tamaño de la informalidad, dado que se contabiliza únicamente el ingreso no reportado y no la totalidad (ingreso de trabajador informal), adicionalmente indica que no existe registro de los trabajadores con los estratos sociales más altos de esta sociedad, ya que las remuneraciones que perciben elevarían significativamente el monto promedio de los ingresos de los asalariados independientes (formales). Una evidencia de la ausencia de estos trabajadores es la baja participación de las remuneraciones con porcentajes del PBI 35%.



Mostacero & Vásquez (2014) Concluyen que: El análisis comparativo de la aplicación del programa de orientación y cultura tributaria en estudiantes de quinto año de secundaria de la ciudad de Trujillo, muestra que existe poco interés sobre temas tributarios en la educación impartida por las I.E. Públicas y Privadas. Deduciendo así que existe la necesidad de crear cultura tributaria que permita a los futuros ciudadanos concebir las obligaciones tributarias como un deber sustantivo. Un mayor nivel de conciencia cívica respecto al cumplimiento tributario, permitirá al país disminuir los elevados índices de evasión, informalidad y corrupción.

Santiago, Coronel, & Galarza (2011) Concluyen que: El desagio tributario es una estrategia eficaz que debe ser usada por la administración tributaria (SUNAT) para incrementar los niveles de ingresos en el país, obteniendo un efecto esperado mediante una adecuada educación y concientización de los contribuyentes, lo que repercutiría en mayores ingresos que serán aplicados en el desarrollo del país.

Quintanilla (2014) Concluye que: la evasión de impuestos, ocasiona disminución en el nivel de fondos que maneja el gobierno, el análisis de los datos permitió conocer que el ocultar bienes o ingresos para pagar menos impuestos, determinan el nivel de ingresos tributarios, también permitió determinar que el incumplimiento de obligaciones tributarias, influye en el conjunto de gravámenes, impuestos y tasas que pagan las personas y que infracciones administrativas, obedecen a las sanciones impuestas por la administración tributaria.

Grández (2014) Concluye que: El estudio ha demostrado que existe una relación directa entre la presión tributaria y la recaudación fiscal, medida a través del ingreso tributario real per cápita, para el Perú durante el periodo 1990 y 2012; es



decir que ante un incremento de la presión tributaria la recaudación real por persona también incrementó durante este periodo.

Para el periodo 1990-2012, los ingresos tributarios representan en promedio el 87,6% del total de los ingresos del Gobierno Central. Se observa claramente a la tendencia creciente de la recaudación por impuesto a la Renta e Impuesto General a las Ventas (IGV) ha ido incrementando su participación, representando en la actualidad el 97% de los ingresos tributarios del Gobierno Central.

### **2.2.3. A nivel Local**

Cabana (2017) Concluye que la principal causa de la evasión tributaria en el mercado central del Distrito de Santa Lucia es la falta de cultura tributaria, ya que dicho análisis indica que el 53% de comerciantes muestran muy poca cultura tributaria, las causas políticas que genera la evasión tributaria se da por la desconfianza del contribuyente con el estados, resultando así que el 35% de la población encuestada afirma que es consiente que evade impuestos, proponiendo de esa manera incrementar programas sobre educación tributaria, basada en prácticas de valores con resultados a futuro favorables para la administración tributaria y el país.

Coila (2013) Concluye que el IGV, IR e ISC con el PBI, ha aumentado en relación al periodo de estudio del 2000 al 2012, Es decir, con un incremento de un 1% en el PBI la recaudación del IGV, IR e ISC aumenta en un 1.97%, 3.04% y 0.23%, respectivamente. Indicando que a su vez, el coeficiente de regresión es mayor al coeficiente de la constante, por lo tanto esto muestra que las variaciones en el PIB no han sido determinantes para el IGV, IR y entonces existen otras variables que las explican. La recaudación tributaria está constituida por el impuesto general a las



ventas (49%) y el impuesto a la renta (42%), indicando que la posición fiscal ha mejorado significativamente en los últimos años, sin embargo, la economía en los últimos los ha tenido en un superávit convencional pero que ha llegado a presentar un difícil estructural.

Condori (2014) Concluye que el modelo estacional autorregresivo integrado de la media móvil SARIMA (0, 1, 1) (0, 1, 1)<sup>4</sup> es confiable de pronóstico para los pagos del impuesto predial para la gerencia de administración tributaria Puno, la cual satisface las pruebas del Ljung – box. Price y el criterio de Akaike del mejor modelo.

Los antecedentes de la investigación citados contribuyen en el desarrollo del trabajo, proporcionando información que se utiliza en la discusión sobre todo aquellas que tienen resultados similares al nuestro o aquellas que no llegan a conclusiones parecidas, en ambos casos aportan sustancialmente a mi proyecto de tesis.

## 2.2 MARCO TEÓRICO

**Base imponible:** Es el valor sobre el cual se aplica el impuesto (Farfán, 2006).

**Impuesto:** Es el tributo cuyo cumplimiento no origina una contraprestación directa en favor del contribuyente por parte del Estado (Farfán, 2006).

**Obligación Tributaria:** La obligación tributaria, que es de derecho público, es el vínculo entre el acreedor y el deudor tributario, establecido por ley, que tiene por objeto el cumplimiento de la prestación tributaria, siendo exigible coactivamente (Farfán, 2006).

**Pago:** Es la forma común de extinción de la obligación tributaria (Farfán, 2006).



**Tesoro Público:** Es el patrimonio público a cargo del Gobierno Central. Los ingresos del tesoro público provienen de la aplicación de impuestos, tasas, contribuciones, rentas de propiedad, multas y otras sanciones, etc. También se le denomina erario o fisco (Farfán, 2006).

**Unidad Impositiva Tributaria UIT:** Es un valor de referencia que puede ser utilizado en las normas tributarias para determinar las bases imponibles, deducciones, límites de afectación y demás aspectos de los tributos que considere conveniente el legislador. También podrá ser utilizada para aplicar en otras obligaciones formales (García, 2018).

**Contribuyente:** Aquél que realiza, o respecto del cual se produce el hecho generador de la obligación tributaria. Se considera deudor por cuenta propia, ya que realiza el hecho imponible. Son los obligados directos al pago del tributo por mandato de ley (García, 2018).

**Régimen tributario:** Cualquiera de los regímenes establecidos para el cumplimiento de la obligación tributaria: Régimen Único Simplificado - RUS, Régimen Especial del Impuesto a la Renta - RER o Régimen General del Impuesto a la Renta. Son ficciones legales que determinan formas de afectación al Impuesto a la Renta teniendo en cuenta la actividad económica que el contribuyente realice, a la que adicionalmente se añadirá el Impuesto General a las Ventas u otros tributos ya sea en forma disgregada o unificándose como un solo tributo (Decreto Legislativo N° 1270, 2004).

**SUNAT:** La Superintendencia Nacional de Aduanas y de Administración Tributaria, es un organismo técnico especializado, adscrito al Ministerio de Economía y



Finanzas, con personería jurídica de derecho público, con patrimonio propio y goza de autonomía funcional, técnica, económica, financiera, presupuestal y administrativa (García, 2018).

**Impuesto General a las Ventas (IGV):** Es el impuesto que se aplica en las operaciones de venta e importación de bienes, así como en la prestación de distintos servicios comerciales, en los contratos de construcción o en la primera venta de inmuebles (Decreto Legislativo N° 1270, 2004).

**Impuesto a la Renta (IR):** Es aquél que se aplica a las rentas que provienen del capital, del trabajo o de la aplicación conjunta de ambos (Decreto Legislativo N° 1270, 2004).

**Impuesto Selectivo al Consumo (ISC):** Es el impuesto que se aplica sólo a la producción o importación de determinados productos como cigarrillos, licores, cervezas, gaseosas, combustibles, etc. (Decreto Legislativo N° 1270, 2004).

**Régimen Especial del Impuesto a la Renta (RER):** Es un régimen tributario dirigido a personas naturales y jurídicas, sucesiones indivisas y sociedades conyugales domiciliadas en el país que obtengan rentas de tercera categoría provenientes de las actividades de comercio y/o industria; y actividades de servicios.

**Nuevo Régimen Único Simplificado (NRUS):** Es un régimen simple que establece un pago único por el Impuesto a la Renta y el Impuesto General a las Ventas (incluyendo al Impuesto de Promoción Municipal). A él pueden acogerse únicamente las personas naturales o sucesiones indivisas, siempre que desarrollen actividades generadoras de rentas de tercera categoría (bodegas, ferreterías, bazares, puestos de



mercado, etc.) y cumplan los requisitos y condiciones establecidas (Decreto Legislativo N° 1270, 2004).

**Tributo:** Es una prestación de dinero que el Estado exige en el ejercicio de su poder de imperio sobre la base de la capacidad contributiva en virtud de una ley, y para cubrir los gastos que le demande el cumplimiento de sus fines (Decreto Legislativo N° 133-2013-EF, 2013).

El Código Tributario establece que el término TRIBUTO comprende impuestos, contribuciones y tasas.

**Impuesto:** Es el tributo cuyo pago no origina por parte del Estado una contraprestación directa en favor del contribuyente. Tal es el caso del Impuesto a la Renta (Decreto Legislativo N° 133-2013-EF, 2013).

**Contribución:** Es el tributo que tiene como hecho generador los beneficios derivados de la realización de obras públicas o de actividades estatales, como lo es el caso de la Contribución al SENCICO (Decreto Legislativo N° 133-2013-EF, 2013).

**Tasa:** Es el tributo que se paga como consecuencia de la prestación efectiva de un servicio público, individualizado en el contribuyente, por parte del Estado. Por ejemplo, los derechos arancelarios de los Registros Públicos (Decreto Legislativo N° 133-2013-EF, 2013).

## **SERIES DE TIEMPO**

Los análisis estadísticos disponen de dos tipos de series de datos; los de corte transversal y los de corte longitudinal (series de tiempo).



Una serie de tiempo es un conjunto de datos observados en periodos de tiempo distintos, generalmente de forma secuencial, que reflejan cambios en el tiempo a determinado sujeto para determinado caso.

(Spiegel & Stephens, 2001, pág. 436) Una serie de tiempo es un conjunto de observaciones tomadas en momento o tiempos específicos, generalmente a intervalos iguales. Matemáticamente, una serie de tiempo se define por medio de valores  $Y_1, Y_2, \dots$  de una variable  $Y$  en los tiempos  $t_1, t_2, \dots$ . Por lo tanto,  $Y$  es una función de  $t$ ; esto se denota por  $Y = F(t)$ .

(Gujarati & Porter, 2010, pág. 22) Una serie de tiempo es un conjunto de observaciones sobre los valores de una variable en diferentes momentos. Tal información debe recopilarse en intervalos regulares, es decir, en forma diaria (precios de acciones, informes del tiempo, etc.), semanal (como cifras de oferta monetaria), mensual (tasa de desempleo, Índice de Precios al Consumidor [IPC], etc.), trimestral (como el PIB), anual (como los presupuestos del gobierno), quinquenal (como el censo de la industria manufacturera), o decenal (como los censos de población). Algunas veces los datos están disponibles por trimestre y por año, como los datos del PIB y del consumo.

## **MOVIMIENTOS O PATRONES DE UNA SERIE TEMPORAL.**

Los movimientos característicos de las series de tiempo pueden clasificarse en cuatro tipos principales, llamados componentes de las series de tiempo.

### **Movimiento a largo plazo o seculares (Tendencia)**

Esto se refiere a la dirección general en la que la gráfica de series de tiempo parece seguir sobre un intervalo grande de tiempo. (Spiegel & Stephens, 2001)



### **Movimientos cíclicos o variaciones cíclicas**

Estos tienen que ver con las oscilaciones o los movimientos respecto a una recta o curva de tendencia. Estos ciclos, como se les denomina en ocasiones, son o no periódicos; es decir, pueden o no seguir patrones exactamente similares, después de intervalos iguales de tiempo. En los negocios y las actividades económicas, los movimientos son considerados cíclicos solo si se repiten después de intervalos mayores a un año. (Spiegel & Stephens, 2001)

### **Movimientos estacionales o variaciones estacionales**

Estos se relacionan con los patrones idénticos o casi idénticos que las series de tiempo parecen seguir durante los meses o trimestres correspondientes de años sucesivos. Tales movimientos se deben a eventos recurrentes que suceden anualmente, como el repentino incremento en las ventas de una tienda de departamento, previo a la Navidad. Los movimientos estacionales suelen conocerse en la teoría de negocios o economía como periodicidad anual, las ideas implicadas llegan a extenderse para un periodo determinado (como días, horas, semanas), dependiendo del tipo de datos disponibles. (Spiegel & Stephens, 2001)

### **Movimientos irregulares o aleatorios**

Estos se refieren a los movimientos esporádicos de las series de tiempo, debidos a eventos aleatorios tales como: inundaciones, huelgas o elecciones. Aunque generalmente se considera que tales eventos producen variaciones que duran poco tiempo, cabe la posibilidad de que sean tan intensos que resulten en nuevos movimientos cíclicos o de otro tipo. (Spiegel & Stephens, 2001)



## CLASIFICACIÓN DESCRIPTIVA DE LAS SERIES TEMPORALES

### **Estacionariedad en varianza.**

Una serie será estacionaria en varianza cuando pueda mantenerse el supuesto de que existe una única varianza para toda la serie temporal, es decir, cuando la variabilidad de la serie en torno a su media se mantenga constante a lo largo del tiempo. Si la serie no es estacionaria en varianza, se utilizan las transformaciones estabilizadoras de varianza, es decir, las transformaciones Box-Cox. (González Casimiro, 2009, pág. 52)

### **Estacionarias**

Una serie de tiempo estacionara es aquella cuya propiedades estadísticas básicas, tales como la media y la varianza, permanecen constantes a lo largo del tiempo. En consecuencia, una serie que varía alrededor de un nivel fijo a lo largo del tiempo (sin crecimiento ni decremento) se dice que es estacionaria. (Guerrero Guzmán, 2003, pág. 65)

### **No estacionarias**

Una serie que contiene una tendencia es no estacionaria. Los coeficientes de autocorrelación para una serie estacionaria caen hacia cero bastante rápido, en general, después del segundo o tercer retraso del tiempo. Por otro lado, las autocorrelaciones de muestras permanecen bastante grandes en una serie no estacionaria durante varios periodos. A menudo para analizar las series no estacionarias se elimina la tendencia antes de realizar modelos adicionales. (Guerrero Guzmán, 2003, pág. 65)



## MODELO

Los datos de series temporales se han de desarrollar modelos específicos que recojan y aprovechen la dependencia entre las observaciones ordenadas de una serie temporal.

El conjunto de técnicas de estudio de series de observaciones dependientes ordenadas en el tiempo se denomina Análisis de Series Temporales. El instrumento de análisis que se suele utilizar es un modelo que permita reproducir el comportamiento de la variable de interés. Los Modelos de Series Temporales pueden ser:

**Univariantes:** Solo se analiza una serie temporal en función de su propio pasado.

**Multivariantes:** Se analizan varias series temporales a la vez. Un ejemplo muy popular en la literatura son las series de número de pieles de visón y rata almizclera capturadas en Canadá. Se sabe que existe una relación víctima-depredador entre ambos animales lo que se supone que afecta a la dinámica de ambas series. La forma de reflejar estas interacciones dinámicas entre ambas series es construir un modelo multivalente. Cuando se construye un modelo multivariante, para casos como este, suponemos que hay cierta dependencia o relación entre los pasados de las diversas series.

Estas interacciones dinámicas también aparecen cuando construimos modelos multivalentes para variables económicas, tales como la renta, consumo e inversión que, como es bien sabido, influyen las unas en las otras. (González Casimiro, 2009, pág. 4)

## TIPOS DE SERIES TEMPORALES SEGÚN SUS COMPONENTES

Son formas teóricas determinísticas y/o aleatorias o la combinación de ambas, para realizar el análisis de una serie de tiempo.



**Variables Temporales:** Son variables que se observan a lo largo del tiempo.

En Donde:  $Y_t$  Indica la variable “Y” en el momento “t”.

**Serie Temporal:** Es el conjunto de “t” observaciones, una observación por cada una de las variables:

$Y_1, Y_2, Y_3, \dots, Y_t$ . También es denominada serie cronológica.

Existen tres tipos de series de tiempo según sus componentes. Que generalmente se aceptan como buenas aproximaciones a las verdaderas relaciones, entre los componentes de los datos observados (García, 2016).

Estos son:

- Aditivo:  $Y(t) = T(t) + E(t) + C(t) + A(t)$
- Multiplicativo:  $Y(t) = T(t) * E(t) + C(t) * A(t)$
- Mixto:  $Y(t) = T(t) * E(t) * C(t) + A(t)$

Donde:

$Y(t)$ : Serie observada en instante t.

$T(t)$ : Componente de Tendencia. Es el movimiento suave y regular de la serie a largo plazo. Refleja la dirección del movimiento de una determinada variable; creciente, decreciente o estable (Catalán, 2004).

$C(t)$ : Componente de Ciclicidad. Consiste en variaciones superiores al año que no son estrictamente periódicas. Se trata de un factor de tipo oscilante



caracterizado por movimientos recurrentes en torno a la tendencia, y suele aparecer fundamentalmente en series de tipo económico (Catalán, 2004).

$E(t)$ : Componente Estacional. Son los movimientos regulares de la misma serie que tienen una periodicidad inferior al año. Recoge las oscilaciones que año a año se repiten en una serie de forma periódica (Catalán, 2004).

$A(t)$ : Componente Aleatoria (accidental). Una suposición usual es que  $A(t)$  sea una componente aleatoria o ruido blanco con media cero y varianza constante (Catalán, 2004).

Los modelos más utilizados para descomponer una serie de tiempo son el aditivo y el multiplicativo. El modelo menos utilizado es el modelo log-aditivo.

La mayoría de las series de tiempo económicas siguen un modelo multiplicativo.

En los casos en que la serie presenta valores negativos o ceros, el único modelo aplicable es el aditivo. (Instituto Nacional de Estadística e Informática (INEI), 2002).

