

**Universidad Nacional de Ingeniería**  
Facultad de Ingeniería Económica, Estadística y  
Ciencias Sociales



TESIS

**Modelos SARIMA, SARIMAX y VAR en la predicción de gastos de  
la Universidad Nacional de Ingeniería  
Por la Fuente de Financiamiento de Recursos Directamente  
Recaudados**

PARA OBTENER EL TÍTULO PROFESIONAL DE INGENIERO ESTADÍSTICO

Elaborado por  
Antonieta Macedo Dávila  
0009-0002-6765-0094

Asesora  
Dra. Anna Sikov  
0000-0001-9869-5952

**LIMA - PERU**  
**2024**

---

Citar/How to cite	(Macedo, 2024)
Referencia/Reference	Macedo, A. (2024). <i>Modelos SARIMA, SARIMAX y VAR en la predicción de gastos de la Universidad Nacional de Ingeniería Por la Fuente de Financiamiento de Recursos Directamente Recaudados</i> . [Tesis de pregrado, Universidad Nacional de Ingeniería] Repositorio institucional Cybertesis UNI.
Estilo/Style: APA (7ma ed.)	

---

## Dedicatoria

Dedicado a mis padres Vicente y María por todo su amor, dedicación y esfuerzo.

Dedicado a mi esposo Fernando y mi hija Iris Maribel por estar juntos como familia y unidos para cualquier proyecto que emprendamos.

Dedicado a la memoria de mis suegros Justiniano y Encarnación, por su cariño, cuidado y su confianza.

## Agradecimiento

Mi agradecimiento es a Dios y la naturaleza por los adorables padres que tengo, por el esposo compañero, amigo que la vida me dio; por ver cada día en la mirada de mi hija que todo valió la pena.

A la Facultad de Ingeniería Económica, Estadística y Ciencias Sociales de la Universidad Nacional de Ingeniería, así como también a los docentes que con sus conocimientos y ejemplo contribuyeron a mi formación profesional.

Agradezco de manera especial a la Dra. Anna Sikov y al Dr. José Javier Cerdá Hernández por su invaluable apoyo en el desarrollo y culminación de la presente tesis.

Agradezco al Mag. Rafael Jimmy Caparó Coronado por sus valiosas sugerencias al presente trabajo.

## **Resumen**

En la presente investigación se ha aplicado tres modelos econométricos: SARIMA, SARIMAX y VAR; los dos primeros, basados en la metodología Box-Jenkins y Tiao y el tercero en el Modelo de Vectores Autorregresivos, debido a la retroalimentación que hay entre las variables de los mismos, para determinar la capacidad predictiva del modelo de ejecución de gastos, por la fuente de financiamiento de recursos directamente recaudados de la Universidad Nacional de Ingeniería se emplearon los datos mensuales de la recaudación de ingresos corrientes y ejecución de gastos para el periodo comprendido entre enero del 2010 a diciembre del 2018 de la mencionada institución.

En la estimación y la capacidad predictiva el modelo SARIMA dio un mejor ajuste con un error cuadrático medio (ECM) de 2.904 en comparación con los modelos SARIMAX y VAR de 3.627 y 4.972, respectivamente. Lo cual se sostiene sobre el hecho de que la ejecución de gastos tiene una estacionalidad en los procesos de adquisición de bienes y servicios durante el año fiscal de acuerdo a la Ley N° 30225 – Ley de contrataciones del estado, artículo 20 y el Decreto Legislativos N° 1439.

### **Palabras clave:**

SARIMA, SARIMAX, VAR, Recaudación de ingresos, Ejecución de gastos, Predicción, Estacionariedad.

## **Abstract**

In the present research three econometric models have been applied: SARIMA, SARIMAX and VAR; the first two, based on the Box-Jenkins and Tiao methodology and the third on the Autoregressive Vector Model, due to the feedback between the variables of the same, to determine the predictive capacity of the expenditure execution model, by the financing source of directly collected resources of the National University of Engineering monthly data of current revenue collection and expenditure execution for the period from January 2010 to December 2018 of the mentioned institution were used. In estimation and predictive ability, the SARIMA model gave a better fit with a mean squared error (MSE) of 2.904 compared to the SARIMAX and VAR models of 3.627 and 4.972, respectively. This is supported by the fact that the execution of expenditures has a seasonality in the procurement of goods and services during the fiscal year according to Law No. 30225 - State Contracting Law, Article 20 and Legislative Decree No. 1439.

## **Key Word:**

SARIMA, SARIMAX, VAR, Revenue collection, Expenditure execution, Prediction, Stationarity.

## Tabla de Contenido

Resumen	v
Abstract	vi
Introducción	xiii
<b>Capítulo I: Parte Introductoria del trabajo</b>	<b>1</b>
1.1. Generalidades	1
1.2. Descripción del Problema de investigación	2
1.2.1. Problema General	2
1.2.2. Problemas Específicos	2
1.3. Objetivos del estudio	2
1.3.1. Objetivo General	2
1.3.2. Objetivos Específicos	3
1.4. Antecedentes Relacionados con la Investigación	3
1.4.1 Antecedentes Internacionales	3
1.4.2 Antecedentes Nacionales	5
<b>Capítulo II: Marco Teórico y Conceptual</b>	<b>7</b>
2.1 Series de Tiempo	7
2.1.1 Proceso Estocástico	9
2.1.2. Condiciones de Estacionariedad	10
2.2 La metodología de Box y Jenkins (1976)	12
2.2.1 Funciones de autocovarianza y autocorrelación	12
2.2.2 Modelos ARMA	13
2.2.3 Modelos para series no estacionarias	18
2.2.4 Identificación del modelo	22
2.2.5 Estimación de los parámetros	25
2.2.6 Predicciones bajo los modelos SARIMA	27
2.2.7 Validación del modelo	29
2.3 Modelos con variables exógenas	29
2.3.1 El modelo de regresión con errores autorregresivos	30
2.3.2 El modelo de regresión con errores ARIMA	31
2.3.3 Modelos SARIMAX	32
2.3.4 Modelos de Función de Transferencia	32
2.4 Modelo Vectorial Autorregresivo (VAR)	35

2.4.1	Definición de un modelo VAR	35
2.4.2	Proceso estacionario en covarianza	37
2.4.3	Autocorrelaciones de un proceso VAR	37
2.4.4	Análisis de raíz unitaria	38
2.4.5	Estimación de un modelo VAR	40
2.4.6	Análisis de los Residuos del Modelo VAR	41
2.4.7	Análisis de Cointegración	42
2.4.8	Análisis de la Función Impulso Respuesta	44
2.4.9	Predicción puntual e intervalo	45
2.4.10	Descomposición de la varianza del error de predicción	46
<b>Capítulo III: Marco Metodológico</b>		<b>48</b>
3.1	Diseño Metodológico	48
3.1.1	Tipo de Investigación	48
3.1.2	Nivel de Investigación	48
3.1.3	Diseño de investigación	48
3.2	Método de investigación	48
3.2.1	Población	49
3.2.2	Muestra	49
3.3	Definiciones operativas	49
3.3.1	Variables de estudio	49
3.3.2	Operacionalización de variables	50
3.4	Formulación de la hipótesis de investigación	50
3.4.1	Hipótesis General	50
3.4.2	Hipótesis Específicas	50
3.5	Técnica e instrumento de recolección de datos	51
3.6	Tratamiento estadístico de la información	51
3.6.1.	Metodología Box -Jenkins y Box Tiao	51
3.6.2.	Metodología del Modelo Vectorial Autorregresivo (VAR)	53
<b>Capítulo IV: Resultados y Discusión</b>		<b>55</b>
4.1	Modelo SARIMA para predicción de gastos futuros	56
4.2	Modelo SARIMAX para predicción de gastos futuros	65
4.3	Modelo VAR para predicción de gastos futuros	67
<b>Capítulo V: Conclusiones y Recomendaciones</b>		<b>74</b>
5.1.	Conclusiones	74



5.1.1. Conclusión General	74
5.1.2. Conclusión Especifica 1	74
5.1.3. Conclusión Especifica 2	74
5.1.4. Conclusión Especifica 3	75
5.2. Recomendaciones	76
<b>Referencias Bibliográficas</b>	<b>77</b>

## Índice de Tablas

<b>Contenido</b>	<b>Pág.</b>	
Tabla 1	Valores críticos de la ACF y PACF	24
Tabla 2	Test de Dickey y Fuller para el modelo (94)	39
Tabla 3	Resultados de aplicación de la prueba de Dickey-Fuller a la serie original de los gastos	57
Tabla 4	Resultados de aplicación de la prueba de Dickey-Fuller a la serie de las diferencias de orden 12	58
Tabla 5	Resultados de ajuste de modelos <i>SARIMA</i> a la serie original de los gastos: estimadores de los coeficientes de los modelos con sus p-valores respectivos (en las parentheses)	60
Tabla 6	Resultados de aplicación de la prueba de Ljung-Box a los residuos de los modelos $M_1$ , $M_2$ y $M_3$	62
Tabla 7	Resultados de aplicación de las pruebas de Shapiro-Wilk y Jarque-Bera a los residuos de los modelos $M_1$ , $M_2$ y $M_3$	63
Tabla 8	Resumen de los errores de predicción bajo los modelos $M_1$ , $M_2$ y $M_3$	65
Tabla 9	Resultados de ajuste de modelos <i>SARIMAX</i> a la serie original de los gastos, $Z_t$ usando los ingresos como variables exógenas, $X_t$ : estimadores de los coeficientes de los modelos con sus p-valores respectivos (en los paréntesis)	65
Tabla 10	Resultados de ajuste de modelos <i>SARIMAX</i> a la serie original de los gastos, $Z_t$ usando los ingresos como variables exógenas, $X_{t+2}$ : estimadores de los coeficientes de los modelos con sus p-valores respectivos (en los paréntesis)	66
Tabla 11	Resumen de los errores de predicción bajo los modelos $M_1$ , $M_2$ , $M_3$ y $\tilde{M}_1$ modificado, en los meses de enero-octubre del año 2018	67
Tabla 12	Prueba de cointegración de Johansen para las variables $(1 - L^{12})G_t$ y $(1 - L^{12})I_{t+2}$	68
Tabla 13	Criterios de información AIC, HQ, SC y FPE	69
Tabla 14	Sesgo, error absoluto y error cuadrático medio para el modelo ajustado <i>VAR</i> (1)	72
Tabla 15	Sesgo, error absoluto y error cuadrático medio de predicción para los meses de enero-octubre del año 2018, para el modelo ajustado <i>VAR</i> (1)	72