### **Procesos estocásticos estacionarios**

Un tipo de proceso estocástico que ha recibido gran atención y ha sido objeto de escrutinio por parte de los analistas de series de tiempo es el proceso estocástico estacionario. En términos generales, se dice que un proceso estocástico es estacionario si su media y su varianza son constantes en el tiempo y si el valor de la covarianza entre dos periodos depende sólo de la distancia o rezago entre estos dos periodos, y no del tiempo en el cual se calculó la covarianza. En la bibliografía sobre series de tiempo, un proceso estocástico como éste se conoce como proceso estocástico débilmente estacionario, estacionario covariante, estacionario de segundo orden o proceso estocástico en amplio



sentido. Para efectos de este capítulo, y en la mayoría de las situaciones prácticas, basta este tipo de estacionariedad. (Gujarati & Porter, 2010, pág. 741)

## RUIDO BLANCO

El proceso estocástico más sencillo es el denominado Ruido Blanco que es una secuencia de variables aleatorias de media cero, varianza constante y covarianzas nulas.

Se denotará habitualmente por:

- i.  $E(\varepsilon_t) = 0 \quad \forall t$
- ii.  $V(\varepsilon_t) = E(\varepsilon_t^2) = \sigma^2 \quad \forall t$
- iii.  $COV(\varepsilon_t) = E(\varepsilon_t, \varepsilon_s) = 0 \quad \forall t \neq s$

Así, un proceso ruido blanco,  $a_t \sim RB(0, \sigma^2)$  es estacionario si la  $\sigma^2$  es finita con función de autocovarianzas (FACV):

$$\gamma_k = \sigma^2; k = 0 \text{ y } \gamma_k = 0; k > 0$$

y función de autocorrelación (FAC):

$$\text{iv. } \gamma_k = 1; k = 0 \text{ y } \gamma_k = 0; k > 0$$

Esto es, la variable " $Y_t$ " presenta una esperanza constante en el tiempo, e igual a 0; varianza constante y autocorrelación nula. Si se agrega la condición de que la variable " $X_t$ " se distribuye normal, la condición (iii) de incorrelación implica independencia.

Supondremos, salvo que se exprese lo contrario, que la distribución de probabilidad asociada a un ruido blanco es normal. Ello significa que las 3 condiciones anteriores pueden sintetizarse en la siguiente expresión:



$$a_t \sim Niid (0, \sigma^2) \forall t$$

*Niid*: Normales independientes e idénticamente distribuidas.

El término ruido ha sido tomado de la teoría de la comunicación. En esta disciplina, un ruido corresponde a perturbaciones aleatorias que vienen en función de la señal enviada, que es diferente de la señal recibida. Se denomina ruido blanco a una perturbación aleatoria con las características antes planteadas, por analogía con la distribución de energía continua en luz blanca de un cuerpo incandescente. (Uriel Jiménez, 1985).

## **METODOLOGÍA BOX JENKINS**

Popularmente conocida como metodología de Box-Jenkins (BJ), pero técnicamente conocida como metodología ARIMA, el interés de estos métodos de pronósticos no está en la construcción de modelos uniecuacionales o de ecuaciones simultáneas, sino en el análisis de las propiedades probabilísticas, o estocásticas, de las series de tiempo por sí mismas según la filosofía de que los datos hablen por sí mismos. A diferencia de los modelos de regresión, en los cuales  $Y_t$  se explica por las  $k$  regresoras  $X_1, X_2, X_3, \dots, X_K$ , en los modelos de series de tiempo del tipo BJ,  $Y_t$  se explica por valores pasados o rezagados de sí misma y por los términos de error estocásticos. (Gujarati & Porter, 2010, pág. 775)

## **MODELO UNIVARIANTE DE BOX-JENKINS NO- INTEGRADOS**

Los procesos de Medias móviles MA ( $q$ ), autorregresivos AR ( $p$ ), y procesos mixtos ARMA ( $p, q$ ) son considerados como los modelos no Integrados.



**a) Proceso autorregresivo de primer orden (AR(p))**

$$Y_t = \varphi_1 Y_{t-1} - \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + e_t$$

Donde:

$Y_t$ : Variable respuesta (dependiente) en el tiempo t.

$Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots, Y_{t-p}$  : Variable respuesta en los retardos  
 $Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots, Y_{t-p}$  desempeña la función de las variables independientes.

$\varphi_1, \varphi_2, \varphi_3, \dots, \varphi_p$  : Coeficientes que serán estimados.

$p$ : Retardo máximo (Orden del modelo).

$e_t$  : Terminio de error en el tiempo t.

**b) Media móvil MA(q)**

$$Y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

Donde:

$Y_t$  : Variable respuesta (dependiente) en el tiempo t.

$a_{t-1}, a_{t-2}, a_{t-3}, \dots, a_{t-p}$ : Errores que siguen el proceso puramente aleatorio desplazados en q periodos a la que se ajusta la serie.

$\theta_{t-1}, \theta_{t-2}, \theta_{t-3}, \dots, \theta_{t-p}$ : Los parámetros a estimar.

$q$ : Retardo máximo (Orden del modelo).

$a_t$ : Variable llamada ruido blanco.

## MODELO UNIVARIANTE DE BOX-JINKINS INTEGRADOS

A los procesos mixtos integrados ARIMA (p, d, q), los procesos estacional mixto integrado ARIMA (p, d, q) x (P, D, Q), proceso de medias móviles exponenciales EWMA, y los procesos de auto agregación, se considera como modelos integrados por qué intervenga la estacionalidad de la serie en estudio. (Uriel Jiménez, 1985)

### Mixtos integrados ARIMA (p,d,q)

$$\varphi(L) = \varphi_p(L)(1 - L)^d(Y_t - \mu) = \theta_q(L)a_t$$

Donde:

$\varphi(L)$ : Es el operador autorregresivo generalizado.

$\varphi_p(L)$ : Operador autorregresivo estacionario de la serie  $Y_t$

$d$ : Numero de diferencias necesarias para alcanzar estacionariedad en la serie.

## ESTACIONALIDAD

De no ser por la estacionalidad, el análisis de las series temporales se convertiría en algo muy simple. Cabe mencionar que la mayoría de los series temporales, ciencias psicológicas y sociales, quedarían bien representadas por los modelos de orden inferior, tales como los modelos ARIMA(1,0,0), ARIMA(0,0,1) ARIMA(0,1,0), ARIMA(1,1,0) Y ARIMA(0,1,1). Dichos modelos cubren gran parte de los procesos temporales fácilmente identificables, pero por desgracia, suele ocurrir que las series temporales



presentan ciclos de carácter mensual, trimestral, anual, etc. De tal manera que se encuentran afectados por fuertes componentes estacionales (Melo & Santana, 2016).

La serie estacionaria se describe por una secuencia de datos o valores que no presentan ningún cambio sistemático en la media, ni cambio en la varianza, esto quiere decir que un proceso es estacionario cuando, en cada uno de los puntos del tiempo, la observación registrada puede ser considerada una variable aleatoria a la que está asociada una función de densidad de probabilidad (Melo & Santana, 2016).

Se puede apreciar que muchas series no son estacionarias; pero si sus primeras y segundas diferencias. El propósito de diferenciar una serie es volver estacionaria al diferencial de dicha serie. Se debe recordar que si toma diferencias también serán estacionarias; luego puede darse una sobre diferenciación de las series.

## MODELO ARIMA ESTACIONAL

La metodología para construir el modelo ARIMA (p; d; q) (P;D;Q)<sub>s</sub> apropiado para la serie estacional cuenta con 4 fases: Identificación, Estimación, Validación y Predicción.

A partir de lo expuesto en el párrafo precedente, es posible expresar el modelo estacional ARIMA, en términos generales, por ARIMA (p, d, q) (P, D, Q)<sub>S</sub>, donde P, D, Q, son análogos a los parámetros p, d, q. el parámetro S simboliza la longitud del periodo o ciclo. De esta forma. Con los datos mensuales S=12. Al modelizar procesos estacionales, solemos incorporar tanto estructuras regulares *como como estacionales de forma aditiva* (Uriel Jiménez, 1985).



## Modelos Autorregresivos (AR)

Definimos un modelo como autorregresivo si la variable endógena de un periodo “t” es explicada por las observaciones de ella misma correspondientes a periodos anteriores añadiéndose, como en los modelos estructurales, un término de error. En este caso de procesos estacionarios con distribución normal, se afirma que, bajo determinadas condiciones previas, toda " $Y_t$ " puede expresarse como una combinación lineal de sus valores pasados (Uriel Jiménez, 1985).

El valor presente de la variable se representa en función de su propio pasado más una innovación contemporánea.

Los modelos autorregresivos se abrevian con la palabra AR tras la que se indica el orden del modelo: AR (1), AR (2), etc. el orden del modelo expresa el número de observaciones retrasadas de las series de tiempo analizadas que intervienen en la ecuación (Uriel Jiménez, 1985).

La expresión genérica de un modelo autorregresivo AR (p) sería de la siguiente forma:

$$Y_t \sim AR(p) \rightarrow Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t$$

(Modelo que solo presente parte autoregresiva, es decir; el polinomio de medias móviles es de orden 0) (Uriel Jiménez, 1985).

Dónde:  $\varepsilon_t$  es una variable “ruido blanco”, siempre y cuando cumpla lo siguiente:

- i. Media Nula

- ii. Varianza Constante
- iii. Covarianza nula entre errores correspondientes a observaciones diferentes.

Los procesos AR (p), también se puede escribir de forma abreviada como:

$$\phi_p(L)Y_t = \delta + \varepsilon_t$$

En donde  $\phi_p(L)$  es lo que se conoce como operador polinomial de retardos:

$$\phi_p(L) = 1 - \phi_1L - \phi_2L^2 - \phi_3L^3 - \dots - \phi_pL^p$$

En donde el término “L”, es el que se conoce como operador retardo tal que, implicado al valor de una variable en “t”, dé como resultado el valor de esa misma variable en “t-1” (Melo & Santana, 2016):

$$LY_t = Y_{t-1}$$

Y aplicando sucesivamente “p” veces retarda, en el valor de “p” periodos.

$$L^pY_t = Y_{t-p}$$

Se suele trabajar con modelos autorregresivos de ordenes bajos: AR (1), AR (2), o bien con órdenes coincidentes con la prioridad de datos de la serie analizada, si es trimestral AR (4), si es mensual AR (12).

Si resulta que  $p=12$  para datos mensuales, el modelo autorregresivo establece un modelo de índices estacionales que son los coeficientes estimados. Como se mencionó previamente, puede eliminarse el propio patrón estacional para investigar si hay otro modelo que abarca varios años, o si el modelo se extiende a un plazo más largo.



Naturalmente, el modelo autorregresivo puede también revelar variaciones cíclicas menores de doce meses. Se debe tener en cuenta que es necesario imponer ciertas restricciones a los valores de los parámetros de este modelo para que funcione correctamente estacionario (Uriel Jiménez, 1985).

### MODELO AR (1)

Según Uriel Jiménez (1985) Un modelo AR (1) viene definido por:

$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t$$

Por lo tanto utilizando el operador de retardos para un modelo AR (1) sería de la siguiente forma:

$$(1 - \phi_1 L)Y_t = \varepsilon_t$$

En donde: “ $\varepsilon_t$ ” es el ruido blanco.

### MODELO AR (2)

Según Uriel Jiménez (1985) un modelo AR (2) viene definido por:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$$

Así mismo aplicando para un modelo AR (2), su operador de retardos sería de la siguiente forma:

$$(1 - \phi_1 L - \phi_2 L^2)Y_t = \varepsilon_t$$

## PROCESOS DE MEDIAS MÓVILES (MA).

Según Melo & Santana (2016) un modelo denominado de medias móviles, es aquel que explica el valor de una determinada variable en un periodo “t”, en función de un término independiente.

Las medias móviles son indicadores que suavizan, en mayor o menor medida, de tal manera que eliminan determinadas fluctuaciones a corto, medio o largo plazo.

El valor presente de la variable se representa en función de todas las innovaciones presente y pasadas.

Estos modelos se denotan normalmente con las siglas MA, seguidos como en el caso de los modelos autorregresivos, del orden entre paréntesis. La notación, MA (q) que se refiere a un modelo de media móvil de orden q. Por lo tanto la expresión genérica de un modelo autorregresivo MA (q) sería de la siguiente forma:

$$Y_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

En donde “ $a_t$ ” es un ruido blanco con las propiedades ya definidas.

Obsérvese que el proceso de medias móviles corresponde a una combinación lineal de variables ruido blanco, siendo los coeficientes “theta” “ponderadores” de la combinación lineal. Esto se asemeja a la definición de un “promedio” de las variables ruido blanco.

### MODELO MA (1)

Según Uriel Jiménez (1985) un modelo MA (1) viene definido por:

$$Y_t = a_t - \theta_1 a_{t-1} = (1 - \theta_1 L) a_t$$

En dónde: " $a_t$ " es un ruido blanco con las propiedades ya definidas.

### **MODELO MA (2)**

Según Uriel Jiménez (1985) un modelo MA (2) viene definido por:

$$Y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} = (1 - \theta_1 L - \theta_2 L^2) a_t$$

En dónde: " $a_t$ " es el ruido blanco.

### **Proceso Autorregresivo de Media Móvil (ARMA)**

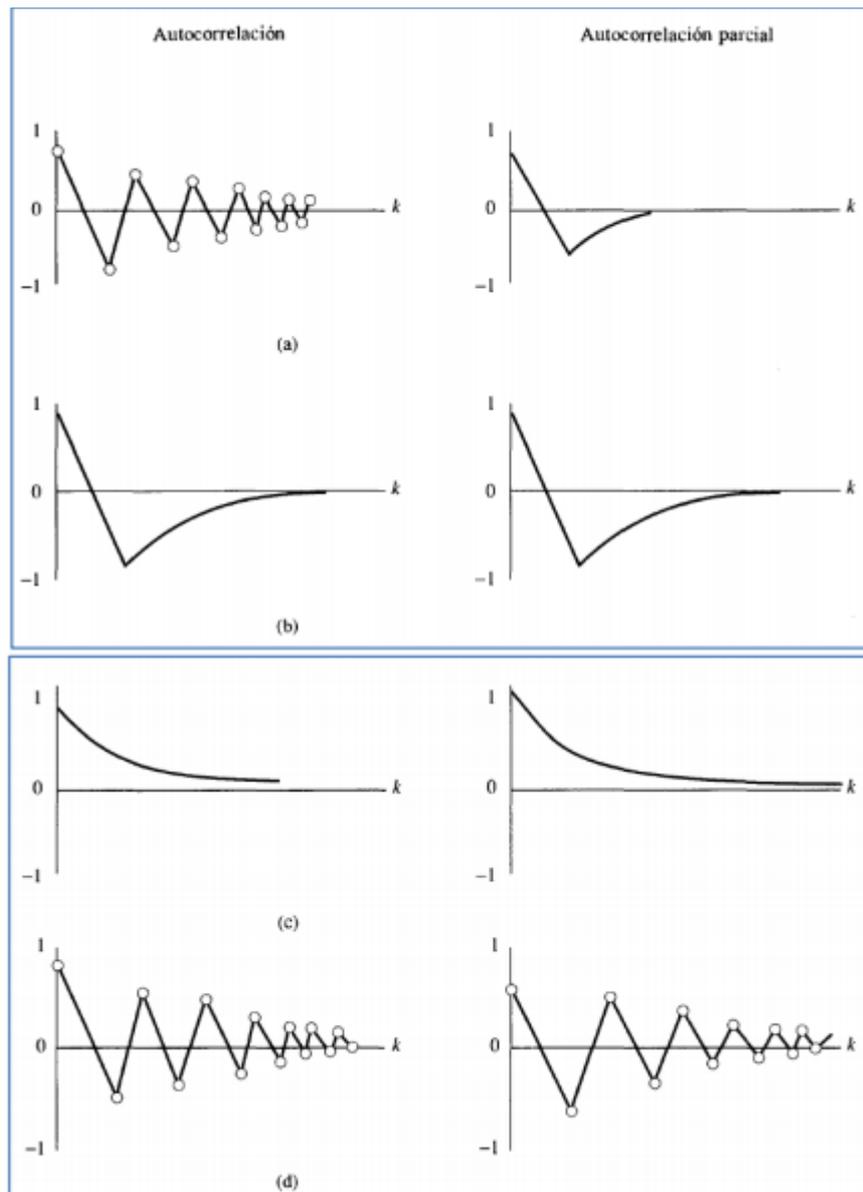
La combinación de procesos AR y MA da lugar a los procesos mixtos ARMA. La notación la notación de ARMA (p, q), se refiere a un modelo con "p" términos autorregresivo y "q" términos de media móvil (Melo & Santana, 2016).

La formulación general de un proceso ARMA, ARMA (p, q), es:

$$Y_t = C + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

Los momentos del proceso se derivan de la misma forma que se vio para los procesos AR y MA.

Distribuciones teóricas de los coeficientes de correlación



**Figura 1** Coeficiente de Autocorrelación y de Autocorrelación Parcial

**Fuente** MELO-MORÍN, Julia y SANTANA-ESPARZA, Gil. Minado de series de tiempo utilizando la metodología ARIMA. Revista de Investigación y Desarrollo 2016, 2-5: 21-31

### Condiciones y Raíces Unitaria para los Modelos

Para que un proceso estocástico estacionario admita una formulación, se deben de cumplirse dos condiciones importantes.



Primero: El proceso no debe ser anticipante (hipótesis de recursividad temporal), lo que quiere decir que los valores de una variable en un momento “ $t$ ”, no dependerán de los esta misma tome en “ $t + k$ ”, siendo “ $k$ ” cualquier valor superior a cero.

Segundo: El proceso ha de ser invertible, lo que supone que la correlación entre una variable y su pasado va reduciéndose a medida que nos alejamos en el tiempo del momento para el que estamos considerando dicha correlación.

### **La Estacionariedad de las Series Temporales en la Realidad**

Son los procesos estocásticos que cumplan al menos de forma débil, la restricción de la estacionariedad. Cuando en la realidad queremos inferir a partir de una serie temporal (muestra), la estructura del proceso estocástico mediante modelos AR o MA, debemos asegurar de que la serie temporal, como muestra del proceso estocástico es estacionaria, caso contrario hay que transformar la serie temporal original de tal forma que la nueva serie transformada si lo sea (Melo & Santana, 2016).

### **PROCESOS ARIMA- NO ESTACIONARIOS**

La mayor parte de las series económicas corresponden a procesos no estacionarios. Así, si se desea obtener un tratamiento de las series basado en el “análisis de series de tiempo” (modelos ARMA), es necesario discutir mecanismos de transformación de las series a procesos estacionarios.

En 1970, en su libro Box y Jenkins desarrollaron un cuerpo metodológico destinado a identificar, estimar y diagnosticar modelos dinámicos de series temporales en los que la variable tiempo tiene un papel muy importante.