## Índice de Figuras

Contenido	Pág.
Figura 1 Serie con presencia de tendencia	8
Figura 2 Serie con presencia de estacionalidad	8
Figura 3 Serie con presencia de factores cíclicos	8
Figura 4 Serie con presencia de cambios de varianza a lo largo del tiempo	9
Figura 5 Correlograma de ruido blanco	11
Figura 6 Serie de ruido blanco con tendencia determinística	19
Figura 7 Serie de ruido blanco con tendencia estocástica	19
Figura 8 Serie con tendencia lineal y comportamiento estacional	21
Figura 9 Correlograma de la función de autocorrelación de la Serie $s_1$	24
Figura 10 Correlograma de la función de autocorrelación parcial de la Serie $s_1$	24
Figura 11 Serie histórica de la Recaudación de Ingresos Corrientes (En millones de soles constantes)	56
Figura 12 Serie histórica de la Ejecución del gasto (En millones de soles constantes)	56
Figura 13 Gastos observados en distinto meses	57
Figura 14 La serie de los gastos. después de aplicación de la transformación de diferencia de orden 12	59
Figura 15 Autocorrelaciones de la serie de los gastos después de aplicación de la transformación de diferencia de orden 12	59
Figura 16 Autocorrelaciones parciales de la serie de los gastos después de aplicación de la transformación de diferencia de orden 12	60
Figura 17 Autocorrelaciones de los residuos del modelo $M_1$	61
Figura 18 Autocorrelaciones de los residuos del modelo $M_2$	61
Figura 19 Autocorrelaciones de los residuos del modelo $M_3$	62
Figura 20 Q-Q plot de los residuos del modelo $M_1$	62
Figura 21 Q-Q plot de los residuos del modelo $M_2$	63
Figura 22 Q-Q plot de los residuos del modelo $M_3$	63
Figura 23 Comparación de los verdaderos gastos, realizados en el año 2018 con las predicciones bajo los modelos $M_1$ , $M_2$ y $M_3$	64
Figura 24 Ingresos y gastos: correlaciones cruzadas	66
Figura 25 Ajuste y residuos para I, usando un modelo $VAR(1)$ . La línea azul es el ajuste del modelo	70

Figura 26	Ajuste y residuos para G, usando un modelo $VAR(1)$ . La línea azul es el ajuste del modelo	70
Figura 27	Gráfica de los residuos del modelo $VAR(1)$ ajustado para las variables (a) Ingreso y (b) Gasto	71
Figura 28	Gasto observado vs estimación usando modelo VAR	73

## Introducción

La responsabilidad fiscal de toda institución pública, tiene por objetivo optimizar el empleo de los recursos públicos haciéndolos cada vez más predecibles y disminuyendo el nivel de decisión que se base en el criterio individual y subjetivo. Con este propósito se establecen normas que fijan objetivos a cumplir sobre variables como el gasto. Para ello es esencial tener una perspectiva presupuestaria a mediano plazo que posibilite ver un panorama más amplio que el presupuesto anual y poder predecir qué ingresos se obtendrán en el mediano plazo. Si bien el porcentaje de países que elaboran un marco presupuestario de mediano plazo en América Latina es considerable pocos lo hacen con métodos de pronóstico fiables y sólidos.

Toda institución, ya sea pública o privada, debe conocer el comportamiento futuro de ciertos fenómenos para poder planificar y prevenirlos en este contexto. La Universidad Nacional de Ingeniería implementa una política de generación de recursos propios, lo que la convierte en una de las primeras productoras de bienes y servicios entre las universidades públicas del país. La generación de recursos propios implica la recaudación de ingresos corrientes y la ejecución de gastos. El presente estudio determina un modelo econométrico que permite el pronóstico de las series futuras de la recaudación de ingresos corrientes y la ejecución de gastos por la fuente de financiamiento de recursos directamente recaudados de la Universidad Nacional Ingeniería entre 2010 y 2018 mediante la comparación de modelos SARIMA, SARIMAX y VAR (Modelo Vectorial Autorregresivo).

El principal aporte es demostrar que un modelo propuesto es capaz de obtener aproximaciones cercanas a la realidad en la predicción de la ejecución de gastos de esta casa de estudios.

La investigación consta de cinco capítulos:

En el capítulo I se encuentra la parte introductoria del trabajo que consta de generalidades del estudio donde se desarrolla la descripción la realidad problemática, descripción del problema de investigación, objetivos del estudio y antecedentes investigados.

El capítulo II contiene al marco teórico y conceptual.

En el capítulo III se encuentra el desarrollo del trabajo de investigación que consta del marco metodológico donde se determina el tipo, nivel diseño y método de la investigación, así como la definición de la población, muestra y variables del estudio, formulación de la hipótesis, técnica e instrumento de recolección de datos, técnicas de procesamiento de datos y tratamiento estadístico de la información.

El capítulo IV contiene Resultados y Discusión de la investigación.

Finalmente se encuentra las conclusiones a las que se llegó en el estudio, las recomendaciones y referencias bibliográficas.

## Capítulo I: Parte Introductoria del trabajo

### 1.1. Generalidades

Las universidades en el Perú pueden ser creadas por el estado o por iniciativa privada dentro de un marco legal sostenido en la constitución y la ley N° 30220, ley universitaria, (2014). Esta ley establece con más especificación el marco legal para la gestión y el financiamiento de las universidades. Destaca la autonomía universitaria de elaborar, reconocer su propio estatuto y gobernar de acuerdo con él, de estructurar su sistema académico, económico y administrativo, así como administrar sus bienes y rentas, realizar su presupuesto y administrar sus fondos de acuerdo a ley. Así mismo esta ley permite a las universidades la generación de sus propios ingresos a través de actividades y órganos dedicados a la producción de bienes y prestación de servicios.

El marco legal antes mencionado tiene una relación directa con las formas de financiamiento de la universidad que es una combinación de financiamiento público (transferencias directas, subvenciones diversas, exoneraciones tributarias, etc.) y privado (matriculas, donaciones filantrópicas, venta de servicios y otras actividades empresariales). Por otro lado, el Presupuesto Público como principal mecanismo de programación financiera y económica que tiene como finalidad cumplir con las funciones del Estado de asignación, distribución y estabilización genera cambios en la gestión pública proponiendo reformas en el desempeño de las instituciones del Estado. En este contexto la Dirección General de Presupuesto Público (DGPP), como ente rector del Proceso Presupuestario estableció a través de Normas Legales el enfoque de Presupuesto por Resultados a partir del año 2011. Dentro de este contexto la estrategia Institucional de la Universidad Nacional de Ingeniería está orientada a la mejora continua de la gestión presupuestal con la captación de ingresos, generados como consecuencia de la gestión de los centros de producción y ejecución de gastos; además la Directiva para la Programación y Formulación Anual del Presupuesto del Sector Público, con una Perspectiva de Programación Multianual, considera que “los Pliegos presupuestarios deberán registrar información sobre los flujos de ingresos tomando en cuenta los factores estacionales que pudieran incidir en su comportamiento, tales como: información estadística de la captación (tendencia al incremento o su disminución), captación de nuevos conceptos de ingresos, variación en el monto de las tasas, y cambios producidos en el marco legal, entre otros” (Directiva N° 0002-2022-EF/50.01 – Resolución Directoral N° 0005-2022-EF/50.01) por la fuente de financiamiento de recursos directamente recaudados.

## **1.2. Descripción del Problema de investigación**

Dado que la programación y formulación anual de presupuesto del sector público considera que, en las proyecciones, las entidades deben tener en cuenta las tendencias y los factores estacionales que influyan en su comportamiento a lo largo del tiempo, se presenta la necesidad de formular y estimar modelos econométricos óptimos para predecir la ejecución del gasto por Recursos Directamente Recaudados, para la toma de futuras decisiones en su financiamiento. Luego en este contexto se plantea el problema de la presente investigación.

### **1.2.1. Problema General**

¿Cuál es el mejor modelo SARIMA, SARIMAX o VAR para la predicción de la ejecución de gastos, por la fuente de financiamiento de recursos directamente recaudados, de la Universidad Nacional de Ingeniería en el periodo 2010 al 2018?

### **1.2.2. Problemas Específicos**

¿Cuál es el modelo SARIMA, para predecir la ejecución de gastos por la fuente de financiamiento de recursos directamente recaudados, de la Universidad Nacional de Ingeniería en el periodo 2010 al 2018?

¿Cuál es el modelo SARIMAX, para predecir la ejecución de gastos por la fuente de financiamiento de recursos directamente recaudados, de la Universidad Nacional de Ingeniería en el periodo 2010 al 2018?

¿Cuál es el modelo VAR, para predecir la ejecución de gastos, por la fuente de financiamiento de recursos directamente recaudados, de la Universidad Nacional de Ingeniería en el periodo 2010 al 2018?

## **1.3. Objetivos del estudio**

### **1.3.1. Objetivo General**

Obtener el mejor modelo SARIMA, SARIMAX o VAR para la predicción de la ejecución de gastos, por la fuente de financiamiento de recursos directamente recaudados, de la Universidad Nacional de Ingeniería en el periodo 2010 al 2018.



### **1.3.2. Objetivos Específicos**

Obtener el modelo SARIMA, para predecir la ejecución del gasto, por la fuente de financiamiento de recursos directamente recaudados, de la Universidad Nacional de Ingeniería en el periodo 2010 al 2018.

Obtener el modelo SARIMAX, para predecir la ejecución del gasto, por la fuente de financiamiento de recursos directamente recaudados, de la Universidad Nacional de Ingeniería en el periodo 2010 al 2018.

Obtener el modelo VAR para predecir la ejecución del gasto, por la Fuente de Financiamiento de Recursos Directamente Recaudados de la Universidad Nacional de Ingeniería en el periodo 2010 al 2018.

### **1.4. Antecedentes Relacionados con la Investigación**

Diversos autores de otras latitudes han tratado de identificar y explicar el comportamiento de la recaudación de ingresos en diferentes instituciones públicas y privadas. No obstante, de estas investigaciones se desprende la falta de consenso ocasionada por la inflación doméstica propia de cada país como manifiestan Gómez, Jiménez y Martner (2017).

#### **1.4.1. Antecedentes Internacionales**

##### **1.4.1.1. Título de investigación: Causalidad entre Ingresos y Gastos: Evidencia Empírica para España**

La tesis fue elaborada por Rocio Cueto Naredo en la Facultad de Economía y Empresa de la Universidad de Oviedo en septiembre 2015 en España.

El objetivo que plantea el estudio es analizar la relación existente entre ingresos y gastos en España entre los años 1964 y 2013, para lo cual se aplicó la metodología del test de raíces unitarias para conocer el orden de integración de las variables ingreso y gasto; obteniéndose que los ingresos y gastos españoles son integrados de orden 1 pues tienen una diferencia cada uno, seguidamente se realizó el test causalidad en el sentido de Granger (aproximaciones VEC) para conocer la dirección de la causalidad entre las variables; obteniéndose que en el sentido de Grange hay una relación unidimensional entre

los ingresos y gastos españoles y siendo estos últimos los que determinan los primeros y finalmente se procedió a realizar el test de cointegración de Johansen (aproximaciones VAR) para conocer si entre ambas variables existe una relación a largo plazo, los resultados confirmaron una relación a largo plazo por lo cual para reducir el déficit español se debe disminuir los gastos públicos.