En principio pueden presentarse distintas formas por las que se introduce la no estacionariedad en un proceso estocástico. Sin embargo, se debe considerar solo algunas formas de la no estacionariedad que sean adecuados para describir el comportamiento de las series temporales, al mismo tiempo las posibilidades de ser transformados en procesos estacionarios.

### **MODELO ARIMA (p, d, q) SARIMA (P, D, Q)**

Generalmente se expresa como ARIMA (p, d, q) donde los parámetros p, d y q son números enteros no negativos que indican el orden de las distintas componentes del modelo respectivamente, los componentes autorregresivo integrados y de media móvil. Cuando alguno de los tres parámetros es cero, es común omitir las letras correspondientes del acrónimo AR para la componente autorregresivo, “I” para la integrada y “MA” para la media móvil.

La forma general expresada del modelo ARIMA (p, d, q) ARIMA (P, D, Q),s podría escribirse como:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_{P_S+p+D_S+d} Y_{t-P_S-p-S_D-d} + \delta + U_t + \theta_1 U_{t-1} + \dots + \theta_{Q_S+q} U_{t-S_Q-q}$$

Entendiendo que puede haber más de un proceso generador de la serie a su vez escribiendo una combinación de los modelos MA (q) y AR (p) que han precisado de una serie de diferenciaciones "d" en la parte regular o "D" en la parte estacional para que fueran estacionarios.

El modelo ARIMA (p,d,q) indica que consta de un a polinomio Autorregresivo de orden “p”, de una diferenciación en la variable de estudio " $Y_t$ " de orden “d”, y de un polinomio de promedios móviles de orden “q”.

## **FUNCIONES DE UN PROCESO ESTOCÁSTICO ESTACIONARIO**

Un proceso estocástico como estacionario debe de cumplir las condiciones en sentido estricto, también cumple las condiciones en sentido débil. Teniendo así el proceso definido si conociéramos su media constante ( $\mu$ ), una varianza constante ( $\sigma$ ) y la covarianza entre cada par de momentos diferentes en el tiempo.

La función de autocovarianza vendrá definida por los distintos valores que tomaría dicha covarianza cuando se cambie el lapso temporal entre las observaciones de la serie en estudio.

Analíticamente, se podría expresar como:

$$\gamma_k = Cov(y_t, y_{t-k}) = E[(y_t - \mu)(y_{t-k} - \mu)]$$

En donde, evidentemente, cuando el valor de "k" es cero, tendríamos la varianza de la función:

$$\gamma_0 = Cov(y_t, y_{t-0}) = E(y_t - \mu)^2 = \sigma^2$$

La función de autocorrelación se define igualmente como:

$$\rho_1 = \frac{Cov(y_t, y_{t-k})}{\sqrt{var(y_t)}\sqrt{var(y_{t-k})}}$$



Como nos encontramos ante un proceso definido como estacionario, la varianza es constante, por lo que podemos escribir:

$$\rho_k = \frac{\gamma_k}{\gamma_k} \quad \forall k$$

Las propiedades de la covarianza y la definición del coeficiente de correlación, se tiene:

$$\rho_k = \rho_{-k} \quad \forall k$$

$$\rho_0 = \frac{\gamma_0}{\gamma_0} \equiv 1$$

El conjunto de los valores de los coeficientes de autocorrelación se denomina función de autocorrelación (FAC), que es una función en “j”, cuyos valores son las autocorrelaciones definidas.

## **FUNCIÓN DE AUTOCORRELACIÓN**

El cociente de autocorrelación de orden k de un proceso estocástico estacionario que mide el grado de asociación lineal existente entre dos variables aleatorias del proceso separadas k periodos:

El coeficiente de autocorrelación entre la variable “ $y_t$ ” y la misma variable un período antes, “ $y_{t-1}$ ” al que denominaremos coeficiente de autocorrelación de primer orden, se formula como:

$$\rho_1 = \frac{Cov(y_t, y_{t-1})}{\sqrt{var(y_t)}\sqrt{var(y_{t-1})}}$$

Dado el supuesto de estacionariedad, se tiene que  $var(y_t) = var(y_{t-1})$ , por lo que:

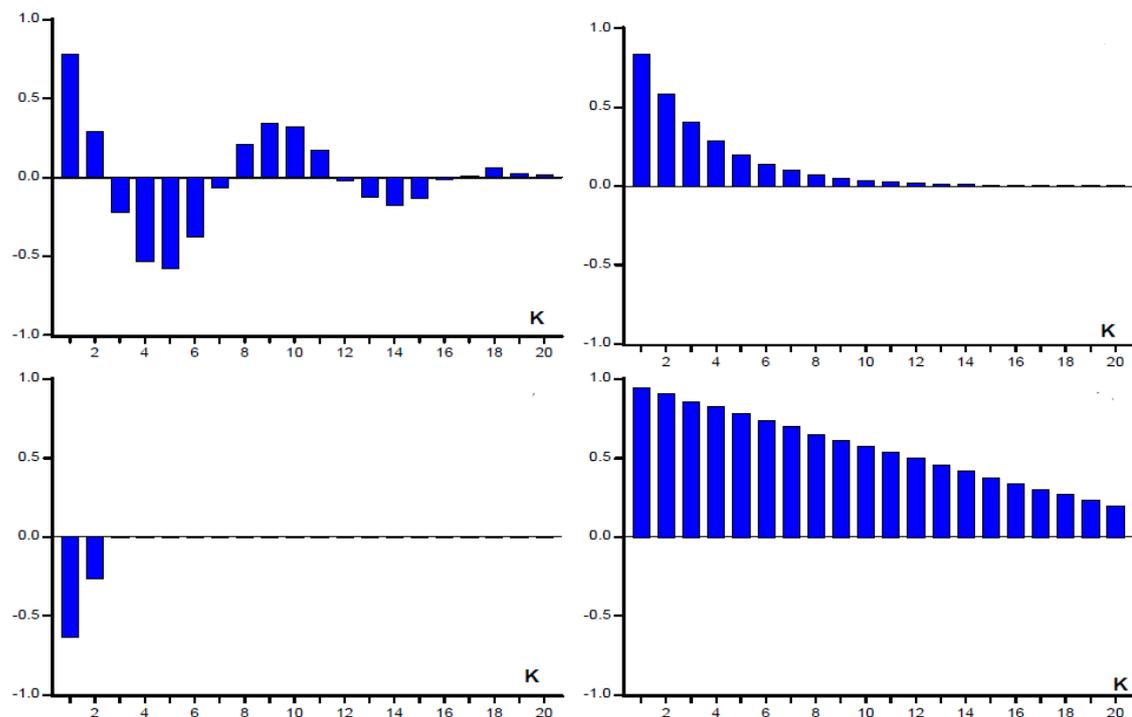
$$\rho_1 = \frac{Cov(y_t, y_{t-1})}{Var(y_t)}$$

En general, para un desfase de k períodos se tiene que:

$$\rho_k = \frac{Cov(y_t, y_{t-k})}{Var(y_t)}$$

Y cuando  $k=0$ ,

$$\rho_0 = \frac{Cov(y_t, y_t)}{Var(y_t)} = \frac{Var(y_t)}{Var(y_t)} = 1$$



**Figura 2** Funciones de Autocorrelación.

**Fuente** Montes Páez, Erik & Calvete, Fernando & Mantilla, Carlos. (2016).

Aplicación de series de tiempo en la realización de pronósticos de producción. Revista Fuentes el Reventón Energético. 14. 79-88. 10.18273/revfue.v14n1-2016007.

## FUNCIÓN DE AUTOCORRELACIÓN PARCIAL

La función de autocorrelación parcial mide la aportación que a las variaciones de una variable como “ $y_t$ ” tiene otra variable, digamos “ $y_{t-2}$ ” aislados los efectos de las posibles restantes variables, por ejemplo “ $y_{t-1}$ ”, Por el contrario, la función de autocorrelación ignora el hecho de que parte de la correlación que pueda existir entre, por ejemplo “ $y_t$ ” y “ $y_{t-2}$ ”, se debe a que ambas están correlacionadas con “ $y_{t-1}$ ”, Pues bien, los distintos coeficientes de autocorrelación parcial de los modelos teóricos se denotan como  $\phi_{kk}$ , y los estimados para nuestra muestra como  $\hat{\phi}_{kk}$

La Esta función entre el rango  $[-1,+1]$ , donde +1 indica una correlación perfecta (la señal se superpone perfectamente tras un desplazamiento temporal de K) y -1 indica una anticorrelación perfecta (De Arce & Mahía, 2009).

Función de autocorrelación parcial, con el fin de tener en cuenta los valores de correlación entre dos variables aleatorias separadas entre sí "k" períodos y en función de los valores intermedios entre ellas (De Arce & Mahía, 2009).

Es decir:

$$\Pi_k = \text{corr}(y_t, y_{t-k}, \dots, y_{t-1}, y_{t-2}, \dots, y_{t-k+1})$$

Si planteamos las mejores predicciones de “ $y_t$ ” y “ $y_{t-k}$ ” como los resultantes de plantear Mínimos Cuadrados Ordinarios (MCO) siendo el primero de ellos del siguiente modo:

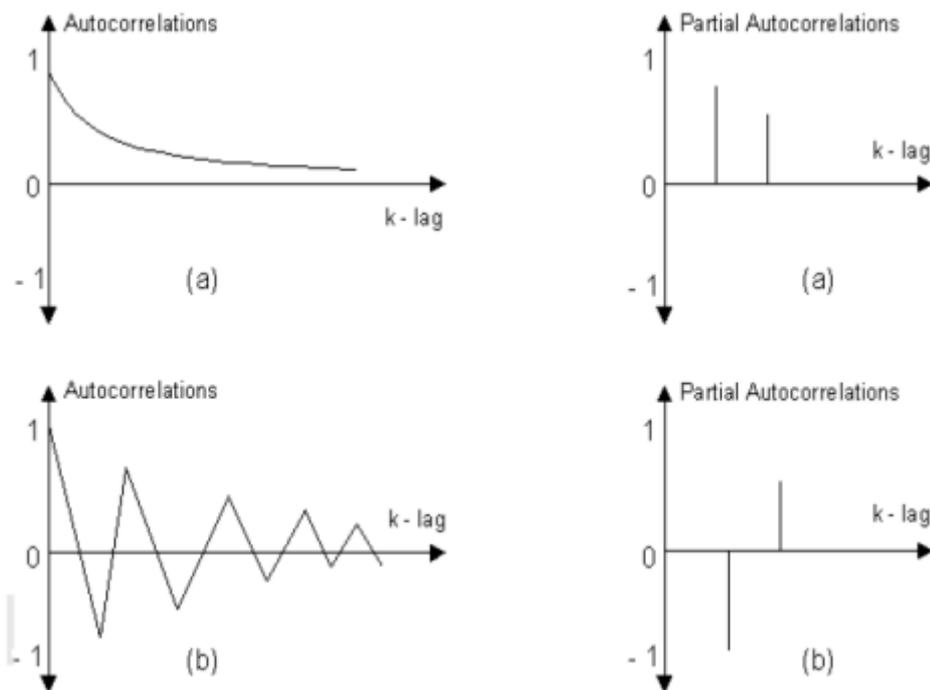
$$\hat{y}_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_{k-1} y_{t-k+1}$$

Se puede escribir la función de autocorrelación parcial, si la media es nula, como:

$$\Pi_k = \frac{Cov[(y_t - \hat{y}_t)(y_{t-k} - \hat{y}_{t-k})]}{\sqrt{var(y_t - \hat{y}_t)}\sqrt{var(y_{t-k} - \hat{y}_{t-k})}}$$

Pudiendo demostrarse que:

$$\Pi_k = \frac{\rho_k - \alpha_1\rho_{k-1} - \alpha_2\rho_{k-2} - \dots - \alpha_{k-1}\rho_1}{1 - \alpha_1\rho_1 - \alpha_2\rho_2 - \dots - \alpha_{k-1}\rho_{k-1}}$$



**Figura 3** Función de Autocorrelación y de Autocorrelación Parcial

**Fuente** MELO-MORÍN, Julia y SANTANA-ESPARZA, Gil. Minado de series de tiempo utilizando la metodología ARIMA. Revista de Investigación y Desarrollo 2016, 2-5: 21-31



**Tabla 1** Comportamientos de las FAC y FACP

| VARIABLES | FAC  | FAP  |
|-----------|--|--|
| MA (q)    | Se anulan para retardos superiores a q     | Decrecimiento rápido sin llegar a anularse |
| AR(p)     | Decrecimiento rápido sin llegar a anularse | Se anulan para retardos superiores a p     |
| ARMA(p;q) | Decrecimiento rápido sin llegar a anularse | Decrecimiento rápido sin llegar a anularse |

## HETEROCEDASTICIDAD

La heterocedasticidad es la existencia de una varianza no constante en el tiempo en las perturbaciones aleatorias en un modelo estadístico. Cuya situación se da en modelos de corte temporal en los que la evolución histórica haya manifestado diferentes períodos en cuanto a los valores de una variable en relación a su media, agrupando en algún subperíodo valores altos y en otros valores pequeños (De Arce & Mahía, 2009).

## PROCESOS ARIMA ESTACIONALES

Muchas series de tiempo económicas que se basan en datos mensuales o trimestrales presentan pautas estacionales (movimiento oscilatorio regular); por ejemplo, las ventas de las tiendas de departamentos en la época de Navidad y otras festividades importantes, la demanda de dinero (saldos de efectivo) por parte de las familias en épocas de vacaciones, la demanda de helado y bebidas gaseosas durante el verano y los precios de los cultivos justo después de la época de cosecha, la demanda de viajes en avión, etc.

A menudo es útil eliminar el factor o componente estacional de las series de tiempo con el fin de concentrarse en los demás componentes, como la tendencia. (Gujarati & Porter, 2010, pág. 290)

## **TRANSFORMACIÓN DE BOX-COX**

Box-Cox definió una transformación instantánea en el sentido de que no está involucrado simultáneamente varios periodos de tiempo de carácter más general, para estabilizar la varianza se utilizan las transformaciones Box-Cox, la cual se define por:

$$Y_t^\lambda = \begin{cases} (Y_t^\lambda - 1)/\lambda & , \lambda \neq 0 \\ \ln y_t & , \lambda = 0 \end{cases}$$

Donde “ $\lambda$ ” es el parámetro de transformación. Es interesante señalar que, usualmente, las transformaciones Box-Cox no solo estabilizan la varianza sino que también mejoran la aproximación a la distribución normal del proceso  $y_t$ . (González Casimiro, 2009, pág. 44)

## **PROCEDIMIENTOS DE DETECCIÓN DE RAÍCES UNITARIAS - CONTRASTE DE DICKEY-FULLER**

Los contrastes de raíces unitarias proporcionan unos contrastes estadísticos que permiten, a partir del conjunto de información, hacer inferencia sobre la existencia o no de una raíz unitaria en una serie, es decir, sobre la no estacionariedad de la serie. Si se rechaza la hipótesis nula de existencia de raíz unitaria en  $\Delta^{d-1}y_t$ , no se diferenciaría más la serie. En caso contrario, si no se rechaza la hipótesis nula se tomara una diferencia más de orden 1.



A continuación, se desarrollan brevemente, los contrastes de raíces unitarias más utilizados en la práctica del análisis de series temporales y que, a la vez, fueron los pioneros en este campo.

El planteamiento más sencillo de Dickey - Fuller es el siguiente. Plantean un modelo autorregresivo AR (1) como:

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

En donde “ $\mu$ ” y “ $\rho$ ” son parámetros a estimar y “ $\varepsilon_t$ ” es un término de error para el que se asume que cumple las propiedades de ruido blanco. Dado que “ $\rho$ ” es un coeficiente de autocorrelación que, por tanto, toma valores entre  $-1 < \rho < 1$ , si  $\rho = 1$ , la serie  $y$  es no estacionaria. Si el valor de “ $\rho$ ” es mayor que la unidad, entonces se dice que la serie es explosiva. De esta forma, la hipótesis de estacionariedad de una serie puede ser evaluada analizando si el valor absoluto de “ $\rho$ ” es estrictamente menor que 1. Pues bien, el test Dickey - Fuller plantea, sencillamente, contrastar estadísticamente si  $\rho = 1$ .

Puesto que en economía las series explosivas no tienen mucho sentido, esta hipótesis nula se analiza frente a la alternativa que establece que:

$$H_1: \rho < 1$$

Se puede demostrar que una expresión equivalente a la anterior, mediante la toma de diferencias, viene dada por:

$$\Delta y_t = \mu + \gamma y_{t-1} + \varepsilon_t$$

En donde  $\gamma = \rho - 1$  y las hipótesis nula y alternativa son:

$$H_0: \gamma = 0$$

$$H_1: \gamma < 0$$

Aunque el test calcula un estadístico “t” sobre el valor estimado de “ $\gamma$ ”, bajo la hipótesis nula de existencia de una raíz unitaria, este estadístico no tiene la distribución convencional de la t-Student. Tomando en cuenta lo mencionado, Dickey - Fuller simularon los valores críticos para una selección de distintos tamaños muestrales, simulación ampliada posteriormente por MacKinnon.

Por lo tanto, el primer paso para la detección de raíces unitarias debe ser especificar un test Dickey - Fuller simple, y después contrastar estadísticamente las características de ruido blanco de los residuos (estadístico Q de Ljung-Box).

Para estos casos disponemos del test ampliado de Dickey-Fuller (ADF), el cual permite considerar otros esquemas de autocorrelación. Para ello, el test ADF añade términos diferenciados de la variable dependiente y en el lado derecho de la regresión:

$$\Delta y_t = \mu + \gamma y_{t-1} + \delta_1 \Delta y_{t-q} + \delta_2 \Delta y_{t-2} + \dots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t$$

Esta especificación aumentada del test se basa en las siguientes hipótesis nula y alternativa:

$$H_0: \gamma = 0$$

$$H_1: \gamma < 0$$

Un resultado importante obtenido por Dickey - Fuller es que la distribución asintótica del estadístico t sobre  $\gamma$  es independiente del número de retardos de la serie en primeras diferencias incluidos en la regresión del test ADF.

## PREDICCIÓN DE UNA SERIE DE DIFERENCIA

El objetivo es obtener predicciones óptimas de  $Y_t$  en algún momento futuro basadas en un conjunto de información dado que en el caso del análisis de series temporales univariante está formado por el pasado disponible de la serie temporal.