#### **1.4.1.2. Título de investigación: Proyección de Series de Tiempo para el Consumo de la Energía Eléctrica a Clientes Residenciales en Ecuador**

El estudio fue elaborado por Manuel Aurelio Chong Fuentes en la Facultad de Ciencias Económicas y Administrativas de la Universidad Concepción, Chile en el año 2016.

El objetivo del estudio es presentar modelos que permitan la proyección del consumo mensual de energía eléctrica por parte de los usuarios del sector residencial en Ecuador entre 2013 y 2022. Se utilizó la metodología ARIMA para el caso univariado y ARIMAX para el caso multivariado de Box-Jenkins. Se encontró que el modelo ARIMA es mejor que el modelo ARIMAX para la proyección del consumo mensual de energía por parte de los usuarios del sector residencial debido. Se utilizaron las variables explicativas: los precios medios de electricidad para los usuarios residenciales, el ingreso y los precios del barril de petróleo.

#### **1.4.1.3. Título de investigación: Modelos de Ecuaciones Múltiples Modelos VAR y Cointegración**

La investigación fue realizada por Wbaldo Londoño en el Departamento de Ciencias Básicas de la Universidad EAFIT en noviembre del 2005 en Colombia.

El objetivo de la investigación es crear modelos de ecuaciones simultáneas, modelos VAR para aplicaciones y modelos de simulación de dos y tres variables. Para lograr esto, examina una variedad de modelos VAR, como la cointegración y la teoría de raíces unitarias. En este último, se examina cómo funcionan los valores de la matriz de coeficientes del modelo VAR. Se realizan aplicaciones en cada caso para confirmar que las series están cointegradas, lo que confirma la validez de la teoría explicada.

### **1.4.2 Antecedentes Nacionales**

#### **1.4.2.1. Título de investigación: Resumen Aplicativo sobre Econometría de Series de Tiempo.**

El estudio fue elaborado por Econ. Johann Paúl Lastra Chacón en el Instituto Nacional de Estadística e Informática (INEI) en octubre del 2001 en Perú.

El objetivo del estudio es presentar un documento que resume y aplica conocimientos básicos de la econometría moderna de series temporales como son el modelo ARIMA, VAR, la cointegración y modelos de corrección de errores para lo cual se desarrolla pronósticos con los Modelos ARIMA conteniendo un proceso autorregresivo, proceso de Media Móvil e Integrado de orden, para ello siguió la Metodología de Box-Jenkins: Identificación, estimación del modelo ARIMA, verificación del diagnóstico y Pronóstico y se usa datos económicos de los estados Unidos y datos del PBI peruano. En este último se llega a demostrar que el PBI peruano tiene un comportamiento no estacionario a lo largo del tiempo. Sin embargo, la primera diferencia de la serie de PBI peruano no presenta raíz unitaria es decir es estacionaria. Basándose en ese resultado se demuestra que el modelo más apropiado para los datos es el ARIMA (2,1,1) y emplean ese modelo para pronósticos futuros.

#### **1.4.2.2. Título de Investigación: Proyección de las Importaciones Peruanas – Una aplicación de la Superintendencia Nacional de Administración Tributaria SUNAT.**

La investigación fue realizada por Edgard Julio Ortega Huayta en Facultad de Ingeniería Económica y Ciencias Sociales de la Universidad Nacional de Ingeniería en el 2004.

El objetivo de la tesis es la proyección del valor de las importaciones. Para este propósito se utilizó el modelo ARIMA y modelos casuales (análisis de regresión). El autor demuestra que la proyección más adecuada de las importaciones se alcanzó con el uso combinado de la metodología causal y de series de tiempo, pues las predicciones obtenidas como una combinación de las proyecciones tuvo como resultado menores niveles de error en los periodos de evaluación en comparación con las predicciones basados en cada uno de los modelos.

### **1.4.2.3. Título de Investigación: La Inversión en Infraestructura Pública y el Crecimiento Económico en el Perú, periodo 1950-2013**

La tesis fue realizada por José Luis Andrés Mayurí Sánchez en la Facultad de Economía y Planificación de la Universidad Nacional Agraria de la Molina en el año 2015.

Este trabajo evalúa el impacto de la inversión pública en infraestructura en el crecimiento económico del Perú entre 1950 y 2013. El autor empleó un Modelo Vectorial Autorregresivo (VAR). Se llevó a cabo un análisis de los factores del producto bruto interno y la inversión bruta fija pública en Perú. Según el autor, la tasa de crecimiento de la inversión bruta fija pública en Perú tiene un impacto positivo en la tasa de crecimiento del producto bruto interno, y el 16.4% de las variaciones en el tiempo del crecimiento económico se debe a los choques producidos por la tasa de crecimiento de la inversión en infraestructura pública.

## Capítulo II: Marco Teórico y Conceptual

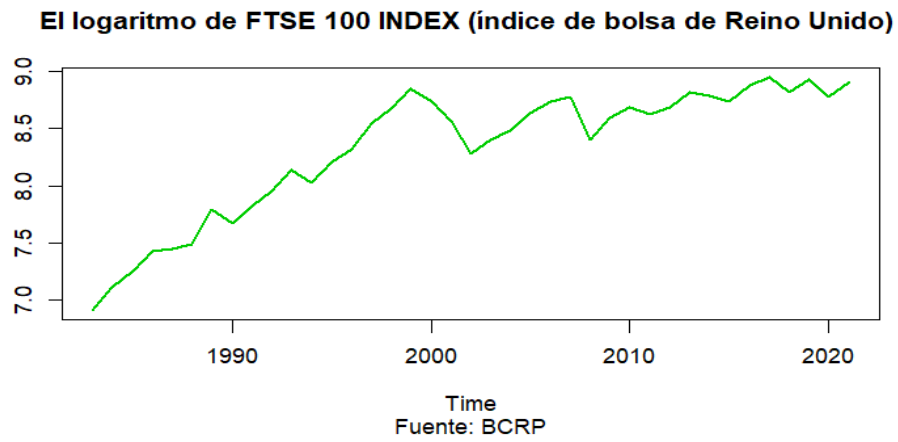
### 2.1 Series de Tiempo

Primero, definamos una serie de tiempo de forma general. De manera más sencilla, la serie temporal  $Z_t, t = 1, \dots, T$ , donde  $T$  es el tamaño de la serie, puede ser definida como una sucesión de observaciones ordenadas en el tiempo, de manera que la distancia cronológica entre dos observaciones consecutivas es constante. Cabe precisar que la definición se refiere a una serie temporal en el tiempo discreto, en el sentido que los valores de la serie pueden ser observados en tiempos definidos, por ejemplo, diariamente, semanalmente, mensualmente, anualmente, etc. Las series temporales en el tiempo discreto son muy comunes, principalmente en las áreas de economía y de ciencias sociales: ventas de un producto o servicio específico, PBI, el nivel de inflación, nivel de pobreza, de desempleo, consumo de energía, número de clientes atendidos, etc. Todas las series mencionadas pueden ser mensuales, trimestrales o anuales, además las series de ventas y del número de clientes atendidos pueden ser semanales o diarias. En principio, una serie puede ser continua en el tiempo. Por ejemplo, el voltaje de una señal eléctrica, dado que la misma pueda ser registrada en cada instante del tiempo. De igual manera, las variables climatológicas como la temperatura, radiación, humedad pueden generar series temporales en el tiempo continuo. Este tipo de series temporales está fuera del enfoque de este trabajo.

Esencialmente, una serie de tiempo puede ser considerada como una realización de un proceso estocástico  $\{Z_t\}$ , donde en cada tiempo  $t = 1, \dots, T$  se observa una sola realización de ese proceso. La pregunta primordial en el análisis de una serie temporal es ¿Si el proceso estocástico que generó dicha serie es estacionario?. Generalmente, es imposible definir las pruebas exactas para establecer la estacionariedad de la serie, puesto que a cada tiempo corresponde una única observación. En la mayoría de las aplicaciones se analiza la presencia de la tendencia, estacionalidad, los factores cíclicos y la dispersión de la serie a lo largo del tiempo, en la serie estudiada. Las Figuras 1, 2, 3 y 4 demuestran ejemplos de series de tiempo con tendencia, estacionalidad, factores cíclicos, cambios de varianza a lo largo del tiempo, respectivamente.