Entonces, es claro que:

$$E_t Y_{t+k} = E_t Y_{t-k} - E_t Y_{t+k+1}$$

Por lo que:

$$\begin{aligned} E_t Y_{t+k} &= E_t Y_{t-k} - E_t Y_{t+k+1} \\ &= Y_{t+k} + Y_{t+k+1} + Y_{t+k+2} + \dots + Y_{t+k+l} + e_t \end{aligned}$$

## ERROR DE PREDICCIÓN

El error de predicción es la diferencia entre un valor  $Y$  y su estimación  $Y'$  en una recta de regresión. El error cometido en la predicción de " $Y_{t+k}$ " depende del periodo en que dicha predicción se realiza.

### 2.3 MARCO CONCEPTUAL

#### 2.3.1. Muestra

Parte o cantidad pequeña de una cosa que se considera representativa del total y que se toma o se separa de ella con ciertos métodos para someterla a estudio, análisis o experimentación.



### **2.3.2. Estacionalidad**

Puede definirse como la repetición de un cierto patrón de comportamiento en forma periódica.

### **2.3.3. Estacionariedad**

Es el comportamiento de la variable en el tiempo.

### **2.3.4. Población**

Se define población como la colección completa de elementos acerca de los cuales se desea conocer la información.

### **2.3.5. Unidad Muestral**

Cada objeto sobre las que se realizará el estudio.

### **2.3.6. Serie**

Conjunto de observaciones ordenadas en el tiempo (o en alguna otra dimensión).

### **2.3.7. Ruido Blanco**

Es un proceso puramente aleatorio en donde las variables son distribuidas con media cero, varianza constante y ausencia de autocorrelación entre observaciones.

### **2.3.8. Pronostico**

Son las predicciones de los hechos y condiciones futuras.



### **2.3.9. Variable Dependiente**

Variables que influyen en el conjunto de relaciones y a su vez están influenciados por las variables independientes.

### **2.3.10. Variable Independiente**

Variables que influyen en el conjunto de relaciones pero no están influenciados por ella.

### **2.3.11. Modelo**

Es la representación matemática de las variables de estudio y los parámetros que son estimados, con fines de predicción del comportamiento futuro de las variables.

### **2.3.12. Modelo Box-Jenkins**

Es uno de los métodos predicativos y se fundamenta en la estimación eficiente de los parámetros por medio de los procesos iterativos.

### **2.3.13. Modelo Univariante de Box-Jenkins**

Es una serie de tiempo, basado en la información existente el pasado.

### **2.3.14. Modelo Univariante de Box-Jenkins Integrado**

A los procesos mixtos integrados ARIMA (p, d, q), proceso estacional mixto integrado SARIMA (p, d, q)\*(P, D, Q), proceso de medias móviles exponenciales porque interviene la estacionalidad de la serie en estudio.



### **2.3.15. Modelo Univariante De Box-Jenkins No Integrado**

Son los procesos de Medias Móviles MA (q), Autorregresivos AR (p) y Procesos Mixtos ARMA (p, q).



## CAPÍTULO III

### MATERIALES Y MÉTODOS

#### 3.1. UBICACIÓN GEOGRÁFICA DEL ESTUDIO

- Distrito: Puno
- Provincia: Puno
- Departamento: Puno

#### 3.2. MÉTODO DE ANÁLISIS DE DATOS

Para la presente investigación se ha hecho uso de la metodología de Box-Jenkins, conocido comúnmente como el enfoque de Box-Jenkins en las series de tiempo en los denominados modelos ARIMA.

Para el modelo de ajuste, se aplicó la metodología Box-Jenkins (teoría de WIENER-KOLMOGOROV) sigue un proceso que consta de cuatro fases:

- Identificación del modelo:

Para determinar si una serie es estacionaria se debe analizar primeramente el gráfico de la serie en un ploteo de dispersión. Si la serie no es estacionaria se puede convertir a una serie estacionaria mediante el método de diferenciación, para lo cual se debe especificar el grado de diferenciación y el algoritmo de BOX- JENKINS convierten los datos en una serie estacionaria y realiza los cálculos subsecuentes utilizando los datos ya convertidos.



- Estimación y prueba de adecuación

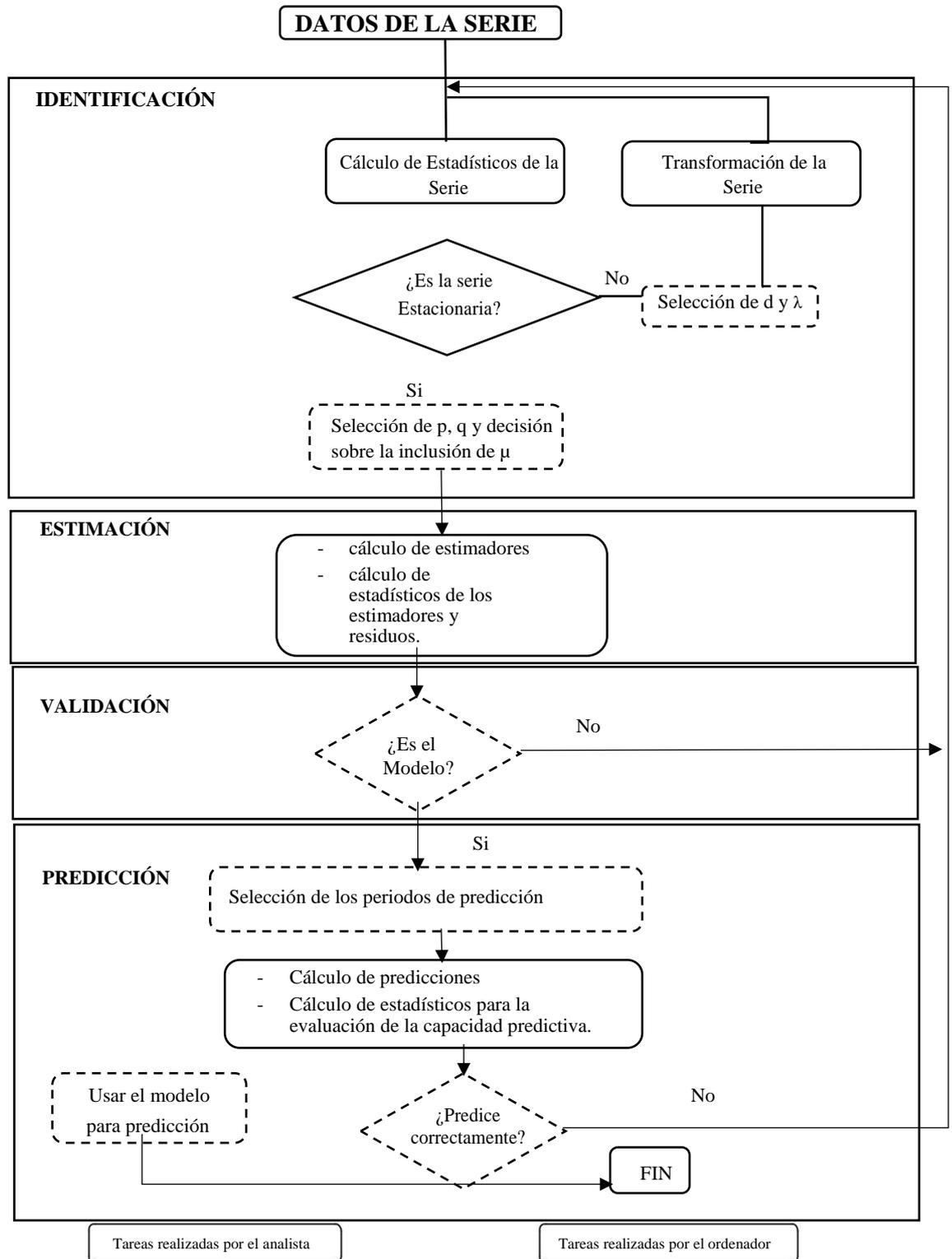
Habiendo identificado los valores apropiados de  $p$ ,  $d$  y  $q$ , la siguiente etapa es estimar los parámetros de los términos autoregresivos y de media móvil incluidos en el modelo. Algunas veces, este cálculo puede hacerse mediante mínimos cuadrados simples, pero otras se tendrá que recurrir a métodos de estimación no lineal (en parámetros). Puesto que esta labor es llevada a cabo a través de rutinas en diversos paquetes estadísticos, en la práctica no es preciso preocuparse por los desarrollos matemáticos de la estimación

- Verificación de diagnóstico

Después de seleccionar un modelo ARIMA particular y de estimar sus parámetros, se trata de ver si el modelo seleccionado ajusta a los datos en forma razonable buena, ya que es posible que exista otro modelo ARIMA que también lo haga, es por eso que el diseño ARIMA de Box Jenkins es un arte más que una ciencia; se requiere gran habilidad para seleccionar el modelo ARIMA correcto.

- Predicción

Una de las razones de la popularidad del proceso de modelación ARIMA en su éxito en la predicción. En muchos casos las predicciones obtenidas por este método son más confiables que aquellas obtenidas de la elaboración tradicional de modelos particularmente para predicciones de corto plazo y teniendo el modelo definido se procede a realizar los pronósticos a futuro haciendo uso del modelo obtenido en el procedimiento.



**Figura 4** Procedimiento del método

**Fuente** Y. Lu, S.M. AbouRizk, Automated Box–Jenkins forecasting modelling,

Automation in Construction, Volume 18, Issue 5, 2009, Pages 547-558,



### **3.3. POBLACIÓN Y MUESTRA**

#### **3.3.1. Población**

La población para efectos de la investigación estará constituida por el total de ingresos registrados mes a mes en la base de datos de la Administración Tributaria, con respecto a los tributos internos en la Región de Puno, que se encuentren registrados y publicados por SUNAT.

El tipo de muestreo es no probabilístico, por conveniencia y disponibilidad de la información.

#### **3.3.2. Muestra**

La muestra está constituida por los registros mensuales en el periodo de enero del 2004 a diciembre del 2017 respecto a los tributos internos en la Región de Puno.

#### **3.3.3. Unidad Muestra**

La unidad muestral estará conformada por el ingreso mensual recaudado por la SUNAT representada en miles de soles (S/.)

### **3.4. MÉTODO DE RECOLECCIÓN DE DATOS**

Con la finalidad de asegurar la efectividad y veracidad de los datos, estos fueron recopilados directamente del registro de Estadística por Regiones de la superintendencia nacional de Administración tributaria correspondiente entre el periodo de 2004 al 2017, la cual se encuentra en su página web.



### 3.5. OPERACIONALIZACIÓN DE VARIABLES

**Tabla 2** Operacionalización de variables

| VARIABLES   | INDICADORES   | UNIDAD DE MEDIDA   |
|---|---|--|
| Variable Dependiente:<br>Ingreso Mensual en Recaudación Tributaria          | <ul style="list-style-type: none"><li>• Ingreso corriente de la recaudación tributaria interna</li></ul>                        | <ul style="list-style-type: none"><li>• (Miles de soles)</li></ul><br>(S/) |
| Variable Independiente:<br>Tiempo que transcurre de una recaudación a otra. | <ul style="list-style-type: none"><li>• Ingreso corriente de la recaudación tributaria interna rezagada en el tiempo.</li></ul> | <ul style="list-style-type: none"><li>• (Miles de soles)</li></ul><br>(S/) |

**FUENTE:** Elaboración propia

## CAPITULO IV

### RESULTADOS Y DISCUSIÓN

A continuación, se presenta la serie mensual de la recaudación tributaria, con respecto a los tributos internos de la Región de Puno, correspondiente a los periodos enero de 2004 a diciembre de 2017, los cuales fueron obtenidos de la página web de la SUNAT.

**Tabla 3** Serie histórica mensual en miles de soles de la Recaudación Tributaria, con respecto a Tributos Internos de la Región de Puno, periodo 2004 – 2017.

| AÑO  | ENE.     | FEB.     | MAR.     | ABR.     | MAY.     | JUN.     | JUL.     | AGO.     | SEP.     | OCT.     | NOV.     | DIC.     |
|------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 2004 | 5 197.3  | 4 244.7  | 4 281.0  | 3 748.7  | 4 826.4  | 4 704.0  | 4 548.2  | 3 191.4  | 5 510.7  | 3 402.4  | 4 320.9  | 3 730.2  |
| 2005 | 5 105.0  | 4 159.6  | 3 659.7  | 4 789.1  | 4 469.9  | 4 608.7  | 4 911.1  | 4 566.8  | 5 733.8  | 5 319.4  | 5 340.0  | 6 077.0  |
| 2006 | 5 798.0  | 5 640.2  | 5 160.7  | 6 386.3  | 6 206.3  | 5 732.7  | 6 709.8  | 6 293.5  | 6 582.8  | 5 786.9  | 5 796.6  | 5 649.3  |
| 2007 | 6 053.7  | 4 702.9  | 5 561.1  | 5 926.8  | 5 646.9  | 5 026.0  | 5 576.7  | 7 113.2  | 6 299.1  | 6 700.4  | 6 764.8  | 7 627.6  |
| 2008 | 8 022.5  | 7 314.6  | 7 022.7  | 8 005.7  | 6 239.5  | 6 827.6  | 6 583.3  | 7 242.7  | 6 917.7  | 7 531.6  | 7 851.1  | 7 795.6  |
| 2009 | 9 361.0  | 9 291.4  | 10 495.4 | 20 884.6 | 13 797.3 | 10 336.5 | 10 645.5 | 10 772.5 | 10 403.1 | 13 408.6 | 10 890.2 | 10 945.3 |
| 2010 | 12 794.6 | 10 797.6 | 10 359.4 | 22 845.7 | 13 469.9 | 12 716.0 | 14 318.1 | 12 603.2 | 12 277.4 | 11 840.4 | 13 661.0 | 13 303.3 |
| 2011 | 17 495.2 | 12 459.3 | 15 957.0 | 17 878.7 | 13 612.0 | 13 312.9 | 13 900.6 | 12 893.1 | 16 592.0 | 15 589.0 | 16 873.1 | 17 052.3 |
| 2012 | 22 015.3 | 17 670.1 | 18 862.3 | 18 345.4 | 16 139.1 | 14 978.7 | 20 364.3 | 18 411.0 | 19 554.6 | 19 059.0 | 20 563.8 | 22 024.0 |
| 2013 | 27 631.8 | 21 858.3 | 23 073.7 | 32 360.5 | 20 218.0 | 19 839.0 | 20 758.0 | 20 106.2 | 20 970.5 | 19 622.8 | 20 969.9 | 24 695.1 |
| 2014 | 28 439.0 | 22 716.0 | 21 827.9 | 30 442.3 | 29 357.4 | 21 353.0 | 22 128.2 | 23 890.3 | 25 681.7 | 26 954.9 | 25 961.8 | 27 507.2 |
| 2015 | 30 204.5 | 24 188.3 | 26 651.1 | 24 579.0 | 21 691.5 | 21 447.2 | 22 000.4 | 25 376.8 | 23 200.3 | 24 560.7 | 28 007.3 | 28 753.1 |
| 2016 | 27 270.7 | 27 513.3 | 30 094.7 | 32 442.7 | 26 408.3 | 25 079.0 | 26 108.9 | 30 105.5 | 28 388.7 | 24 330.7 | 23 690.0 | 24 557.3 |
| 2017 | 28 040.6 | 27 088.3 | 30 932.7 | 24 537.8 | 31 680.1 | 22 338.9 | 21 163.3 | 22 512.8 | 23 137.5 | 25 220.4 | 23 096.5 | 23 253.8 |

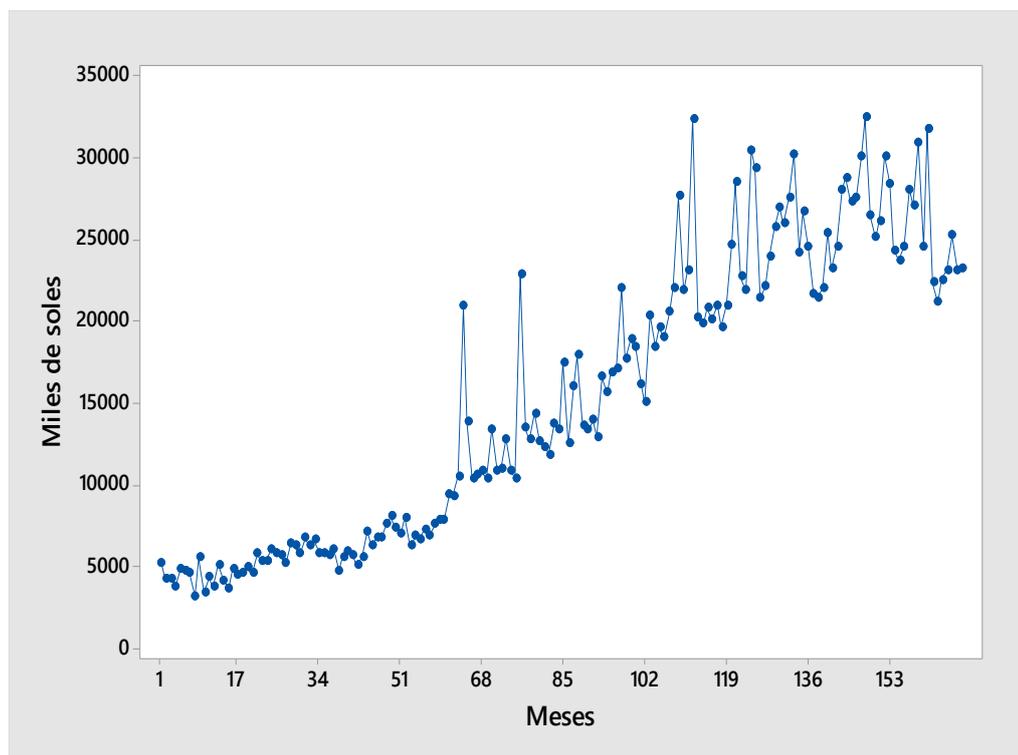
**Fuente:** Superintendencia Nacional de Aduanas y de Administración Tributaria

En la tabla 3 de datos de recaudación tributaria, se puede observar que el mes con mayor ingreso tributario es el mes de abril del 2016 con una recaudación de tributos internos en la Región de Puno de S/.32442.68303 miles de soles y el ingreso tributario con menor ingreso es el mes de agosto del 2004 con una recaudación de tributos internos en la Región de Puno de S/. 3191.43969 miles de soles. Esto es resultado de un conjunto

de elementos que hacen reflejar la recaudación de tributos por la SUNAT en la Región de Puno.