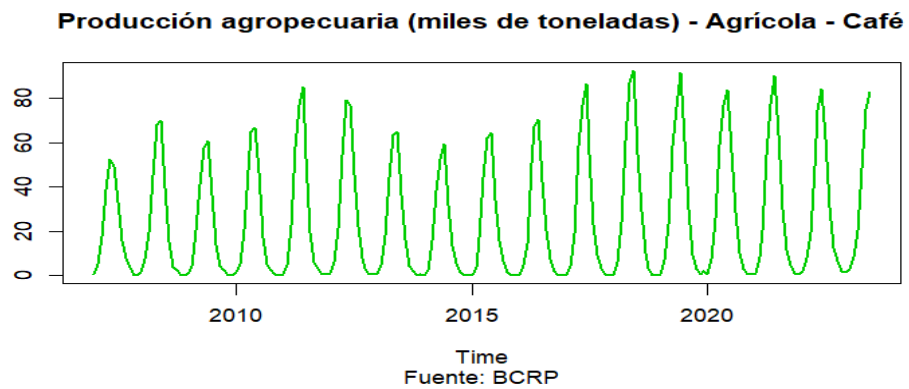
**Figura 1**

*Serie con presencia de tendencia*



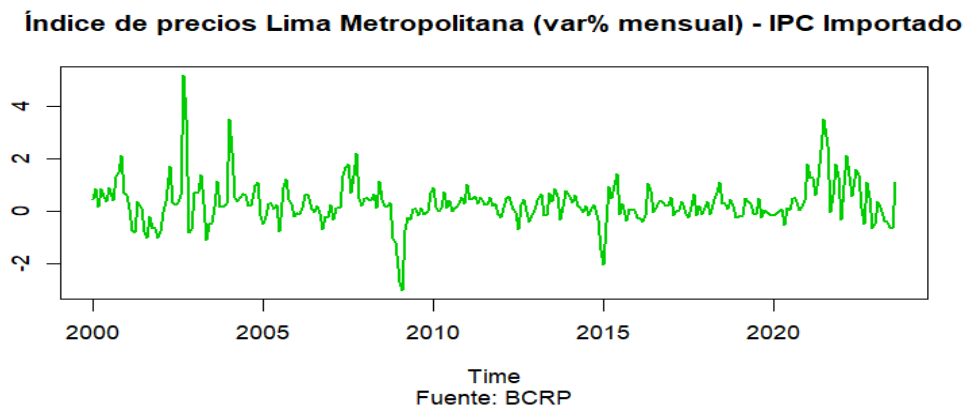
**Figura 2**

*Serie con presencia de estacionalidad*



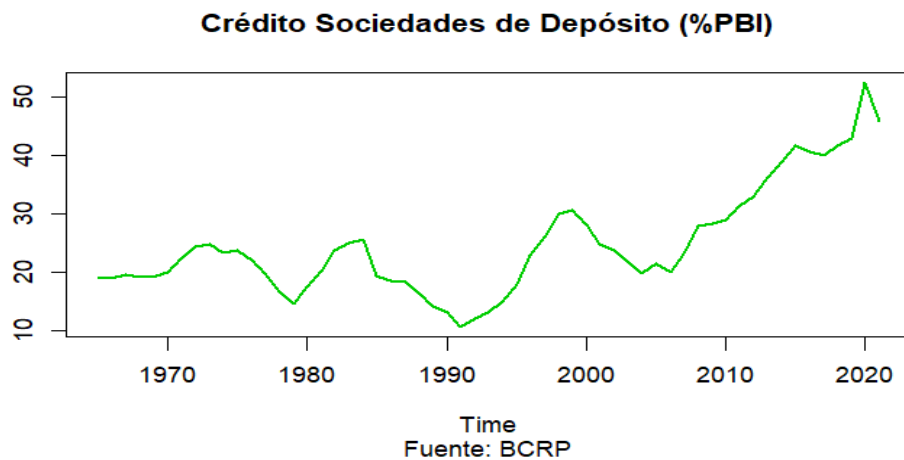
**Figura 3**

*Serie con presencia de factores cíclicos*



**Figura 4**

*Serie con presencia de cambios de varianza a lo largo del tiempo*



La presencia de por lo menos uno de los tres componentes especificados implica que la serie no es estacionaria, sin embargo, la ausencia de dichos componentes no comprueba la estacionariedad de la serie. En las siguientes secciones vamos a definir los procesos estocásticos y el concepto de estacionariedad de forma matemática.

### **2.1.1 Proceso Estocástico**

Siguiendo a Karatzas y Shreve (2012) un proceso estocástico es un conjunto de variables aleatorias  $\{Z_t\}_{t \in I}$ , indexadas por un conjunto  $I$  que representa la evolución temporal de la variable. Por ejemplo,  $I = \{0, 1, 2, \dots\}$ ,  $I = [0, \infty)$ ,  $I = \{\dots, -2, -1, 0, 1, 2, \dots\}$ . Un ejemplo más simple de un proceso aleatorio es el ruido blanco, que se define de la siguiente manera.

Un ruido blanco es un proceso estocástico  $\{\epsilon_t\}_{t=1}^{\infty}$  tal que las variables  $\epsilon_t$  son independientes e idénticamente distribuidos, donde  $E(\epsilon_t) = 0$  y  $Var(\epsilon_t) = 1$ ,  $\forall t$ . Si adicionalmente, los  $\epsilon_t$  siguen una distribución normal el proceso se denomina un ruido blanco gaussiano.

Otro ejemplo sencillo de un proceso estacionario es el paseo aleatorio, que se define como un proceso  $\{Z_t\}$ , cuyas primeras diferencias, es decir,  $\Delta_t = Z_t - Z_{t-1}$ , es un ruido blanco.

La estructura probabilística de un proceso estocástico se describe completamente mediante las distribuciones finito dimensionales,

$$F_{(t_1, \dots, t_K)}(z_1, \dots, z_K) = P(Z_{t_1} \leq z_1, \dots, Z_{t_K} \leq z_K), \forall K \in \{1, 2, \dots\}, \quad (1)$$

donde  $t_1 < t_2 < \dots < t_K$ . No obstante, para poder estimar esta distribución necesitamos una muestra aleatoria de realizaciones del proceso  $\{Z_t\}$ . Puesto que en la realidad disponemos de una sola realización, el uso de (1) no es factible.

Alternativamente, la determinación de las propiedades de un proceso estocástico puede realizarse mediante los momentos de las distribuciones marginales de  $Z_t$ . Ese abordaje sirve como la base en la teoría desarrollada por Box and Jenkins (1976), que utiliza los dos primeros momentos. Vale notar que, de forma general, los dos primeros momentos no son suficientes para describir una distribución. Sin embargo, este enfoque es el principal en las aplicaciones prácticas, por su simplicidad y factibilidad (Uriel, Peiró, 2000).

Ahora, definamos los primeros dos momentos de la distribución de  $Z_t$ . La función de medias o momento de primer orden, para cada instante del tiempo se precisa de la siguiente forma:

$$\mu_t = E(Z_t) \quad (2)$$

La función de varianzas, para cada instante del tiempo se define como

$$\gamma_t = Var(Z_t) = E(Z_t - \mu_t)^2 \quad (3)$$

La autocovarianza, para cada instante de los tiempos  $t$  y  $t + l$ , se define como

$$\gamma_{t,t+l} = Cov(Z_t, Z_{t+l}) = E(Z_t - \mu_t)(Z_{t+l} - \mu_{t+l}) \quad (4)$$

Para caracterizar un proceso estocástico se utilizan también los coeficientes de autocorrelación.

$$\rho_{t,t+l} = \frac{Cov(Z_t, Z_{t+l})}{\sqrt{Var(Z_t)Var(Z_{t+l})}} = \frac{\gamma_{t,t+l}}{\sqrt{\gamma_t \gamma_{t+l}}} \quad (5)$$

Como ya fue mencionado, teniendo sólo una realización de  $\{Z_t\}$ , es imposible estimar las expresiones del (2) al (5). En la siguiente sección definimos el concepto de estacionariedad de un proceso estocástico. Veremos más adelante, que el supuesto de estacionariedad es crucial para ajustar un cierto tipo de modelos (Box y Jenkins, 2016).

### 2.1.2 Condiciones de Estacionariedad

En esta sección definamos los conceptos de estacionariedad, siguiendo Karatzas y Shreve (2012).

Se manifiesta que un proceso estocástico  $\{Z_t\}$  es estacionario en sentido estricto si para toda  $m$ -upla  $t_1, t_2, \dots, t_m$  y todo entero  $h$ , los vectores aleatorios  $(Z_{t_1}, Z_{t_2}, \dots, Z_{t_m})$  y  $(Z_{t_1+h}, Z_{t_2+h}, \dots, Z_{t_m+h})$  tienen la misma distribución de probabilidad conjunta.

Esta definición es demasiado fuerte, puesto que, en las aplicaciones prácticas, donde se observa una sola realización del proceso, esta condición no puede ser verificada. Por lo tanto, se utilizan definiciones más débiles.



Un proceso estocástico  $\{Z_t\}$  es estacionario en media, o débilmente estacionario de primer orden si y sólo si

$$E(Z_t) = \mu, \quad \forall t \quad (6)$$

Es decir, un proceso estacionario en media, es un proceso cuya media se conserva constante a lo largo del tiempo.

La siguiente definición de estacionariedad es la definición que generalmente se utiliza en la práctica.

Se dice que un proceso estocástico  $\{Z_t\}$  es estacionario de autocovarianza o débilmente estacionario de segundo orden si y sólo si es estacionario en media y se cumplen las dos condiciones siguientes:

1º) La varianza es finita y se mantiene constante a lo largo del tiempo, es decir:

$$E(Z_t - \mu_t)^2 = \sigma^2 < \infty, \quad \forall t \quad (7)$$

2º) La autocovarianza entre observaciones en dos puntos de tiempo distintos dependen únicamente del lapso de tiempo transcurrido entre esos dos puntos, es decir:

$$E[(Z_{t+k} - \mu) \cdot (Z_t - \mu)] = \gamma_k, \quad \forall t, k = \pm 1, \pm 2, \dots \quad (8)$$

La definición implica que el valor,  $\gamma_k$ , es independiente del periodo de tiempo  $t$ .

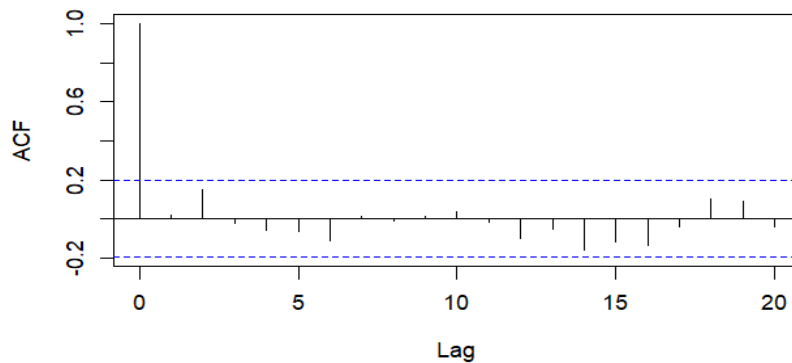
En este trabajo utilizamos la definición de estacionariedad de segundo orden. Es decir, en este trabajo, el concepto de una serie de tiempo estacionaria se refiere a una realización de un proceso estocástico débilmente estacionario de segundo orden.

Obviamente, en un proceso débilmente estacionario de segundo orden, las autocorrelaciones pueden ser obtenidas de la siguiente forma:

$$\rho_k = \frac{\gamma_k}{\gamma_0}, \quad k = \dots -2, -1, 0, 1, 2, \dots \quad (9)$$

**Figura 5**

*Correlograma de ruido blanco*



En el análisis de series temporales se utiliza el correlograma, que es una representación gráfica de autocorrelaciones  $\rho_k$  para  $k = 0,1,2,3, \dots$ . En la Figura (5) se demuestra un ejemplo de correlograma de ruido blanco, donde las líneas azules punteadas corresponden a los valores críticos que corresponden a las hipótesis de  $H_0: \rho_k = 0$  versus la alternativa  $H_1: \rho_k \neq 0$  (ver la sección 2.2.3) para más detalle.

## **2.2 La metodología de Box y Jenkins (1976)**

Según Box y Jenkins (1976), la idea principal en análisis de series temporales es estudiar el patrón del comportamiento de la serie, basándose en las realizaciones observadas de la serie para poder predecir los valores futuros de la misma. El análisis de la estructura de las autocovarianzas de una serie es una herramienta muy importante en el proceso de ajuste de modelos. Box and Jenkins (1976) desarrollan modelos ARMA y los utilizan para realizar predicciones futuras.

La metodología desarrollada por Box y Jenkins (1970) consiste en cuatro pasos, donde en el primer paso, se analiza la estructura de las covarianzas para identificar los modelos posibles que pueden describir bien el comportamiento de la serie.