#### 4.1. IDENTIFICACIÓN DEL MODELO

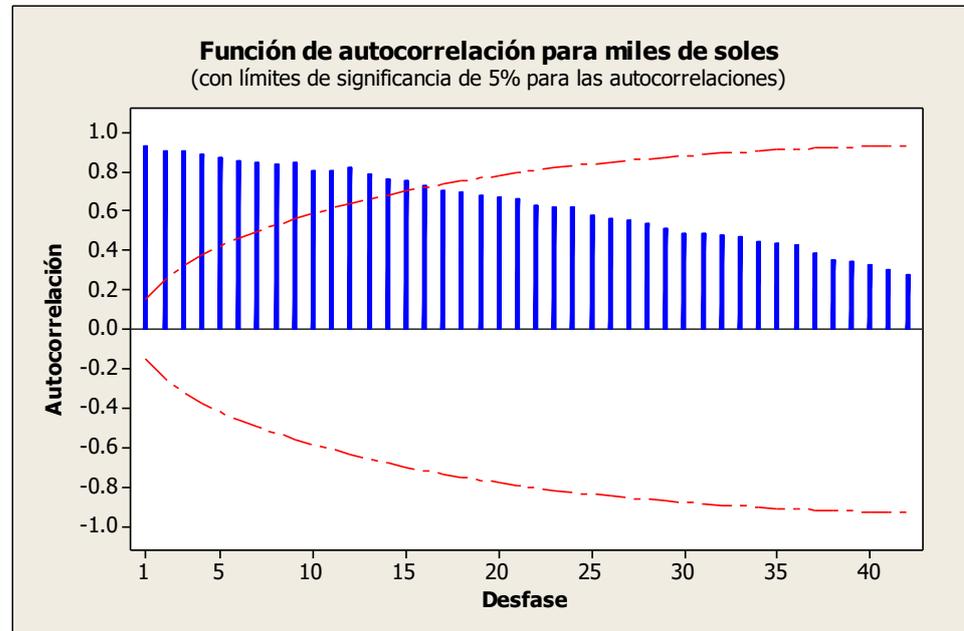
Como primer paso se requiere que la serie de interés sea estacionaria. Una serie estacionaria es la que posee una media y una varianza constante.



**Figura 5** Serie original de recaudaciones tributarias mensuales, con respecto a tributos internos de los años 2004 – 2017

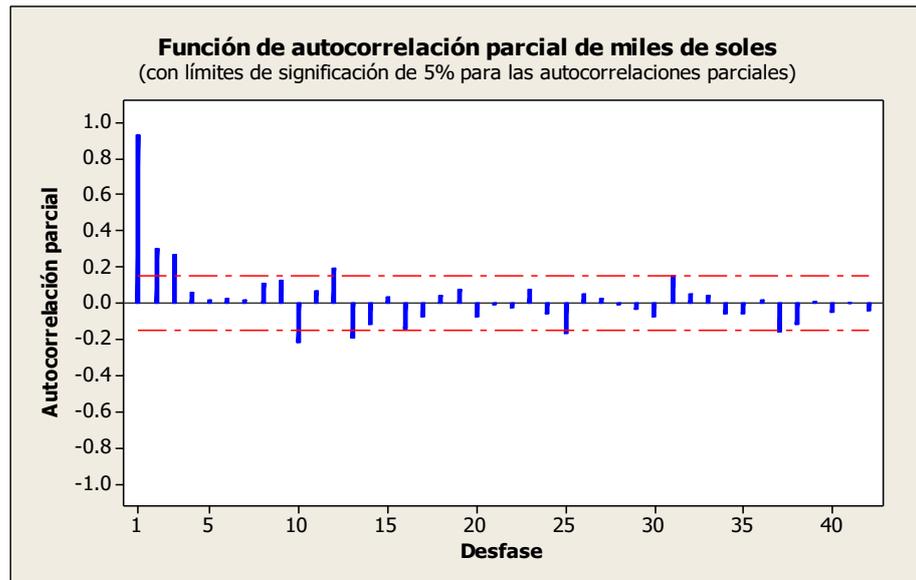
La Figura 5, se puede apreciar que la serie en estudio cuenta con fluctuaciones a lo largo de los años, los cuales tienen una tendencia creciente aditiva ya que presenta un crecimiento aritmético, también se puede apreciar que existen picos los cuales indican que existe variabilidad en la serie de tiempo.

Resultados que concuerdan con Grández (2014) quien afirmó que durante el periodo 1990 y 2012; es decir que ante un incremento de la presión tributaria la recaudación real por persona también incrementó durante este periodo.



**Figura 6** Función de Autocorrelación de la recaudación de tributos internos de la región de Puno

En la Figura 6, se observa en el correlograma de la función de autocorrelación estimada para la serie analizada, cuenta con 16 coeficientes significativos, los cuales tienden a 0 de una manera lenta. Por lo tanto, este grafico corrobora con el grafico anterior afirmando que la serie no es estacionaria.

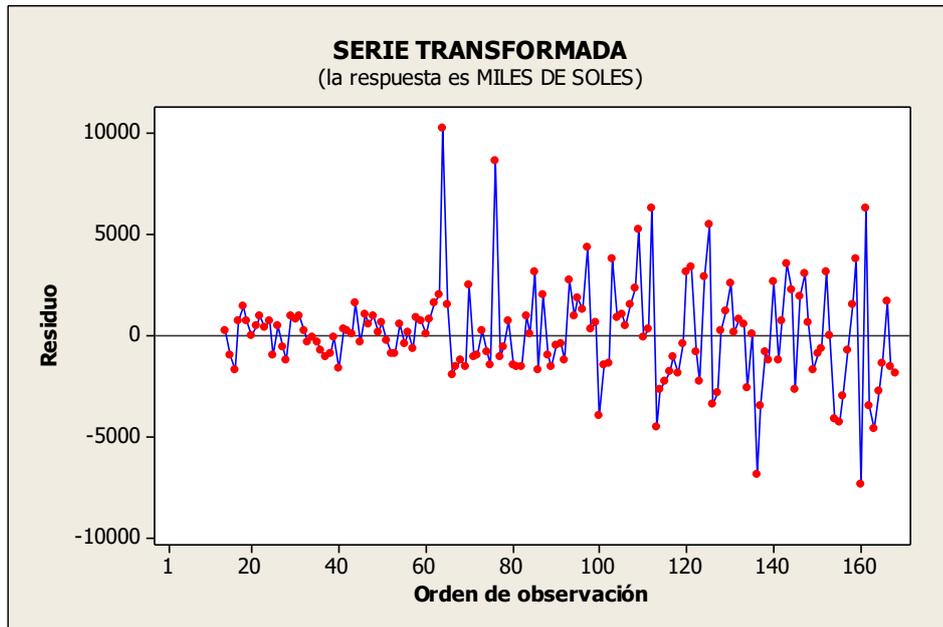


**Figura 7** Función de Autocorrelación parciales de recaudación de tributos internos en la Región de Puno

En la Figura 7, se observa que el correlograma de la serie histórica, donde se advierte el comportamiento de los coeficientes de autocorrelación parcial. Se tiene que los coeficientes en retardo 1, 2, 3, 10, 12, 13, 25 y 37 de los 42, presenta un coeficiente significativo, con un nivel de confianza del 95%, debido a que esta fuera de los límites, adicionalmente se observa que la serie en estudio no es estacionaria.

El análisis de la estacionalidad de las 3 figuras precedentes, en base a una primera inspección se observa un comportamiento no estacional de la serie historia en estudio, dado que la serie cuenta con tendencia y no se desplaza alrededor de una media y una varianza constante a lo largo del tiempo.

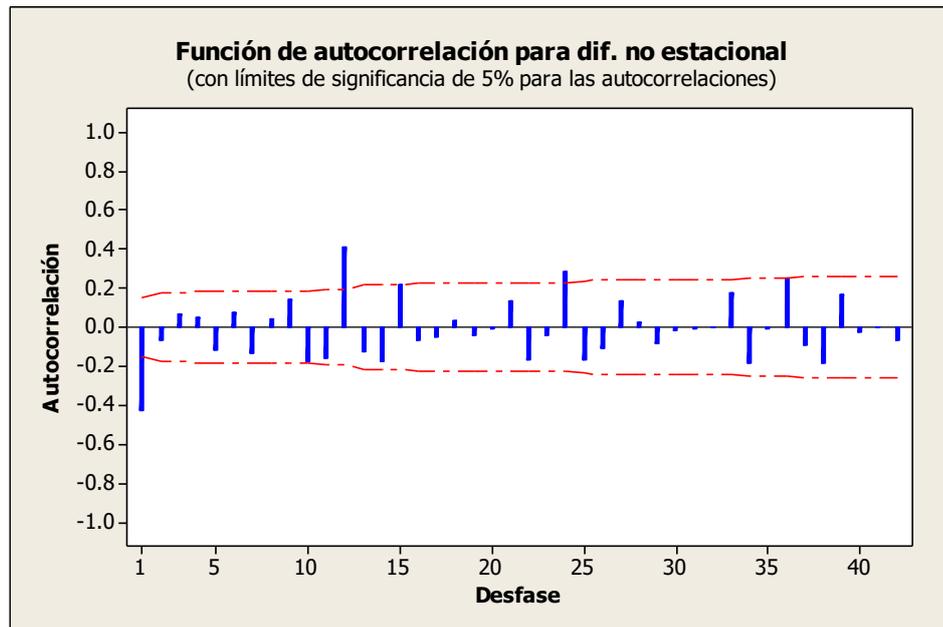
## 4.2. ANÁLISIS DE ESTACIONARIEDAD DE LA SERIE



**Figura 8** Serie transformada (Transf. Box - Cox) de la recaudación de tributos internos en la Región de Puno.

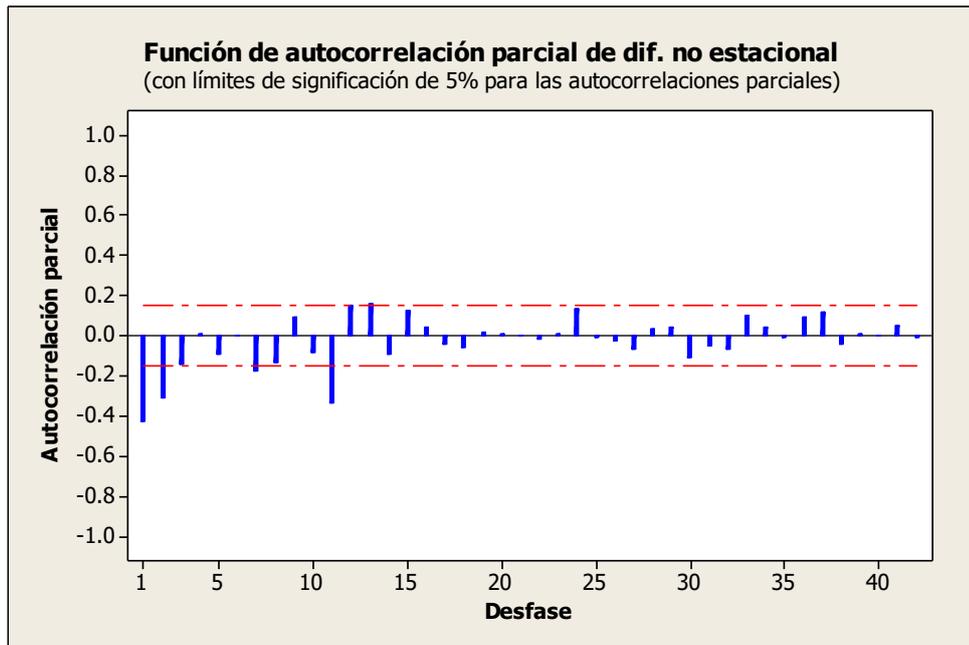
En la Figura 8, se observa que los datos con transformación de Box-Cox con el fin de graficar la serie sin tendencia y observar solamente el patrón de la estacionariedad, se presentan oscilaciones con respecto a la serie en estudio, esto se presume que la serie no es estacionaria.

### 4.3.DIFERENCIA NO ESTACIONAL



**Figura 9** Función de autocorrelación de la primera diferencia no estacional de la serie transformada para la recaudación de tributos internos en la Región de Puno.

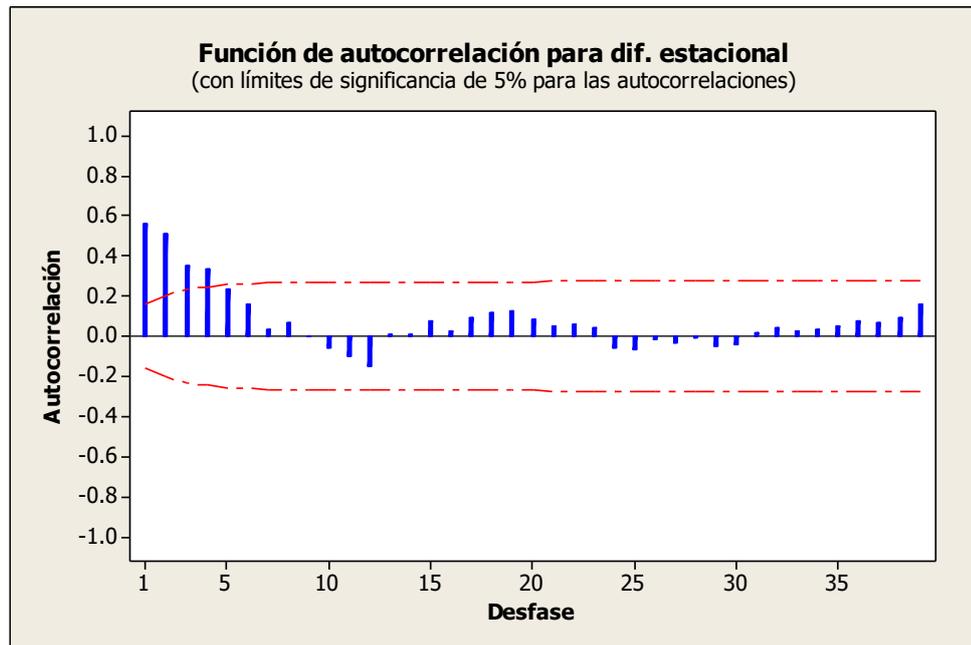
En la Figura 9, observamos el correlograma de la función de autocorrelación estimada para la serie analizada, que existe 3 de los 42 coeficientes de autocorrelación en los retardos 1, 12 y 24 son precisamente grandes, que implica que son estadísticamente significativos a un nivel de confianza del 95% pues sobresalen de los límites, implica que la serie pueden no ser aleatoria.



**Figura 10** Función de autocorrelación parcial de la primera diferencia no estacional de la serie transformada para la recaudación de tributos internos en la Región de Puno.

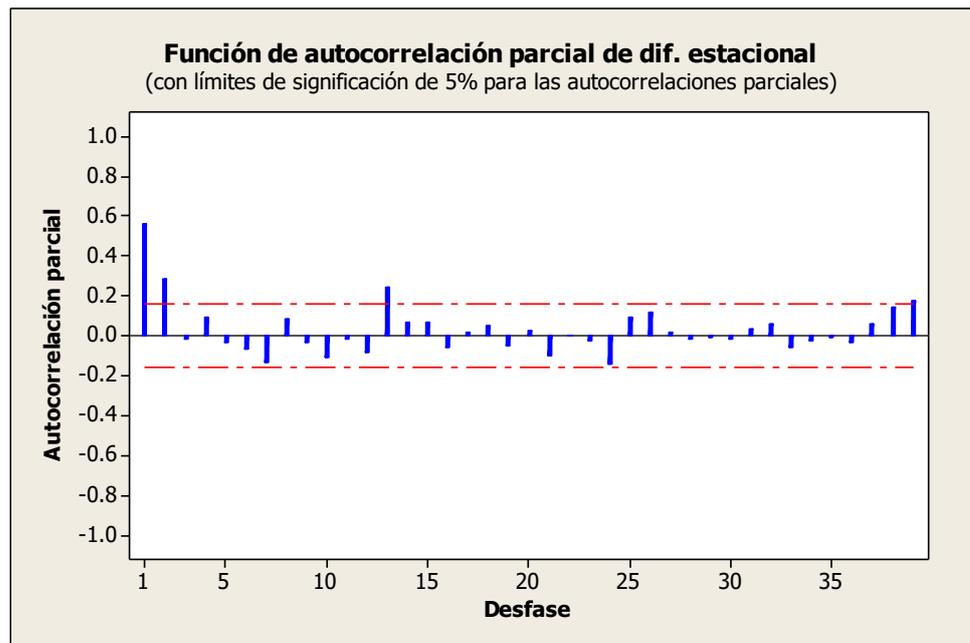
En la Figura 10, observamos el correlograma de la serie histórica, donde advierte el comportamiento de los coeficientes de autocorrelación, se tiene 5 de los 42 coeficientes en los retardos 1, 2, 7, 11 y 13 son estadísticamente significativos a un nivel de confianza del 95%, debido a que esta fuera de los límites; además se observa que la serie muestra un decremento rápido sin llegar a cero, lo que corrobora existe estacionalidad y un comportamiento estacional a lo largo del tiempo.

#### 4.4.DIFERENCIA ESTACIONAL



**Figura 11** Función de autocorrelación parcial de la primera diferencia estacional de la serie transformada para la recaudación de tributos internos en la Región de Puno.

En la Figura 11, se observa que el correlograma de la serie histórica, donde se advierte el comportamiento de los coeficientes de autocorrelación parcial. Se tiene que de los 42 coeficientes, los retardos 1, 2, 3 y 4 son estadísticamente significativos a un nivel de confianza del 95%, debido a que están fuera de los límites; además se observa que la serie muestra un decremento rápido sin llegar a cero, lo que corrobora la existencia de estacionariedad y un comportamiento estacional a lo largo del tiempo



**Figura 12** Función de autocorrelación parcial de la primera diferencia no estacional de la serie transformada para la recaudación de tributos internos en la Región de Puno.

En la Figura 12, se observa el correlograma de la serie histórica, donde advierte el comportamiento de los coeficientes de autocorrelación parcial. Se tiene que 4 de los 42 coeficientes en los retardos 1, 2, 13 y 39 son significativos a un nivel de confianza del 95% debido a que esta fuera de los límites; además se observa un decremento sin llegar a cero, lo que corrobora que existe estacionariedad y un comportamiento estacional a lo largo del tiempo.

Siguiendo la metodología de Box-Jenkins, se presenta los resultados de la serie de la estimación, comprobación del diagnóstico y la predicción, con el software estadístico minitab 16.

La elección del mejor modelo de predicción presenta el menor valor.

#### Modelo Identificado

Se realizó la transformación de la serie, a la cual se aplicó la transformación de Box Cox a la serie en estudio, luego se aplicó la primera diferenciación a la serie transformada

manifestando un mejor acercamiento de estacionalidad durante toda la fase. Una vez realizada la estacionalidad se aplicó de la metodología Box – Jenkins, la cual dio como resultado el modelo identificado es un ARIMA (0,1,1) (0,1,1)12, cuya ecuación es la siguiente:

$$Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} + \theta_1 \varepsilon_{t-1} - \theta_{12} \varepsilon_{t-12} + \theta_{13} \varepsilon_{t-13}$$

$$Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} + 0.6911\varepsilon_{t-1} - 0.8537\varepsilon_{t-12} + 0.5899\varepsilon_{t-13}$$

#### **Estimación del modelo identificado ARIMA (0, 1, 1) (0, 1, 1)**

Se realiza la estimación de los parámetros de los modelos seleccionados. Procederemos a estimar los parámetros del modelo como ya se sabe, en los modelos ARIMA se estiman al minimizar la suma de los cuadrados en los errores de ajuste. Puesto que esta labor es llevada a cabo a través de rutinas en diversos paquetes estadísticos.

**Tabla 4** Resumen de los parámetros del modelo ARIMA (0, 1,1) (0,1,1)

| Parameter      | Estimate | Std. Error | T     | P-value |
|----------------|----------|------------|-------|---------|
| <b>MA(1)</b>   | 0.6911   | 0.0586     | 11.78 | 0.000   |
| <b>SMA(12)</b> | 0.8537   | 0.0544     | 15.68 | 0.000   |

**Fuente:** Minitab 16

Diferenciación: 1 regular, 1 estacional de orden 12

Número de observaciones: Serie original 168, después de diferenciar 155

Se ha seleccionado el modelo de un promedio móvil autorregresivo integrado (ARIMA). Este modelo asume que el mejor pronóstico disponible para datos futuros está

dato por el modelo paramétrico que reacciona el valor más reciente con los valores y ruidos previos

### **Ecuación de pronóstico**

Ecuación de pronóstico estimado para la recaudación de tributos internos

$$\hat{Y}_t = Y_{t-1} + Y_{t-12} - Y_{t-13} + \hat{\theta}_1 \varepsilon_{t-1} - \hat{\theta}_{12} \varepsilon_{t-12} + \hat{\theta}_{13} \varepsilon_{t-13}$$

### **Validación del modelo estimado ARIMA (0,1,1)(0,1,1)**

En esta fase se comprueba la adecuación del modelo estimado se verifica que los residuos siguen un proceso de ruido blanco, el cual es:

#### **1.- Adecuación de los parámetros estimados**

$$|\theta| < 1$$

Los valores de los parámetros son menores a 1, por lo tanto cumplen la condición de invertible.

$$\text{MA (1) } (0.6911) < 1$$

$$\text{SMA (12) } (0.8537) < 1$$

#### **2.- Significancia de los parámetros**

Las probabilidad de los parámetros estimados son significativos  $\rho = (0.000) < \alpha = (0.05)$ , los parámetros son significativos, la cual es condición necesaria de un buen modelo.



### 3.- La bondad de ajuste del modelo

El error medio cuadrado del modelo elegido es menor con respecto a los demás modelos analizados

### 4.- Análisis de los residuales

Los residuos del modelo estimado se aproximan al comportamiento de un ruido blanco.

#### Contraste global de Box Pierce

Planteamiento de hipótesis para el modelo ARIMA (0, 1,1) (0, 1,1)<sub>12</sub>

$H_0$ : Los residuos siguen un proceso de ruido blanco.

$H_a$ : Los residuos no siguen un proceso de ruido blanco.

Nivel de significancia  $\alpha = 0.05 = 5\%$

#### Prueba estadística:

$$Q^* = T(T + 2) \sum_{k=1}^M (T - k)^{-1} P_k^2$$

Donde:

M: Número Máximo de rezagos a analizar.

T: Número total de observaciones

$R_j$ : La función de autocorrelación de los errores del proceso.

$$Q^* = 23.4$$

$$Q^* = 23.4 \rightarrow X^2_{(M-p-q, \alpha)}$$

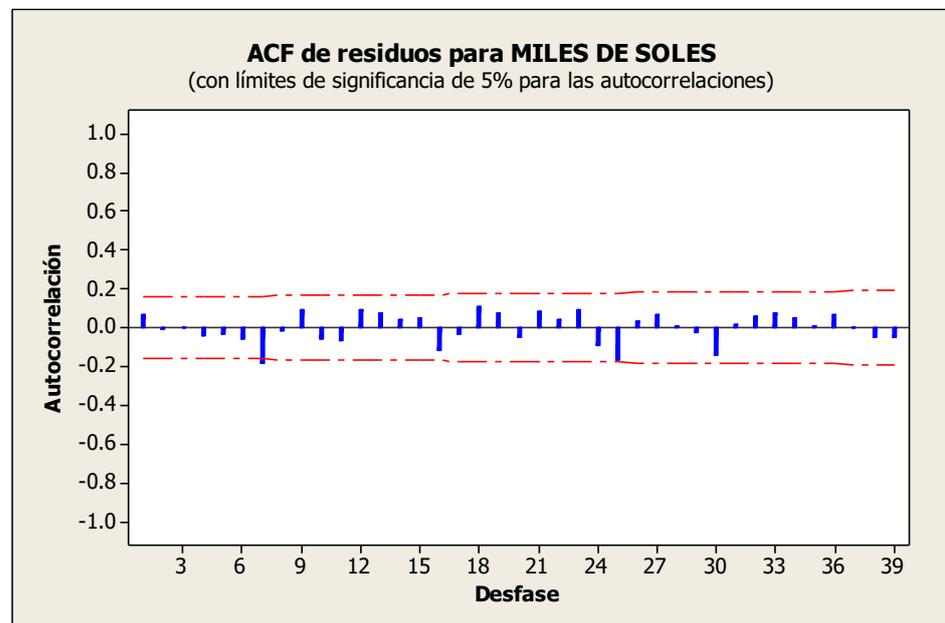
**Tabla 5** Estadística Chi-cuadrada modificada de Box-Pierce (Ljung-Box) Para el modelo de la serie transformada

| Desfase      | 12    | 24    | 36    | 48    |
|--------------|-------|-------|-------|-------|
| Chi-cuadrada | 11.3  | 23.4  | 37.0  | 36.6  |
| GL           | 10    | 22    | 34    | 46    |
| Valor P      | 0.338 | 0.377 | 0.333 | 0.446 |

Fuente: Minitab 16

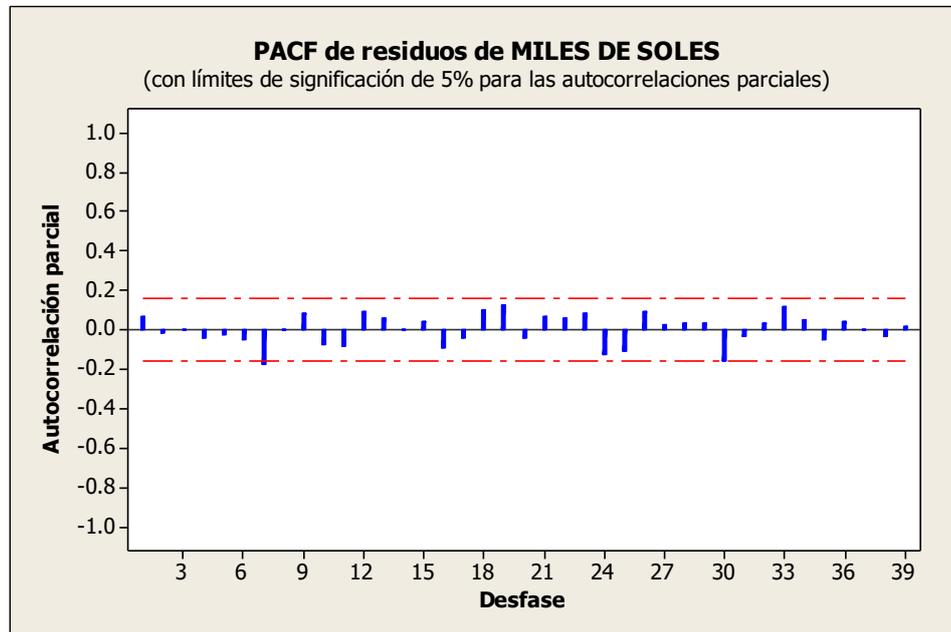
### Decisión

Como  $Q^* = 23.4 < X^2$  con una  $P=0.377$ , se acepta la hipótesis por lo tanto se llega a la conclusión de que la  $\alpha_t$  (errores residuales) siguen el proceso de ruido blanco a un 95% de confianza.



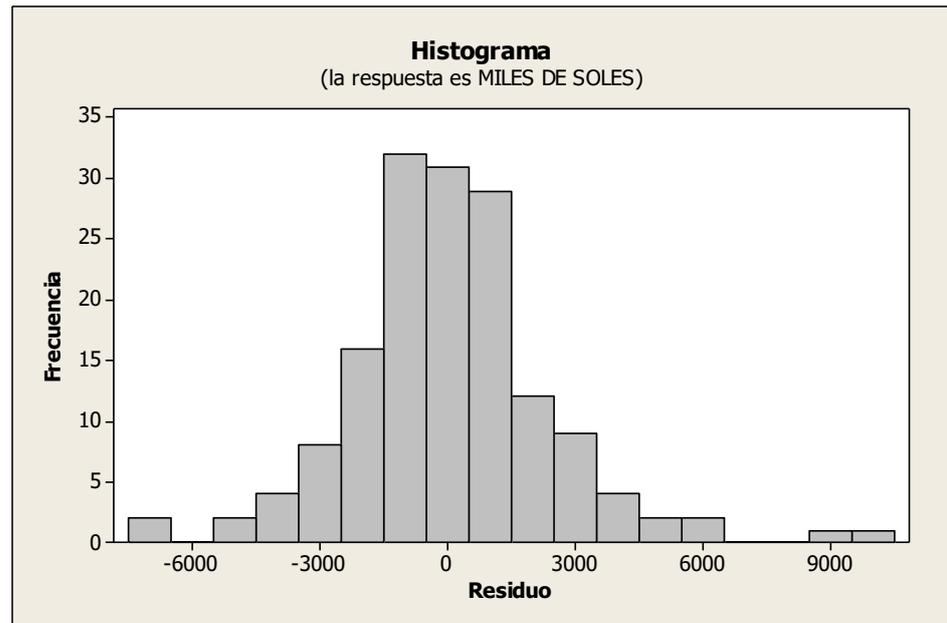
**Figura 13** Función de autocorrelación de los residuos del modelo estimado con límites de significancia del 5% para las autocorrelaciones.

En la Figura 13 solo se tiene un coeficiente de autocorrelación significativo de los residuos del modelo estimado, con esto se confirma que la serie es estacionaria.



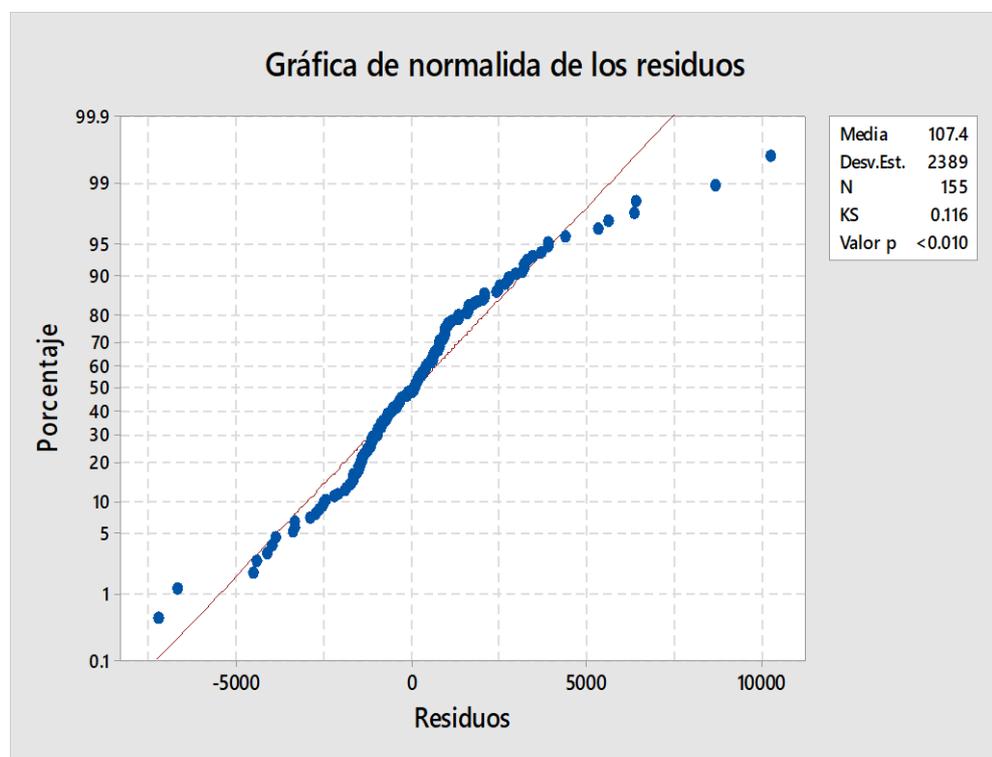
**Figura 14** Función de autocorrelación parcial de los residuos del modelo estimado con límites de significancia del 5% para las autocorrelaciones.

En la Figura 14 se observó que en el gráfico solo hay un coeficiente de autocorrelación simple y parcial de los residuos, con la cual se confirma la independencia de los errores, verificando que el ruido blanco se distribuye normalmente.



**Figura 15** Distribución de normalidad de los residuos del modelo

En la figura 15 el histograma de distribución de los residuales se observa la simetría de la serie en estudio, y se duda de la normalidad porque tiene valores altos y pequeños que podrían causar la no normalidad.



**Figura 16** Ploteo de la Normalidad de los residuos del modelo estimado

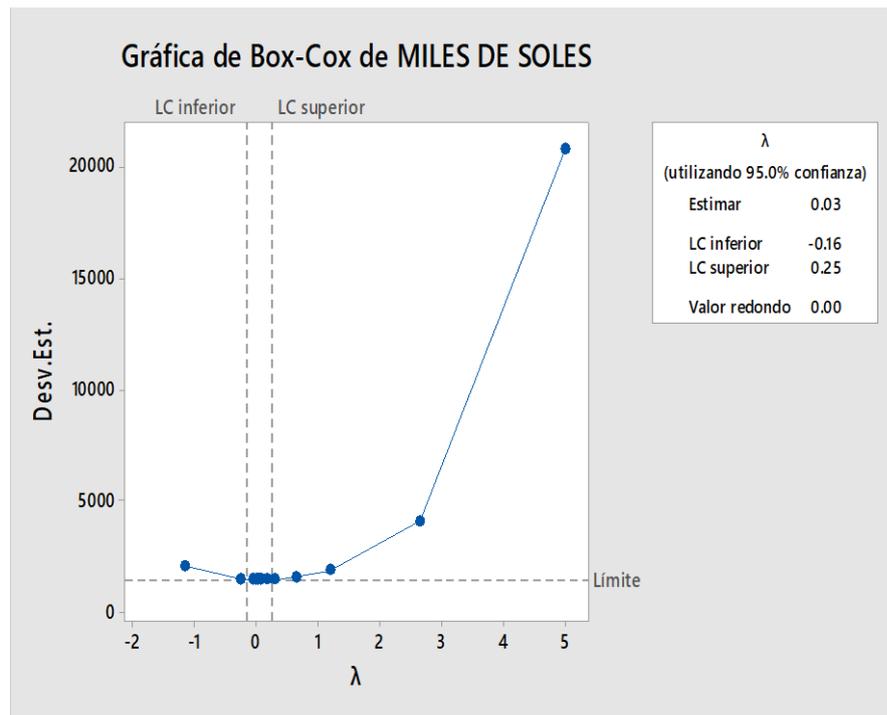


En la Figura 16, Se observa en el grafico que los errores se no distribuyen normalmente utilizando la prueba estadística de Kolmogorov Smirov, a un nivel de confianza del 5% (0.05) mayor al valor calculado 0.01, rechazándose la hipótesis nula de normalidad en los residuos.

La estructura del proceso estocástico mediante modelos AR o MA, debemos asegurar de que la serie temporal, como muestra del proceso, los errores deben destruirse normalmente para garantizar los resultados obtenidos, caso contrario hay que transformar la serie temporal original de tal forma que la nueva serie transformada si lo desea (Melo & Santana, 2016).

Cuya situación se da en modelos de corte temporal en los que la evolución histórica haya manifestado diferentes períodos en cuanto a los valores de una variable en relación a su media, agrupando en algún subperíodo valores altos y en otros valores pequeños.

Por ello se procede a la transformación de Cox-Box para garantizar los resultados y la normalidad de los residuos.



**Figura 17** Transformación de Box-Cox

El valor encontrado por la transformación de Box-Cox es de  $\lambda = 0$ , las transformaciones de Box y Cox son una familia de transformaciones potenciales usadas para corregir sesgos en la distribución de errores, en este caso para aproximar a una normal.