Como veremos más adelante, a cada modelo ARMA corresponde una estructura específica de autocorrelaciones. El uso de estos patrones facilita el proceso de identificación del modelo que se ajusta mejor a los datos. Las autocorrelaciones pueden ser estimadas a partir de los datos. El siguiente paso es la estimación de los parámetros de los modelos identificados en el primer paso y comparación entre los modelos, basándose en significancia de los parámetros, criterio AIC, etc. En este paso pueden ser seleccionados uno o varios modelos candidatos para ser utilizados para hacer inferencia. El tercer paso es la validación de los supuestos del modelo, que se basa principalmente en evaluación de la normalidad y de las autocorrelaciones de los residuos. Los modelos validados en el tercer paso, pueden ser utilizados para realizar predicciones. Cabe notar, que este paso también puede ser utilizado para comparar modelos candidatos mediante la comparación de los errores de la predicción de los valores observados. Finalmente, el mejor modelo se utilizará para realizar predicciones futuras. En este trabajo vamos a presentar todos los avances de la metodología de Box y Jenkins (1970) con detalle.

### **2.2.1 Funciones de autocovarianza y autocorrelación**

El primer paso del abordaje de Box y Jenkins (1970) es el análisis de la estructura de las autocorrelaciones de la serie original.

Dado  $\gamma_k$  se denomina función de autocovarianza  $\{\gamma_k\}$  del proceso estocástico, luego el coeficiente de autocorrelación  $\rho_k$  es una función del rezago  $k$  y se denomina función de autocorrelación  $\{\rho_k\}$  del proceso estocástico. Es propicio mencionar que la función de autocorrelación es independiente de la escala de medida de la serie temporal es decir es adimensional ya que se cumple que  $\gamma_k = \rho_k \sigma_z^2$ , dicho de otra forma, el conocimiento de la función de autocorrelación  $\{\rho_k\}$  y la varianza  $\sigma_z^2$  es equivalente al conocimiento de la función de autocovarianza  $\{\gamma_k\}$  (Box, Jenkins, Reinsel y Ljung, 2016).

Ahora bien, si una serie temporal,  $Z_1, Z_2, \dots, Z_T$  es estacionaria, los estimadores de  $\mu$  y  $\gamma_k$ ,  $k = 0, \pm 1, \pm 2, \dots$  se obtienen de la siguiente manera

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^T Z_t = \bar{Z} \quad (10)$$

$$\widehat{\gamma}_k = \frac{1}{T} \sum_{k=1}^T (Z_t - \bar{Z})(Z_{t-k} - \bar{Z}) \quad (11)$$

Los estimadores para  $\rho_k$  se obtienen como

$$\widehat{\rho}_k = \frac{\widehat{\gamma}_k}{\widehat{\gamma}_0} = \frac{\sum_{k=1}^T (Z_t - \bar{Z})(Z_{t-k} - \bar{Z})}{\sum_{t=1}^T (Z_t - \bar{Z})^2} \quad (12)$$

Existen ciertos teoremas de ergodicidad que demuestran que los estimadores  $\hat{\mu}$  y  $\widehat{\gamma}_k$  son consistentes para  $\mu$  y  $\gamma_k$ . Es decir,  $\hat{\mu}$  y  $\widehat{\gamma}_k$  tienden a  $\mu$  y  $\gamma_k$  respectivamente, en probabilidad, si  $T$  es suficientemente grande. Específicamente, la condición suficiente para que  $\hat{\mu}$  sea consistente para  $\mu$  es que  $\rho_k$  tiende a cero cuando  $k$  es suficientemente grande, lo que implica que si dos valores observados con diferencia de tiempo muy grande, son incorrelacionados.

En las siguientes subsecciones definiremos los modelos ARMA y presentamos los patrones generales de comportamiento de autocorrelaciones en el caso de modelos autorregresivos, *AR* y de medias móviles, *MA*. Además, definiremos autocorrelaciones parciales, que también son utilizados en el primer paso de identificación de modelos candidatos.

### **2.2.2 Modelos ARMA**

Los modelos ARMA son muy comunes en el análisis de series de tiempo estacionarias, desarrollados por Box y Jenkins (1976). El modelo implica que la realización de la serie en el tiempo  $t$ ,  $Z_t$ , puede ser expresado como una combinación lineal de las

realizaciones de la serie y de los errores aleatorios en tiempos pasados. Así el modelo ARMA incluye dos partes: la parte autorregresiva (AR) y la parte de medias móviles (MA). Definamos este modelo de forma matemática.

Sea  $Z_1, Z_2, \dots, Z_T$  una serie temporal estacionaria, tal que  $E(Z_t) = 0$ . Entonces, el modelo  $ARMA(p, q)$  se define de la siguiente forma

$$Z_t - \phi_1 Z_{t-1} - \dots - \phi_p Z_{t-p} = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q} \quad (13)$$

donde  $\phi_1, \dots, \phi_p$  son los coeficientes autorregresivos y  $\theta_1, \dots, \theta_q$  son los coeficientes de medias móviles,  $\epsilon_t \sim N(0, \sigma^2)$  es una serie de ruido blanco. Cabe precisar que el supuesto de normalidad no es necesario para la definición de los modelos, desarrollados por Box y Jenkins (1970). Sin embargo, puesto que los métodos de inferencia para estos modelos se basan en el supuesto de los errores normales, nosotros siempre asumimos dicha distribución.

Si  $E(Z_t) = \mu$ , el modelo (13) incluirá una constante. Específicamente, reescribiendo el modelo (13) como

$$(Z_t - \mu) - \phi_1(Z_{t-1} - \mu) - \dots - \phi_p(Z_{t-p} - \mu) = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q}, \quad (14)$$

obtenemos, que

$$Z_t - \phi_1 Z_{t-1} - \dots - \phi_p Z_{t-p} = C + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q} \quad (15)$$

donde  $C = \mu(1 + \phi_1 + \phi_2 + \dots + \phi_p)$ .

El modelo (15) puede ser presentado como

$$\phi(L)Z_t = \theta(L)a_t \quad (16)$$

donde  $\phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$ ,  $\theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q$  y  $L$  es el operador retardo, tal que  $L^k Z_t = Z_{t-k}$  y  $L^k \epsilon_t = \epsilon_{t-k}$ .

Los modelos  $ARMA(p, q)$  tienen dos subclases particulares: los modelos autorregresivos y los modelos de medias móviles. En las siguientes secciones vamos a presentar estos modelos y describir sus estructuras de las autocovarianzas correspondientes.

### 2.2.2.1 Modelos de Media Móvil $MA(q)$

El modelo de media móvil de orden  $q$ ,  $MA(q)$  se define como:

$$Z_t = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q} \quad (17)$$

donde  $\epsilon_t \sim N(0, \sigma^2)$  es una serie de ruido blanco, donde  $\sigma^2$  es desconocido.

Observar, que (17) implica que  $E(Z_t) = 0$ . Si  $E(Z_t) = \mu$ , el modelo incluirá una constante  $\mu$ .

Considerando el operador retardo  $L$ , la ecuación que resulta es:

$$Z_t = \theta(L)\epsilon_t \quad (18)$$

donde:

$$\theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q \quad (19)$$

Es importante decir que los conceptos el “modelo  $MA(q)$ ” y el “proceso  $MA(q)$ ” pueden ser intercambiados. Sin embargo, vale precisar que un proceso estocástico  $\{Z_t\}$  es  $MA(q)$  si  $Z_t$ ,  $t = 0, \pm 1, \pm 2, \dots$ , puede escribirse como (17). Por otro lado, una realización  $Z_t$ ,  $t = 1, \dots, T$ , del proceso estacionario  $\{Z_t\}$  sigue el modelo  $MA(q)$ , es decir,  $Z_t \sim MA(q)$ ,  $t = 1, \dots, T$ , si el proceso  $\{Z_t\}$  es  $MA(q)$ .

Los modelos de medias móviles,  $MA(q)$  son estacionarios, donde  $E(Z_t) = \mu$ ,  $Var(Z_t) = \gamma_0 = (1 + \theta^2_1 + \dots + \theta^2_q)\sigma^2$ ,

$$\gamma_1 = Cov(Z_t, Z_{t-1}) = \theta_2\theta_3 + \dots + \theta_{q-1}\theta_q\sigma^2,$$

$$\gamma_2 = Cov(Z_t, Z_{t-2}) = \theta_2\theta_4 + \dots + \theta_{q-2}\theta_q\sigma^2,$$

.....,

$$\gamma_q = Cov(Z_t, Z_{t-q}) = -\theta_q\sigma^2, \gamma_{q+1} = \gamma_{q+2} = \dots = 0.$$

Entonces, los coeficientes de autocorrelación  $\rho_k$  de una serie que es una ejecución de un proceso  $MA(q)$  están acotados (Uriel, Peiró, 2000); (Montgomery, 2015). Es decir, si el comportamiento de una serie puede ser descrito por un modelo, en el correlograma generalmente, se observará un patrón, donde las primeras  $q$  autocorrelaciones son diferentes de cero, y, a partir de la  $(q+1)$ -ésima autocorrelación, todos son iguales a cero. Cabe destacar que en aplicaciones prácticas no siempre el patrón del comportamiento de la serie observado coincide perfectamente con el patrón teórico. Esto es debido a que los estimadores  $\widehat{\gamma}_k$ , y, consecuentemente  $\widehat{\rho}_k$  son variables aleatorias, cuyas varianzas dependen del tamaño de la serie. Nosotros discutiremos este asunto más adelante.

En el caso particular de un modelo  $MA(1)$ ,  $Z_t = \epsilon_t - \theta_1\epsilon_{t-1}$ . Entonces  $Var(Z_t) = \gamma_0 = (1 + \theta^2_1)\sigma^2$ ,  $\gamma_1 = Cov(Z_t, Z_{t-1}) = -\theta_1\sigma^2$ ,  $\gamma_2 = \gamma_3 = \dots = 0$ ,  $\rho_1 = \frac{-\theta_1}{1+\theta^2_1}$ ,  $\rho_2 = \rho_3 = \dots = 0$ .

Siendo siempre estacionarios, los modelos  $MA(q)$  no siempre son invertibles. Por ejemplo, a un modelo  $MA(1)$  con autocorrelación  $\rho_1 = 0.4$ , corresponden dos modelos:  $Z_t = \epsilon_t + 0.5\epsilon_{t-1}$  y  $Z_t = \epsilon_t + 2\epsilon_{t-1}$ .