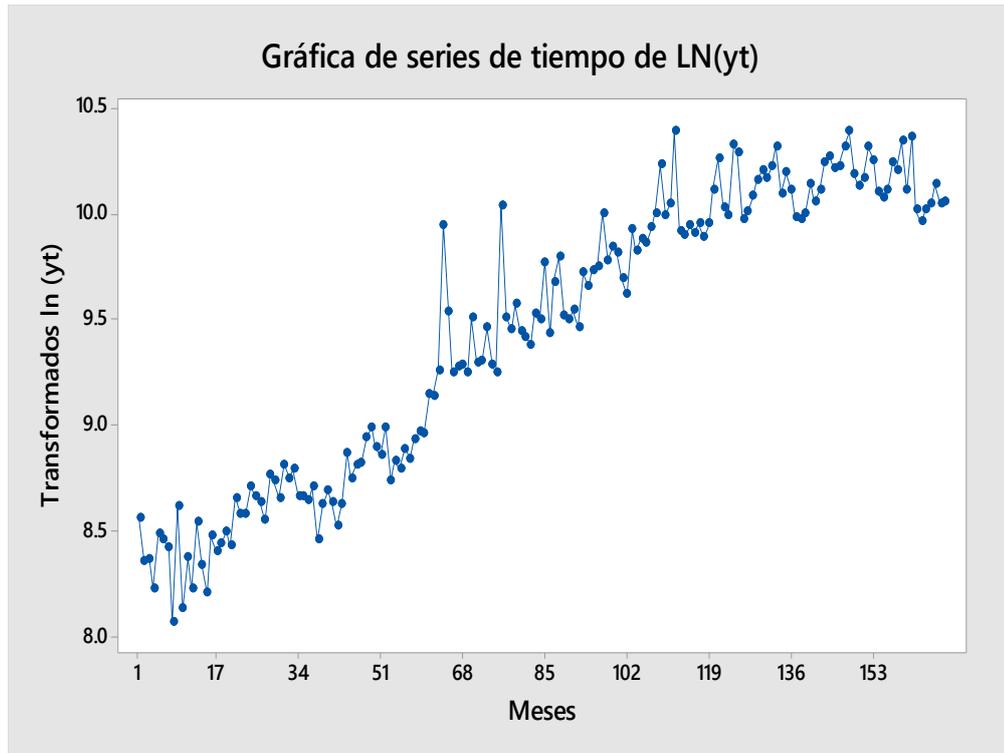
Las transformaciones Box-Cox no solo estabilizan la varianza sino que también mejoran la aproximación a la distribución normal del proceso  $y_t$  (González Casimiro, 2009, pág. 44).

Donde:

$$Y_t^\lambda = \begin{cases} (Y_t^\lambda - 1)/\lambda & , \lambda \neq 0 \\ \ln y_t & , \lambda = 0 \end{cases}$$

Siendo  $\lambda = 0$  esto hace que la transformación Box-Cox sea  $\ln y_t$

$$F(y_t) = \ln y_t$$



**Figura 18** Valores de la transformación Box-Cox de recaudaciones tributarias mensuales, con respecto a tributos internos de los años 2004 – 2017.

En la figura 18 observamos la serie de recaudaciones tributarias internas mensuales con transformación Box-Cox en que los valores oscilan entre 8 y 10.5, para cumplir la normalidad en los residuos.

Recalculando los parámetros del modelo ARIMA (0,1,1) (0,11)12

### Ecuación de pronóstico

Ecuación de pronóstico estimado para la recaudación de tributos internos transformados

$$\hat{Y}_t = Y_{t-1} + Y_{t-12} - Y_{t-13} + \hat{\theta}_1 \varepsilon_{t-1} - \hat{\theta}_{12} \varepsilon_{t-12} + \hat{\theta}_{13} \varepsilon_{t-13}$$

**Tabla 6** Resumen de los parámetros del modelo de la serie transformada  
*ARIMA (0, 1,1) (0,1,1)*

| Parameter      | Estimate | Std. Error | T     | P-value |
|----------------|----------|------------|-------|---------|
| <b>MA(1)</b>   | 0.6261   | 0.0608     | 10.30 | 0.000   |
| <b>SMA(12)</b> | 0.8974   | 0.0497     | 18.04 | 0.000   |

$$Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} + 0.6261\varepsilon_{t-1} - 0.8974\varepsilon_{t-12} + 0.5619\varepsilon_{t-13}$$

### Validación del modelo estimado ARIMA (0,1,1)(0,1,1) de las serie transformada

En esta fase se comprueba la adecuación del modelo estimado se verifica que los residuos siguen un proceso de ruido blanco, el cual es:

#### 1.- Adecuación de los parámetros estimados

$$|\theta| < 1$$

Los valores de los parámetros son menores a 1, por lo tanto cumplen la condición de invertible.

$$\text{MA (1) } (0.6261) < 1$$

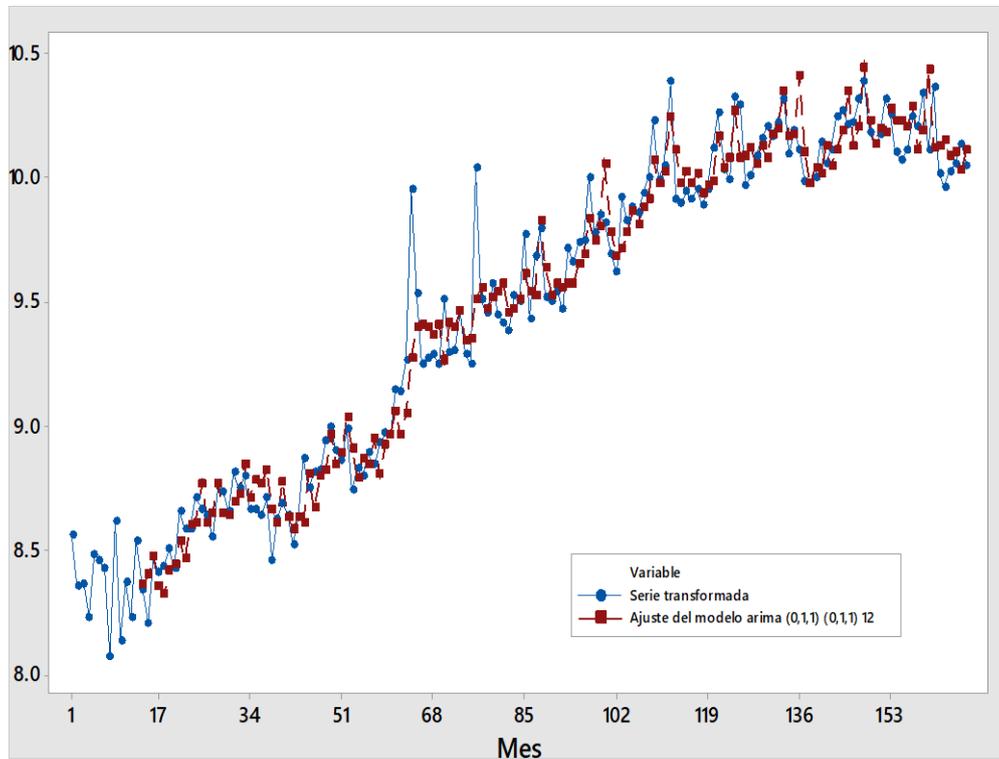
$$\text{SMA (12) } (0.8974) < 1$$

#### 2.- Significancia de los parámetros

Las probabilidad de los parámetros estimados para estimado ARIMA (0,1,1) (0,1,1) de las serie transformada son significativos  $\rho = (0.000) < \alpha = (0.05)$ , los parámetros son significativos, la cual es condición necesaria de un buen modelo.

### 3.- La bondad de ajuste del modelo

El error medio cuadrado y error medio absoluto porcentual de los datos pronosticados con respecto a los datos originales.



**Figura 19** Valores reales y ajustados de la serie transformada.

En la figura 19 se observa el ajuste del modelo ARIMA (0,1,1) (0,1,1)12 donde claramente podemos apreciar que la mayoría de los puntos se ajustan por muy cerca de los datos originales salvo algunos valores ajustados que están por debajo de los reales, sin embargo en el comportamiento general se ve que el modelo tiene pocos errores lo que supondrían una normalidad en ellos.

**Tabla 7** Bondad de ajuste del modelo de la serie transformada

| MAPE       | EMC        |
|------------|------------|
| 0.99045449 | 0.01642438 |

En la tabla 7 observamos que el error medio cuadrático es de 0.0164 lo que indica un error bajo ya que este valor se encuentra en unidades cuadráticas de miles de soles, un



error medio porcentual de 0.99% lo que indica que los pronósticos erraran en un 0.99% en promedio.

Según Lewis (1982) sobre el error medio porcentual (MAPE)

- Menor a 10.0%: Pronósticos de alta precisión
- Entre 10% y 20% : Buenos pronósticos
- Entre 21% y 50%: Pronósticos razonables
- Mayor que 50% : Pronósticos inexactos, dudosos

Con respecto al error medio porcentual (MAPE), los pronósticos son de alta precisión siendo 0.99% el valor calculado.

#### 4.- Análisis de los residuales

Los residuos del modelo estimado se aproximan al comportamiento de un ruido blanco.

#### Contraste global de Box Pierce

Planteamiento de hipótesis para el modelo ARIMA (0, 1,1) (0, 1,1)<sub>12</sub>

$H_0$ : Los residuos siguen un proceso de ruido blanco.

$H_a$ : Los residuos no siguen un proceso de ruido blanco.

Nivel de significancia  $\alpha = 0.05 = 5\%$

#### Prueba estadística:

$$Q^* = T(T + 2) \sum_{k=1}^M (T - k)^{-1} P_k^2$$



Donde:

M: Número Máximo de rezagos a analizar.

T: Número total de observaciones

$R_j$ : La función de autocorrelación de los errores del proceso.

$$Q^* = 23.4$$

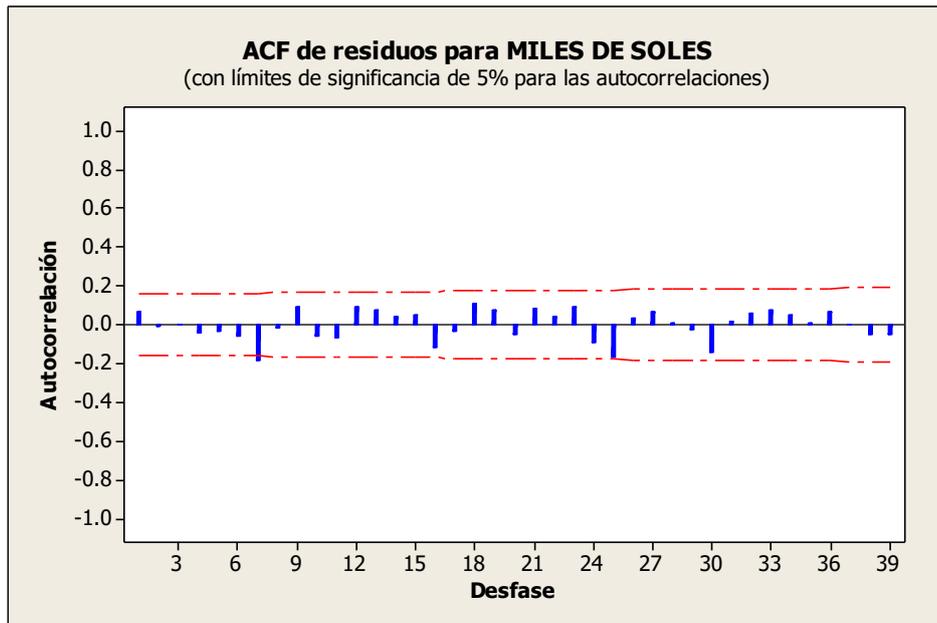
$$Q^* = 23.4 \rightarrow X^2_{(M-p-q, \alpha)}$$

**Tabla 8** Estadística Chi-cuadrada modificada de Box-Pierce (Ljung-Box) Para el modelo de la serie transformada

|              |       |       |       |       |
|--------------|-------|-------|-------|-------|
| Desfase      | 12    | 24    | 36    | 48    |
| Chi-cuadrada | 17.79 | 26.74 | 38.46 | 49.04 |
| GL           | 9     | 21    | 33    | 45    |
| Valor P      | 0.338 | 0.178 | 0.236 | 0.314 |

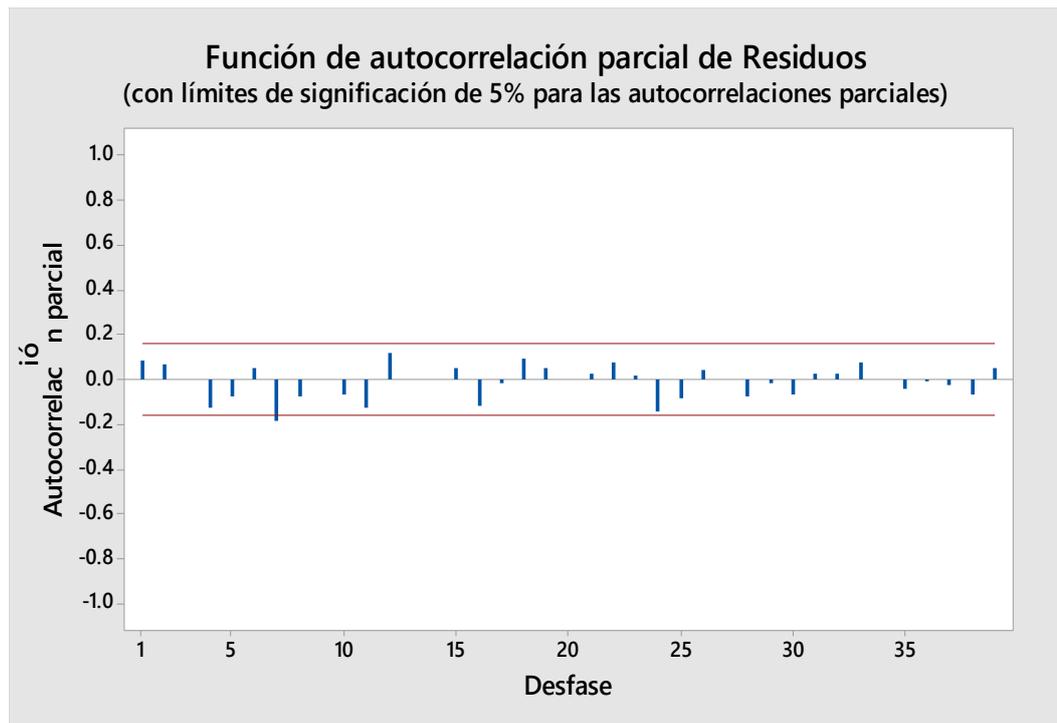
### Decisión

Como  $Q^* = 23.4 < X^2$  con una  $P=0.178$ , se acepta la hipótesis por lo tanto se llega a la conclusión de que la  $\alpha_t$  (errores residuales) siguen el proceso de ruido blanco a un 95% de confianza.



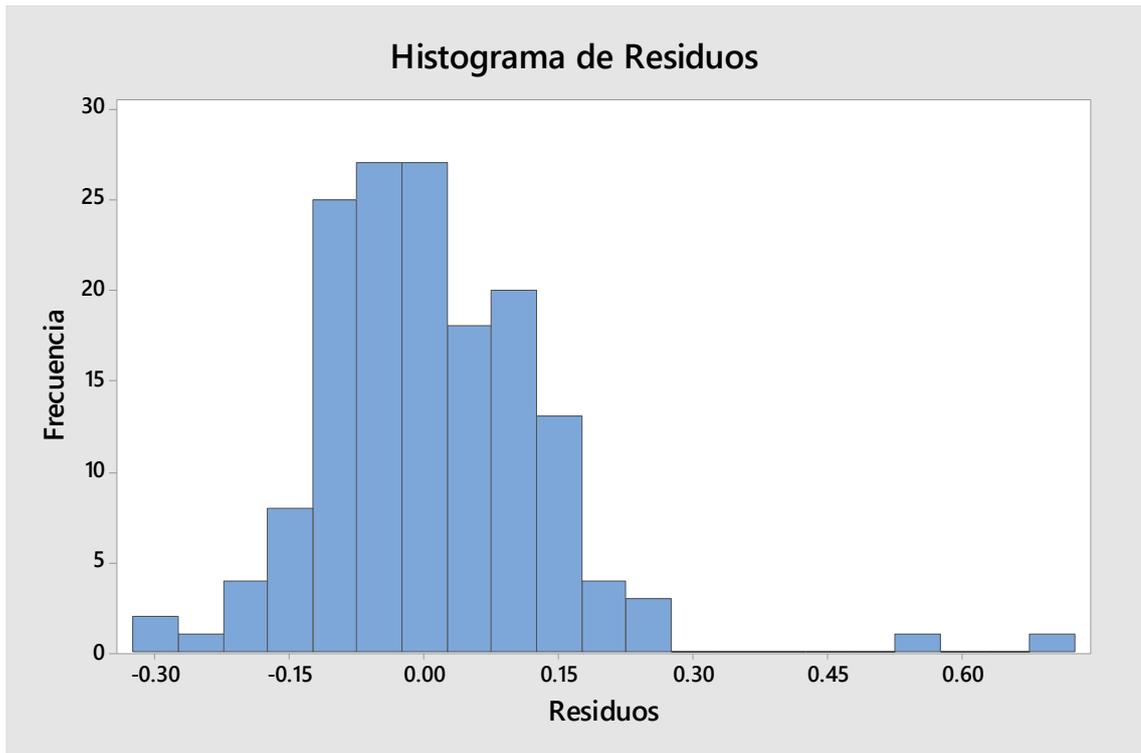
**Figura 20** Función de autocorrelación de los residuos del modelo estimado con límites de significancia del 5% para las autocorrelaciones.

En la Figura 20 solo se tiene un coeficiente de autocorrelación significativo en el desfase 7 de los residuos del modelo estimado de la serie transformada, con esto se confirma que la serie es estacionaria.



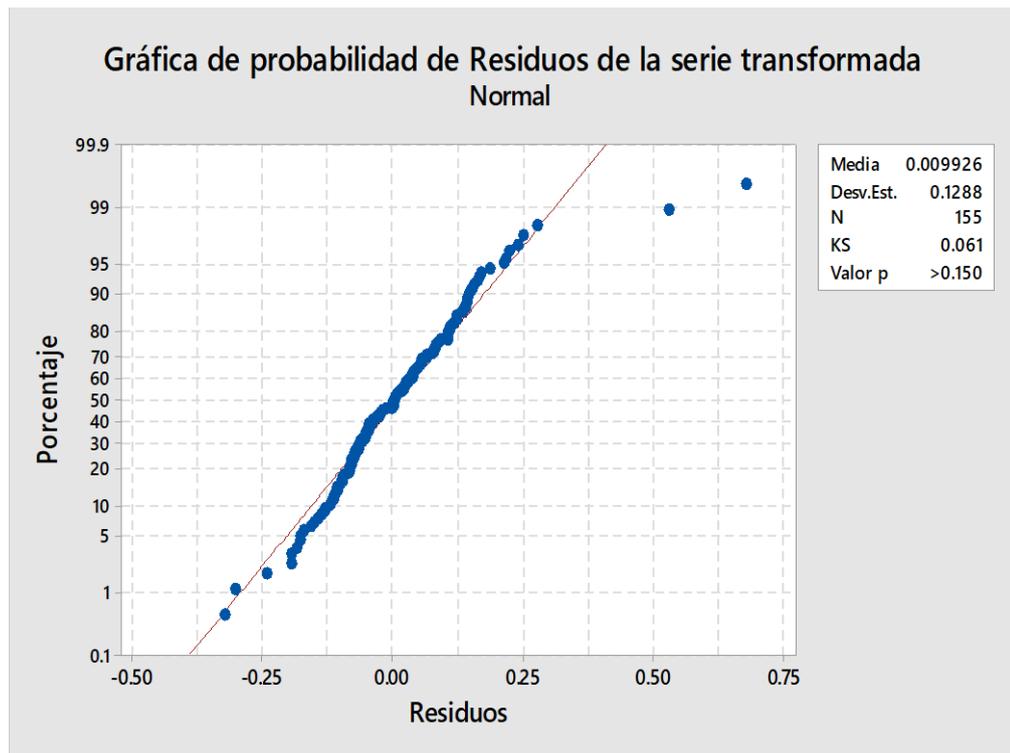
**Figura 21** Función de autocorrelación parcial de los residuos del modelo de la serie transformada con límites de significancia del 5% para las autocorrelaciones.

En la Figura 21 se observó que en el gráfico solo hay un coeficiente de autocorrelación simple y parcial de los residuos que es el desfase 6, con la cual se confirma la independencia de los errores, verificando que el ruido blanco se distribuye normalmente.



**Figura 22** Distribución de normalidad de los residuos de la serie transformada.

En la figura 22 el histograma de distribución de los residuales en la serie transformada se observa la simetría de la serie en estudio y la eliminación de algunos puntos atípicos con respecto a los datos originales.

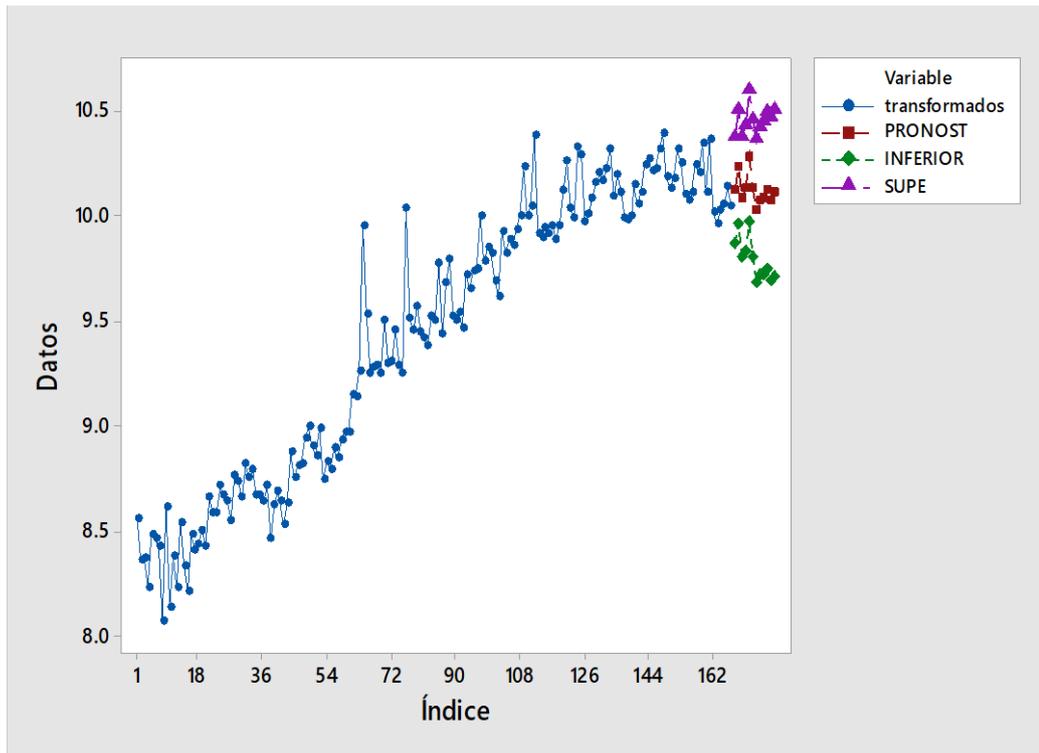


**Figura 23** Normalidad de los residuos nuevo modelo de los datos con transformación Box-Cox

En la Figura 23 se aprecia en el gráfico que los residuos del modelo de la serie con transformación Box-Cox. Siendo  $\lambda = 0$  esto hace que la transformación Box-Cox sea  $\ln y_t$ , distribuyéndose normalmente utilizando la prueba estadística de Kolmogorov Smiroy, a un nivel de confianza del 5% (0.05) menor al valor calculado 0.15, aceptándose la hipótesis alterna de normalidad en los residuos.

Cumpléndose la normalidad en los residuos que nos permite realizar pronósticos garantizando los resultados obtenidos.

Pronóstico con el modelo estimado ARIMA (0, 1,1) (0, 1,1)<sub>12</sub> de la serie con transformación Box-Cox. Siendo  $\lambda = 0$ , la nueva transformación es  $\ln y_t$ .



**Figura 24** Pronostico de la serie transformada

En la Figura 24 se observa los pronóstico de 24 valores que por defecto nos da el programa estadístico de la serie transformada  $\ln y_t$  , también muestra los intervalos de confianza del 95% de estos pronósticos que siguen en el sentido de dichos valores, asumiendo que el modelo ajustado es apropiado.

Convirtiendo la serie a los datos originales.

$$\text{Como } F(y_t) = \ln y_t$$

$$\text{Entonces } y_t = e^{\ln y_t}$$

Eliminándose el numero  $e$  con  $\ln$  quedando solo los datos originales.



**Tabla 9** Pronostico del modelo de la serie transformada de la Recaudación de Tributos Internos en la Región de Puno para el año 2018

| Periodo | Pronostico | Límite Inferior | Límite Superior |
|---------|------------|-----------------|-----------------|
| Ene-18  | 27064.2666 | 20975.7608      | 34920.0457      |
| Feb-18  | 23253.1977 | 17714.1282      | 30524.2911      |
| Mar-18  | 24452.077  | 18328.0011      | 32622.4374      |
| Abr-18  | 28249.5554 | 20852.0278      | 38271.452       |
| May-18  | 24272.5205 | 17656.7071      | 33367.2213      |
| Jun-18  | 21728.2915 | 15586.8042      | 30289.6377      |
| Jul-18  | 22775.6826 | 16120.7158      | 32177.9581      |
| Ago-18  | 23059.4056 | 16112.3821      | 33001.7114      |
| sep-18  | 23877.4115 | 16477.5354      | 34600.4888      |
| Oct-18  | 23012.3774 | 15690.4543      | 33751.0631      |
| Nov-18  | 23352.6159 | 15737.5876      | 34652.3674      |
| Dic-18  | 23949.4823 | 15957.805       | 35943.3959      |

**Fuente:** Superintendencia Nacional de Aduanas y de Administración Tributaria

**Elaborado:** Por el investigador

De la Tabla 9 se puede observar numéricamente la predicción realizada, permitiendo obtener resultados a futuro. La información proyectada es de 24 meses, desde la información existente de la cantidad en miles de soles recaudados por tributos internos de la SUNAT.

Los valores previstos en miles de soles para la recaudación de tributos internos en la Región de Puno, durante el periodo donde los datos reales están disponibles, también se muestran los valores predichos a partir del modelo ajustado y los residuos. Para los periodos de tiempo más allá del final de la serie, se muestra el 95% de predicción para las previsiones correspondientes a la entidad.

Estos límites muestran donde es probable encontrar los verdaderos, valores en un momento futuro seleccionado al 95% de confianza, asumiendo que el modelo ajustado es apropiado para los datos.



Para las observaciones futuras, muestra una predicción constante que el comportamiento razonable estará la futura observación.

Resultados que son apoyados por Condori (2014) quien concluyó que el modelo estacional autorregresivo integrado del media móvil SARIMA (0, 1, 1) (0, 1, 1)<sup>4</sup> es confiable de pronóstico para los pagos del impuesto predial para la gerencia de administración tributaria Puno, la cual satisface las pruebas del Ljung –Bbox Price. Siendo la única diferencia el periodo estacional ya que en este trabajo de investigación el periodo estacional es de 12 y por Condori (2014) utilizo 4 periodos estacionales.

Asimismo por Santiago, Coronel, & Galarza (2011) quienes sostienen que el desagio tributario es una estrategia eficaz que debe ser usada por la administración tributaria (SUNAT) para incrementar los niveles de ingresos en el país, obteniendo un efecto esperado mediante una adecuada educación y concientización de los contribuyentes, lo que repercutiría en mayores ingresos que serán aplicados en el desarrollo del país.

Finalmente Amaya (2014) Afirma que los impuestos constituyen una herramienta vital para alcanzar el bienestar de la sociedad, destinados a cubrir los gastos permanentes del Estado, como la provisión de salud, educación, defensa, administración de justicia, orden interno e infraestructura, entre otros bienes y servicios públicos.

## V. CONCLUSIONES

**PRIMERA:** El modelo estimado ARIMA (0, 1,1) (0, 1,1)<sub>12</sub> de la serie con transformación Box-Cox siendo  $\lambda = 0$ , la nueva transformación es  $\ln y_t$  pudo proyectar numéricamente los pronósticos, permitiendo obtener resultados a futuro. La información proyectada es de 12 meses, desde la información existente de la cantidad en miles de soles recaudados por tributos internos de la SUNAT, siendo este modelo  $Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} + 0.6261\varepsilon_{t-1} - 0.8974\varepsilon_{t-12} + 0.5619\varepsilon_{t-13}$ , resultado ser válido, ya que los errores del modelo transformado tienen una distribución normal, siendo los pronósticos de alta precisión con un MAPE de 0.99%.

**SEGUNDA:** La serie en estudio cuenta con fluctuaciones a lo largo de los años, esta serie cuenta con una tendencia creciente aditiva, también se puede apreciar que existen picos que indican variabilidad en la serie de tiempo, el mes con mayor ingreso tributario es el mes de abril del 2016 con una recaudación de tributos internos en la Región de Puno de S/.32442.68303 miles de soles y el ingreso tributario con menor ingreso es el mes de agosto del 2004 con una recaudación de tributos internos en la Región de Puno de S/. 3191.43969 miles de soles.

**TERCERA:** Se tiene 5 de los 42 coeficientes en los retardos 1, 2, 7, 11 y 13 son estadísticamente significativos a un nivel de confianza del 95%, debido a que esta fuera de los límites y se observa que la serie muestra un decremento



rápido sin llegar a cero, resultados que confirman la existencia de estacionalidad a lo largo del tiempo.



## VI. RECOMENDACIONES

**PRIMERO:** Se sugiere utilizar la transformación Box-Cox cuando un modelo ARIMA cuando los residuos no presentan una distribución normal ya que se pudo evidenciar en este trabajo de investigación que dicha transformación es muy útil para cumplir el supuesto de normalidad en los residuos.

**SEGUNDO:** Se sugiere que el modelo obtenido sea comparado con otras realidades que se presentan en las diversas Regiones de Nuestro País en donde también se realizan este tipo de recaudaciones, comparando la variable tomada en este estudio y agregándole factores de recaudación de aduanas (tributos externos, importaciones, etc.), para conseguir modelos óptimos y tener pronósticos más acertados.

**TERCERO:** Sería interesante la utilización de otras metodologías de predicción alternativas a la que se utilizó en este estudio, de modo que se pueda contrastar cuál de ellas permite obtener mejores estimaciones a partir de series temporales. En este sentido, la utilización de redes neuronales o algoritmos genéticos, entre otras metodologías, puede ser de gran utilidad.

**CUARTO:** Las predicciones deben considerar de mucha importancia para cada entidad, porque dan una visión de lo que puede suceder a futuro con las variables analizadas y lograr una planificación, para tomar acciones, para el mejor desarrollo de la sociedad.



## VII. REFERENCIAS BIBLIOGRÁFICAS

- Amaya González, A. S. (2014). La Recaudación Tributaria en el Ecuador: Análisis de Series de Tiempo con Cambios de Régimen (1993 - 2011). *Tesis de Pregrado*. Escuela Politécnica Nacional, Quito.
- Cabana Huarilloclla, R. (2017). La Cultura Tributaria Y Su Influencia En El Cumplimiento De Las Obligaciones Tributarias De Los Comerciantes Del Mercado Central Del Distrito De Santa Lucia, Periodo 2015. *Tesis Pregrado*. Universidad Nacional del Altiplano, Puno.
- Catalán, C. E. (2004). Series Temporales. *Laboratorio de Estadística. Consejo Superior de Investigaciones Científicas (CSIC). España, 54*.
- Coila Curo, M. (2013). Modelo De Predicción Mensual De Mortalidad General Intrahospitalaria En El Hospital Regional Manuel Nuñez Butrón-Puno, 2008-2016-I. *Tesis Pregrado*. Universidad Nacional del Altiplano, Puno.
- Condori Chambi, R. Y. (2014). Modelo de Pronóstico del Impuesto Predial de Administración Tributaria de la Municipalidad Provincial de Puno - 2014. *Tesis de Pregrado*. Universidad Nacional del Altiplano - Puno, Puno.
- De Arce, R., & Mahía, R. (2009). Conceptos básicos sobre la heterocedasticidad en el modelo básico de regresión lineal tratamiento con e-views. *Madrid España: Departamento de Economía Aplicada, Universidad Autónoma de Madrid*.



- De La Roca, J., & Hernández, M. (2004). Evasión Tributaria E Informalidad En El Perú: Una Aproximación A Partir Del Enfoque De Discrepancias En El Consumo. *Doctorado en Economía*. Universidad Texas A&M, Lima.
- Decreto Legislativo N° 1270. (2004). Delegado en el Poder Ejecutivo la facultad de legislar en materia tributaria referida tanto a tributos internos como aduaneros. Diario oficial el peruano. Lima, 01 de enero de 2004.
- Decreto Legislativo N° 133-2013-EF. (2013). Decreto supremo que aprueba el texto único ordenado del código tributario. Diario oficial el peruano. Lima, 22 de junio de 2013.
- Farfán, Y. M. (2006). Historia de la Contabilidad Pública en el Perú. *Contabilidad y Negocios*, 1(1), 32-35.
- García Díaz, J. C. (2016). Predicción en el dominio del tiempo. Análisis de series temporales para ingenieros. *Colección Manual de referencia*.
- García Estela, N. E. (2018). La incorporación de la norma XVI del título preliminar del código tributario, como medida antielusiva general en el Perú-análisis del artículo 3° del decreto legislativo n° 1121.
- González Casimiro, P. (2009). *Análisis de Series Temporales: Modelos ARIMA*. Lejona: Facultad de Ciencias Económicas y Empresariales.
- Grández Vásquez, E. (2014). La presión Tributaria y su Relación con la Recaudación Fiscal en el Perú: 1990 - 2012. *Tesis de Pregrado*. Universidad Nacional de Trujillo, Trujillo.



Guerrero Guzmán, V. M. (2003). *Análisis Estadístico de Series de Tiempo Económicas*.

Mexico: International Thomson Editores, S. A.

Gujarati, D., & Porter, D. (2010). *Econometría*. Mexico D. F.: MCGRAW-HILL.

Hanke, J. (2006). *Pronósticos en los Negocios*. México: PERSON EDUCACIÓN.

Instituto Nacional de Estadística e Informática (INEI). (1 de Junio de 2002). *INEI*.

Recuperado el 15 de AGOSTO de 2018, de INEI Web site:

<https://www.inei.gob.pe>

Juez Barona, S. Y., & Martillo Jeremias, L. D. (2017). Modelo Económico de la Relación entre la Presión Fiscal y el Crecimiento Económico del Ecuador en el Periodo 2007 - 2016. *Tesis de Pregrado*. Universidad Laica Vicente Rocafuerte de Guayaquil, Guayaquil.

Lewis, C. (1982). *International and Business Forecasting Methods*. London: Butterworths.

Mendieta Mendoza, A. P., & Rodríguez Emperador, R. R. (2004). Estimación de la evasión tributaria para el impuesto al valor agregado (IVA) por consumo interno en el Ecuador. *Tesis Pregrado*. Instituto De Ciencias Humanísticas y Económicas, Guayaquil.

Melo Morín, j., & Santana Esparza, G. (2016). Minado de series de tiempo utilizando la metodología ARIMA. *Revista de Investigación y Desarrollo*, 2(5), 21-31.

Mostacero Miranda, M. L., & Vásquez Mosqueira, K. M. (2014). Análisis Comparativo de la Aplicación de un Programa de Orientación y Cultura Tributaria En Estudiantes del Quinto Año de Secundaria de Instituciones Educativas Públicas y



Privadas de la Ciudad de Trujillo. *Tesis Pregrado*. Universidad Privada Antenor Orrego, Trujillo.

Quintanilla de la Cruz, E. (2014). La Evasión Tributaria y su Incidencia en la Recaudación Fiscal en el Perú y Latinoamérica. *Tesis de Doctorado*. Universidad de San Martín de Porres, Lima.

Ramos Inga, A. P. (2020). Pronóstico de los ingresos tributarios mensuales del gobierno central peruano aplicando redes neuronales y modelos Sarima, en base a los años 2003–2018. Universidad Nacional de Piura.

Santiago Ramos, J. F., Coronel Armas, R. E., & Galarza Aguirre, N. B. (2011). La Aplicación del Desagio Tributario y su Efecto en la Recaudación Tributaria en el Departamento de Junín. *Revista Científica Estudiantil*. Universidad Nacional del Centro del Perú, Huancayo.

Solera Rodríguez, M. L. (2015). Modelo de pronóstico para ingresos tributarios en condiciones de incertidumbre. *Trabajo Final de Graduación*. Universidad de Costa Rica, San Pedro.

Spiegel, M., & Stephens, L. (2001). *Estadística*. Mexico D. F.: McGraw-Hill Interamericana.

Uriel Jiménez, E. (1985). *Análisis de series temporales*. Madrid: Paraninfo S.A.

Villalobos Villegas, P. (2014). Pronóstico de ingresos tributarios recaudados de impuestos generales a las ventas a nivel nacional, por la SUNAT, mediante la metodología de box Jenkins y redes, neuronales artificiales para el año 2014. Universidad Nacional Pedro Ruíz Gallo.