Por otro lado, para un modelo  $MA(1)$  se cumple:

$$Z_t = \epsilon_t - \theta_1\epsilon_{t-1} = \epsilon_t - \theta(Z_{t-1} + \theta\epsilon_{t-2}) = \epsilon_t - \theta Z_{t-1} - \theta^2\epsilon_{t-2} = \dots = \epsilon_t - \sum_{j=1}^{\infty} \theta^j Z_{t-j}$$

Entonces, si  $|\theta| > 1$  las observaciones más lejanas tienen más peso, lo que no tiene sentido. En esta ocasión el modelo no es invertible.

De forma general, el proceso  $MA(q)$  es invertible si él puede escribirse como un proceso autorregresivo infinito  $AR(\infty)$ , es decir  $\widetilde{Z}_t = \epsilon_t + \alpha_1 \widetilde{Z}_{t-1} + \alpha_2 \widetilde{Z}_{t-2} + \dots$ , donde  $\widetilde{Z}_t = Z_t - \mu$  y los coeficientes  $\alpha_1, \alpha_2, \dots$  convergen a cero, es decir,  $\sum_{j=1}^{\infty} |\alpha_j| < \infty$ .

Se puede demostrar Box y Jenkins (1976), que un proceso  $MA(q)$  es invertible si sólo si todas las raíces de la ecuación característica  $\theta(L) = 0$  están fuera del círculo unitario.

### 2.2.2.2 Modelo Autorregresivo $AR(p)$

El modelo autorregresivo de orden  $p$ ,  $AR(p)$  se define como:

$$Z_t - \mu = \phi_1(Z_{t-1} - \mu) + \dots + \phi_p(Z_{t-p} - \mu) + \epsilon_t, \quad (20)$$

donde  $\epsilon_t \sim N(0, \sigma^2)$  es un ruido blanco, con  $\sigma^2$  desconocido y  $\mu = E(Z_t)$ .

El proceso  $AR(p)$  puede expresarse de una forma alternativa, como

$$\Phi(L)\widetilde{Z}_t = \epsilon_t, \quad (21)$$

donde  $\widetilde{Z}_t = Z_t - \mu$  y  $\Phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$  es el polinomio característico del modelo.

En adelante utilizamos el supuesto, sin pérdida de generalidades, de que  $\mu = 0$ .

Un proceso autorregresivo  $AR(p)$  es estacionario si sólo si todas las raíces de la ecuación característica  $\Phi(L) = 0$  están fuera del círculo unitario. De forma alternativa, un proceso autorregresivo  $AR(p)$  es estacionario si sólo si él puede ser presentado como un proceso de medias móviles infinito, es decir,  $MA(\infty)$ ,  $Z_t = \epsilon_t + \beta_1 \widetilde{\epsilon}_{t-1} + \beta_2 \widetilde{\epsilon}_{t-2} + \dots$ , donde los coeficientes  $\beta_1, \beta_2, \dots$  convergen a cero, es decir,  $\sum_{j=1}^{\infty} |\beta_j| < \infty$  (Box, Jenkins, Reinsel y Ljung, 2016).

Fácilmente, esto puede ser observado notando que

$$Z_t = \Phi^{-1}(L)\epsilon_t = \frac{1}{\Phi(L)}\epsilon_t \quad (22)$$

En caso que el polinomio  $\Phi(L) = 0$  tenga raíces con módulo menor que el proceso  $\{Z_t\}$  no podría escribirse como un proceso de medias móviles,  $MA(\infty)$  con coeficientes  $\beta_1, \beta_2, \dots$  que convergen a cero.

Obviamente, los procesos autorregresivos son siempre invertibles, puesto que  $AR(p)$  es un caso particular de  $AR(\infty)$ .

Una propiedad interesante de modelos  $AR(p)$ , que facilita identificación de los mismos es que las autocorrelaciones y los coeficientes de  $AR(p)$  siguen un patrón particular, que puede ser expresado mediante las ecuaciones de Yule-Walker (Box, Jenkins, Reinsel y Ljung, 2016):

$$\rho_1 = \phi_1 + \phi_2 \rho_1 + \dots + \phi_p \rho_{p-1}$$

$$\rho_2 = \phi_1\rho_1 + \phi_2 + \dots + \phi_p\rho_{p-2}$$

..... (23)

$$\rho_p = \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \dots + \phi_p$$

Las ecuaciones se obtienen de la siguiente manera. Si al modelo se multiplica ambos miembros por  $Z_{t-r}$  y se toma esperanzas se obtiene que,

$$\gamma_\tau = \phi_1\gamma_{\tau-1} + \dots + \phi_p\gamma_{\tau-p} + E(\epsilon_t Z_{t-r}) \quad (24)$$

Específicamente, para  $\tau = 0$ , se obtiene:

$$\gamma_0 = \phi_1\gamma_1 + \dots + \phi_p\gamma_p + \sigma_\epsilon^2 \quad (25)$$

Para  $\tau > 0$ , se obtiene la siguiente ecuación:

$$\gamma_\tau = \phi_1\gamma_{\tau-1} + \dots + \phi_p\gamma_{\tau-p} \quad (26)$$

Dividiendo ambas partes por  $\gamma_0$  se obtienen las ecuaciones (23), que, de forma general, pueden ser escritos como

$$\rho_\tau = \phi_1\rho_{\tau-1} + \dots + \phi_p\rho_{\tau-p}, \tau = 1, \dots, p \quad (27)$$

De forma alternativa, las ecuaciones (23) pueden ser obtenidos mediante las ecuaciones,

$$\Phi(L)\rho_k = 0, \text{ donde } \rho_{-k} = \rho_k, k = 0, 1, 2, \dots \quad (28)$$

Resolviendo el sistema de ecuaciones (23), las autocorrelaciones  $\rho_0, \rho_1, \rho_2, \dots, \rho_p$  se obtienen cómo funciones de los coeficientes autorregresivos  $\phi_1, \phi_2, \dots, \phi_p, \tau = 1, \dots, p$ . Además, se obtiene, que

$$\gamma_0 = \sigma^2(1 - \phi_1\rho_1 + \phi_2\rho_2 + \dots + \phi_p\rho_p)^{-1} \quad (29)$$

Como ya hemos mencionado arriba, el patrón de autocorrelaciones de una serie puede ser utilizado para identificar el proceso que la generó. En el caso de modelos autorregresivos  $AR(p)$ , el patrón general está dado por (20). Por ejemplo, para identificar el modelo  $AR(1)$ , notemos que las autocorrelaciones en este caso tienen el siguiente patrón:  $\rho_k = \phi_1^k$ . En caso de  $AR(2)$ , el patrón es  $\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}$ . En caso de  $AR(p)$ ,  $p > 2$ , la identificación es más compleja, sin embargo, basándose en la ecuación general (20), se puede concluir, que contrariamente a los modelos  $MA(q)$ , cuyas autocorrelaciones  $\rho_k = 0$  si  $k > q$ , en el caso de  $AR(p)$ , las autocorrelaciones decaen a cero. Cabe notar, que en el caso de modelos autorregresivos es más difícil identificar el orden  $p$ , basándose solamente en autocorrelaciones observadas.

Para identificar los modelos  $AR(p)$ , generalmente se utilizan autocorrelaciones parciales, puesto que admite identificar el orden  $p$  de manera más exacta. La idea es, ajustar una serie de modelos autorregresivos,  $AR(s)$ , para  $s = 1, 2, \dots$ , para hallar el coeficiente de  $Y_{t-s}$ . Este coeficiente denota  $\phi_{ss}$  y se llama la autocorrelación parcial de

orden  $s$ . Para hallar  $\phi_{ss}$ , primero se obtienen las estimaciones para las autocorrelaciones  $\rho_1, \dots, \rho_s$ , a partir de las realizaciones  $Z_1, \dots, Z_T$  observadas. Enseguida, se utilizan las ecuaciones de Yule-Walker (23) para hallar  $\phi_{ss}$ . Entonces, para un proceso  $AR(p)$ ,  $\phi_{kk} = 0$  para  $k > p$ , implicando que autocorrelaciones parciales se cortan después del retraso  $p$  para un  $AR(p)$ . Cabe destacar que en caso del proceso  $MA(q)$ , las autocorrelaciones parciales decaen a cero, y por ende definir el orden  $q$  para un modelo de medias móviles, a partir de autocorrelaciones parciales es una tarea más difícil.

### 2.2.2.3 Modelo Autorregresivo de Medias Móviles $ARMA(p, q)$

El modelo autorregresivo de medias móviles de órdenes  $p$  y  $q$ , denotado como  $ARMA(p, q)$  se define de la forma siguiente:

$$Z_t - \phi_1 Z_{t-1} - \dots - \phi_p Z_{t-p} = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q} \quad (30)$$

donde  $\epsilon_t \sim N(0, \sigma^2)$  es un ruido blanco, con  $\sigma^2$  desconocido.

$$\Phi(L)Z_t = \Theta(L)\epsilon_t \quad (31)$$

La condición de estacionariedad en caso de modelos  $ARMA(p, q)$  es equivalente a la condición de estacionariedad de la parte autorregresiva, mientras que la condición de invertibilidad es equivalente a la condición de invertibilidad de la parte de medias móviles del modelo.

Identificación del modelo  $ARMA(p, q)$  es relativamente más difícil, comparando con los modelos autorregresivos y de medias móviles, puesto que tanto autocorrelaciones como autocorrelaciones parciales pueden demostrar decaimiento, que en algunos casos puede ser muy despacio (Montgomery, 2015). Sin embargo, en caso que los valores de  $p$  y  $q$  son pequeños, la identificación puede ser más sencilla. Por ejemplo, en el caso que  $Z_t \sim ARMA(p, q)$ , es decir  $Z_t = \phi_1 Z_{t-1} + \epsilon_t - \theta_1 \epsilon_{t-1}$ , se obtiene que:

$$\rho_1 = \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1+\theta_1^2-2\phi_1\theta_1} \quad \text{y} \quad \rho_k = \phi_1 \rho_{k-1} \quad (32)$$

### 2.2.3 Modelos para series no estacionarias

Como ya fue mencionado, los modelos  $ARMA(p, q)$  solamente pueden ser ajustados a una serie estacionaria. Sin embargo, en muchas aplicaciones prácticas las series presentan un comportamiento no estacionario. Por ejemplo, series económicas frecuentemente presentan una tendencia a lo largo del tiempo (ver Figuras 1 y 4), estacionalidad (Figura 2) o ambos. Típicamente, la falta de estacionariedad por presencia de una tendencia se observa en series económicas anuales. Las series económicas mensuales o trimestrales presentan adicionalmente efectos de estacionalidad.



### 2.2.3.1 Modelos Autorregresivos Integrados de Medias Móviles $ARIMA(p, d, q)$

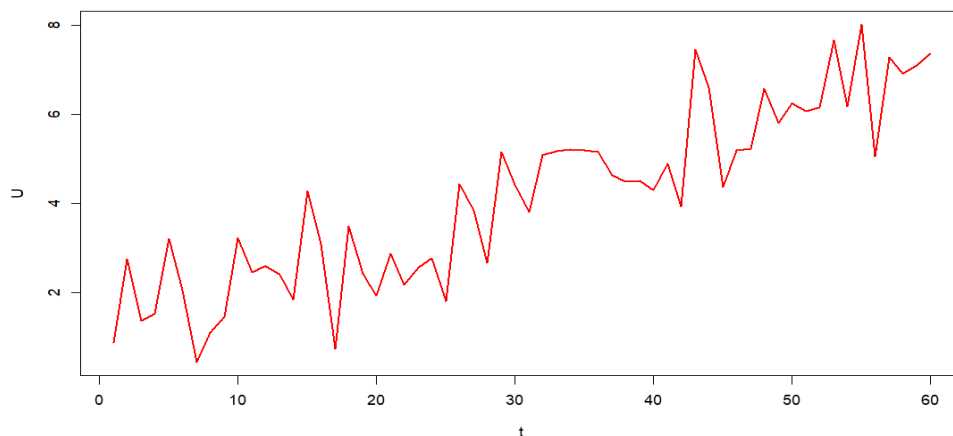
Los modelos  $ARIMA(p, d, q)$  se utilizan para implementar la metodología de Box y Jenkins (1976) a las series que presentan una tendencia y no presentan efectos de estacionalidad. Existen dos tipos de tendencia: la tendencia determinística y la tendencia estocástica. Una tendencia es determinística si es una función no aleatoria del tiempo; caso contrario la tendencia es estocástica. Por ejemplo, consideremos dos siguientes series:

1.  $U_t = a_0 + a_1t + \epsilon_t$ , donde  $\epsilon_t$  es un ruido blanco,  $\epsilon_t \sim N(0, \sigma^2)$ .
2.  $V_t = V_{t-1} + \epsilon_t$ , donde  $\epsilon_t$  es un ruido blanco,  $\epsilon_t \sim N(0, \sigma^2)$ .

La serie  $U_t$  tiene una tendencia determinística, mientras que la serie  $V_t$  tiene una tendencia estocástica (Figuras 6 y 7). En Novales (2016) se detalla el tema de tendencias determinísticas y estocásticas.

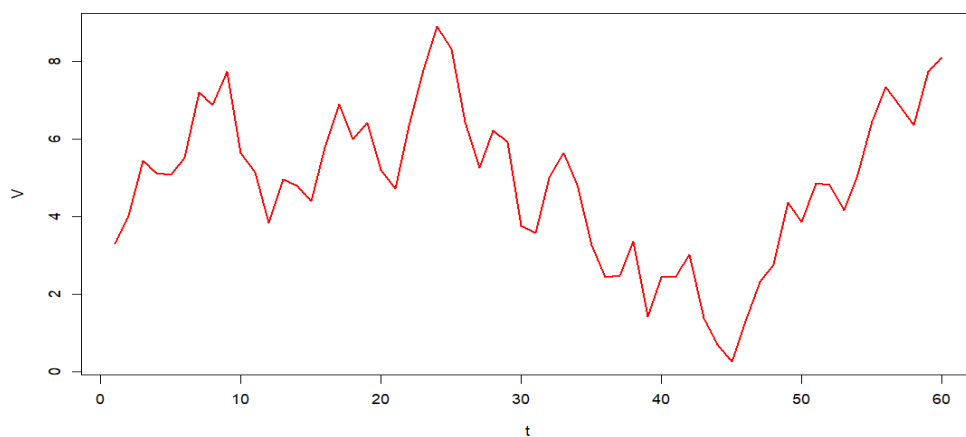
**Figura 6**

*Serie de ruido blanco con tendencia determinística*



**Figura 7**

*Serie de ruido blanco con tendencia estocástica*



Ahora, supongamos que una serie,  $Z_t$  sigue el siguiente modelo:  $Z_t = a_0 + a_1 t + V_t$ , donde  $\Phi(L)(1-L)V_t = \theta(L)\epsilon_t$ ,  $\epsilon_t \sim N(0, \sigma^2)$  es una serie de ruido blanco. La serie  $Z_t$  no es estacionaria puesto que tiene una tendencia lineal determinística, sin embargo, aplicación de la primera diferencia a la serie original  $Z_t$  resulta en una serie estacionaria:

$$\Delta Z_t = (1-L)Z_t = Z_t - Z_{t-1} = a_1 + (1-L)V_t, \quad (33)$$

tal que,

$$\Phi(L)(1-L)Z_t = c + \Phi(L)(1-L)V_t = c + \theta(L)\epsilon_t \quad (34)$$

La diferencia  $(1-L)$  se llama la diferencia regular. De forma general, si la tendencia de la serie original puede ser expresada por un polinomio de orden  $d$ , para obtener una serie estacionaria es necesario aplicar  $d$  diferencias regulares a la serie original (Box, Jenkins, Reinsel Ljung, 2016).

Cabe sobresalir que en el caso que la tendencia de una serie es de origen estocástico, aplicación de una transformación de diferencia regular con el orden  $d$  adecuado se utiliza para transformarla en una serie estacionaria (Box, Jenkins, Reinsel Ljung, 2016).

Entonces, el modelo autorregresivo de medias móviles de órdenes  $p$  y  $q$ , denotado como  $ARIMA(p, d, q)$  se define de la forma siguiente:

$$(1-L)^d \Phi_p(L)Z_t = \theta_q(L)\epsilon_t \quad (35)$$

donde  $\epsilon_t \sim N(0, \sigma^2)$  es un ruido blanco, con  $\sigma^2$  desconocido,  $p$  y  $q$  son las órdenes de los polinomios correspondientes (Box, Jenkins, Reinsel Ljung. 2016).

Es importante notar que en mayoría de aplicaciones prácticas se utiliza  $d = 1$  o  $d = 2$  para obtener una serie estacionaria. Para ajustar un modelo  $ARIMA(p, d, q)$  a una serie temporal, primero se identifica el valor de la diferencia regular, aplicación de la cuál produce una serie estacionaria,  $Y_t = (1-L)^d Z_t$ . El siguiente paso es ajustar el modelo  $ARMA(p, q)$  a la serie  $Y_t$ .

Como hemos mencionado arriba, los modelos  $ARIMA(p, d, q)$  son más útiles para analizar series anuales, puesto que estas series no poseen efectos de estacionalidad, como, por ejemplo, series trimestrales o mensuales. Para analizar este tipo de series se utiliza una extensión de los modelos  $ARIMA$ - los modelos  $SARIMA$ , que toman en cuenta los efectos de estacionalidad.

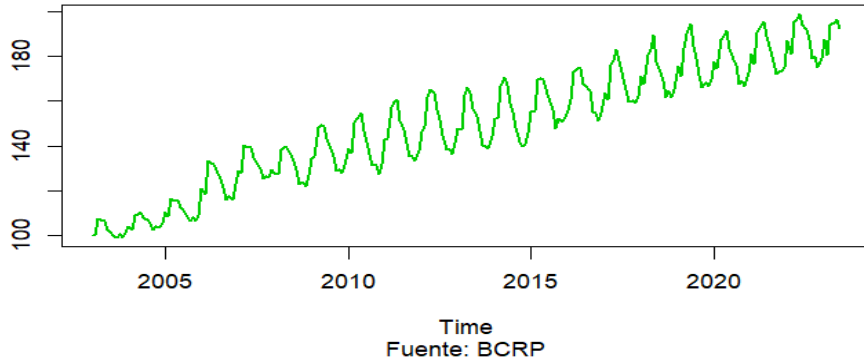
### 2.2.3.2 Modelo Autorregresivo Integrado de Medias Móviles Estacional $SARIMA(p, d, q)x(P, D, Q)_s$

Consideremos una serie mensual (Figura 8). Obviamente, además de una tendencia lineal se observa un comportamiento estacional, donde los picos corresponden a los meses marzo, abril y mayo.

**Figura 8**

*Serie con tendencia lineal y comportamiento estacional*

**Producción agropecuaria (miles de toneladas) - Agrícola - Leche**



Aparentemente, la transformación de diferencia regular  $(1 - L)^d$  no transformará dicha serie en una serie estacionaria, puesto que la periodicidad, observada en la serie original  $Z_t$  se mantendrá presente en la serie de  $(1 - L)^d Z_t$ . A fin de obtener una serie estacionaria en este caso es necesario aplicar una transformación de diferencia estacional del orden  $D$ , es decir  $(1 - L^S)^D Z_t$ , donde  $S$  es el número de periodos. Cabe sobresalir, que, en la realidad, aplicación de transformaciones  $(1 - L)^d$  y  $(1 - L^S)^D$ , siendo una herramienta recomendable y eficiente para resolver el problema de falta de estacionariedad, no garantiza la obtención de una serie estacionaria.

Entonces, el modelo estacional autorregresivo integrado de medias móviles de órdenes  $p, d, q, P, D$  y  $Q$  se define de la siguiente forma:

$$\Phi_p(L^S)\phi_p(L)(1 - L)^d(1 - L^S)^D Z_t = \theta_q(L)\theta_Q(L^S)\epsilon_t \quad (36)$$

donde  $\epsilon_t \sim N(0, \sigma^2)$  es un ruido blanco, con  $\sigma^2$  desconocido,  $d$  y  $D$  son los órdenes de diferencia regular y estacional,  $p, P, q$  y  $Q$  son las órdenes de los polinomios correspondientes (Box, Jenkins, Reinsel Ljung, 2016).

Los polinomios  $\phi_p(L)$  y  $\phi_q(L)$  representan la parte regular del modelo que expresa la dependencia entre las realizaciones en el tiempo  $t$  y los tiempos  $t-1, t-2, \dots$ , mientras que los polinomios  $\Phi_P(L)$  y  $\Phi_Q(L)$  representan la parte estacional del modelo que expresa la dependencia entre las realizaciones en el tiempo  $t$  y los tiempos  $t-s, t-2s, \dots$  (en el caso de una serie mensual,  $s=12$ ).

Obviamente,

$$\Phi_P(L^S) = 1 - \phi_1 L^S - \dots - \phi_P L^{Ps} \quad (37)$$

$$\theta_Q(L^S) = 1 - \theta_1 L^S - \dots - \theta_Q L^{QS} \quad (38)$$

Equivalentemente:

$$\Phi_p(L^S)\phi_p(L)W_t = \theta_q(L)\theta_Q(L^S)\epsilon_t \quad (39)$$

donde:

$$W_t = (1 - L)^d(1 - L^S)^D Z_t \quad (40)$$

Como en el caso de  $ARIMA(p, d, q)$ , para ajustar el modelo  $SARIMA(p, d, q)x(P, D, Q)_S$  a una serie temporal  $Z_t$ , primero se identifican los valores de la diferencia regular  $d$  y estacional  $D$ , tales que la serie  $W_t$  es estacionaria. Cabe notar que en mayoría de aplicaciones prácticas es suficiente aplicar  $D = 1$  o  $D = 2$  para obtener una serie estacionaria.

#### 2.2.4 Identificación del modelo

Como ya hemos mencionado, a cada modelo  $ARMA(p, q)$ , corresponde su patrón particular de las autocorrelaciones regulares y parciales, y, por ende, estas pueden ser utilizadas para identificar el modelo o los modelos posibles que describen el comportamiento de la serie y pueden ser utilizados para inferencia, principalmente sobre el comportamiento futuro de la serie.

En esta sección suponemos que la serie  $Z_t$  es estacionaria o fue obtenida de una serie no estacionaria,  $\tilde{Z}_t$  mediante aplicación de transformación de logaritmo (caso la varianza de  $\tilde{Z}_t$  no es constante a lo largo del tiempo), transformación de diferencia regular (caso la serie presenta una tendencia) y transformación de diferencia estacional (caso la serie presenta un comportamiento estacional).

El siguiente paso es identificar las autocorrelaciones  $\rho_k$  y autocorrelaciones parciales  $\phi_{kk}$ ,  $k = 1, 2, \dots$  significativas de la serie analizada.

Box y Jenkins (1970) demuestran que

$$\hat{\rho}_k \sim N(\rho_k, C_{kk}), \text{ si } T \rightarrow \infty \quad (41)$$

donde  $T$  es el tamaño de la serie,  $\hat{\rho}_k$  es el estimador para  $\rho_k$  que se obtiene utilizando (41),

$$C_{kk} \approx \frac{1}{T} \sum_{i=-\infty}^{\infty} (\rho_{k+i}^2 + \rho_{i+k}\rho_{i-k} - 4\rho_k\rho_i\rho_{i-k} + 2\rho_k^2\rho_i^2) \quad (42)$$

Obviamente, para un proceso de medias móviles,  $MA(q)$ , si  $k > q$ ,

$$C_{kk} \approx \frac{1}{T} (1 + 2 \sum_{i=1}^q \rho_i^2) \quad (43)$$

Entonces, para contrastar  $H_0: \rho_k = 0$ , con el nivel de la prueba,  $\alpha = 0.05$  se utilizan los valores críticos  $\pm 1.96\sqrt{\hat{C}_{kk}}$ , donde  $\hat{C}_{kk}$  se obtiene sustituyendo los  $\rho_j$  en (43) por sus estimadores correspondientes  $\hat{\rho}_j$ . Es decir, se rechaza  $H_0$  si  $|\widehat{\rho}_k| > 1.96\sqrt{\hat{C}_{kk}}$ .

Además, Box y Jenkins (1970) demuestran que para un modelo  $AR(p)$ , para  $k > p$

$$\hat{\phi}_{kk} \sim N(\phi_k, T^{-1}), \text{ si } T \rightarrow \infty \quad (44)$$

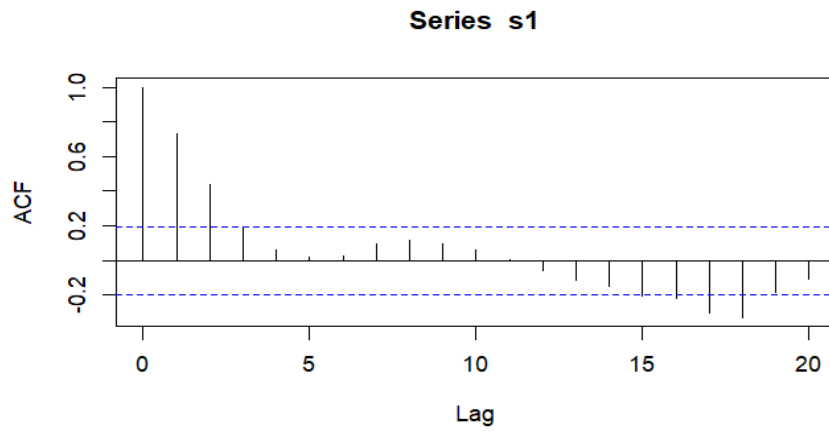
Entonces, para contrastar  $H_0: \phi_k = 0$ , con el nivel de la prueba,  $\alpha = 0.05$  se utilizan los valores críticos  $\pm \frac{1.96}{\sqrt{T}}$ .

Primero identifiquemos, los órdenes  $p$  y  $q$  de los polinomios de la parte regular (no estacional) del modelo, basándose en las autocorrelaciones regulares y parciales. Anderson (1974) sugiere los siguientes criterios. Caso se observa que  $\hat{\rho}_k \approx N(0, C_{kk})$ , donde  $C_{kk}$  se obtiene como (42), para  $k > q$ , es decir, no se rechaza  $H_0: \rho_k = 0$  para  $k > q$ , el modelo  $MA(q)$  puede ser el modelo apropiado. Por otro lado, si el comportamiento de las autocorrelaciones parciales,  $\hat{\phi}_{kk}$  demuestra que  $\hat{\phi}_{kk} \sim N(0, T^{-1})$ , para  $k > p$ , es decir, no se rechaza  $H_0: \phi_{kk} = 0$  para  $k > p$ , el modelo  $AR(p)$  puede ser un modelo candidato razonable. Si ninguno de los dos patrones se observa, deben ser considerados los modelos  $ARMA$ . Box y Jenkins (1970) sugieren que los valores de  $p$  y  $q$  más razonables son aquellos que cumplen  $p + q \leq 2$ . Esto implica que los modelos más razonables son:  $MA(1)$ ,  $MA(2)$ ,  $AR(1)$ ,  $AR(2)$  y  $ARMA(1,1)$ . Efectivamente, los modelos mencionados son los más comunes en muchas aplicaciones prácticas, donde se analizan series anuales.

En el siguiente ejemplo, se analiza una serie de tiempo de tamaño  $T = 100$ , cuyas correlogramas para  $\hat{\rho}_k$  y  $\hat{\phi}_{kk}$ ,  $k = 1, \dots, 20$  se presentan en las Figuras (9) y (10) (para  $k > 20$ , todos los valores de  $\hat{\rho}_k$  y  $\hat{\phi}_{kk}$  han salido no significativos). Además, los primeros seis valores de  $\hat{\rho}_k$  y  $\hat{\phi}_{kk}$  se observan en la Tabla 1. Observemos que las líneas azules punteadas en las Figuras (9) y (10) corresponden a los valores críticos para contrastar las hipótesis  $H_0: \rho_k = 0$  vs  $H_1: \rho_k \neq 0$  y  $H_0: \phi_{kk} = 0$  vs  $H_1: \phi_{kk} \neq 0$ , respectivamente, con el nivel  $\alpha = 0.05$ .

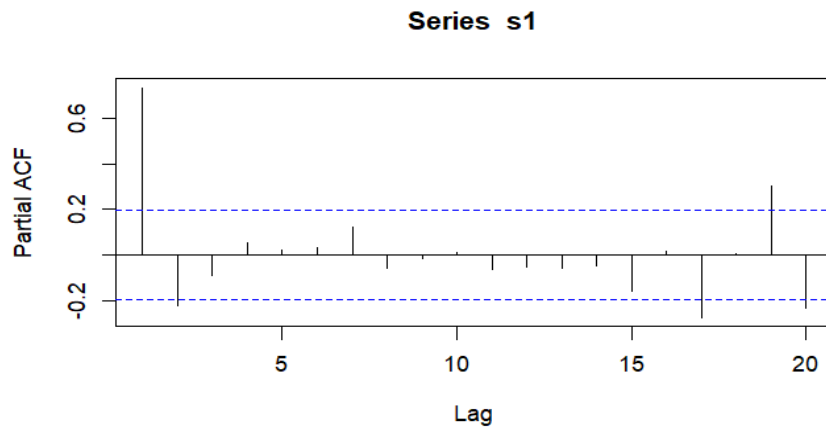
**Figura 9**

Correlograma de la función de autocorrelación de la Serie s1



**Figura 10**

Correlograma de la función de autocorrelación parcial de la Serie s1



**Tabla 1**

Valores críticos de la ACF y PACF

$k$	1	2	3	4	5	6
$\hat{\rho}_k$	0.602	0.345	0.186	0.044	-0.062	-0.100
$\hat{\phi}_{kk}$	0.602	-0.028	-0.018	-0.086	-0.073	-0.016

En este ejemplo, se puede observar que si  $k > 3$ ,  $\hat{\rho}_k \approx 0$ . Esto indica que el modelo  $MA(3)$  podría ajustarse bien a la serie. Por otro lado, si  $k > 1$ ,  $\hat{\phi}_{kk} \approx 0$ , lo que puede indicar que el modelo  $AR(1)$  es también un modelo razonable. Notemos también que las autocorrelaciones significativas cumplen  $\hat{\rho}_3/\hat{\rho}_2 \approx \hat{\rho}_2/\hat{\rho}_1$ , lo que es propio del modelo  $ARMA(1,1)$ . Estos tres modelos pueden ser evaluados como modelos candidatos en los siguientes pasos.

Cabe destacar que en nuestro ejemplo hemos obtenido que  $\hat{\rho}_k$  para  $k = 16, 17, 18$  y  $\hat{\phi}_{kk}$  para  $k = 17, 19, 20$  han salido significativos. Sin embargo, según Novales (2000), esto, probablemente, tiene una naturaleza aleatoria, debido a la varianza de los estimadores y error tipo I, lo que significa que algunas autocorrelaciones  $\hat{\rho}_k$  y  $\hat{\phi}_{kk}$  pueden salir significativas, mientras que los verdaderos valores correspondientes son iguales a cero.

En el caso de modelos *SARIMA*, es necesario analizar el patrón estacional de la serie, para definir los órdenes  $P$  y  $Q$  de los polinomios correspondientes. Para realizarlo, se evalúan las autocorrelaciones  $\hat{\rho}_k$  y  $\hat{\phi}_{kk}$ , para  $k = S, 2S, 3S, \dots$ , aplicando el mismo razonamiento como en el caso anterior. Además, en casos más sencillos es recomendable utilizar los patrones conocidos de los modelos más utilizados. Por ejemplo, uno de los modelos *SARIMA* comunes es el  $SARIMA(0,1,1)(0,1,1)_{12}$ . Entonces, si  $Z_t \sim SARIMA(0,1,1)(0,1,1)_{12}$ , para obtener una serie estacionaria, es necesario aplicar una transformación de diferencia regular y de una diferencia estacional, es decir,  $\bar{Z}_t = (1 - L)(1 - L^{12})Z_t \sim SARIMA(0,0,1)(0,0,1)_{12}$ . Es fácil demostrar que en caso una serie sigue el modelo  $SARIMA(0,0,1)(0,0,1)_{12}$ , solamente las autocorrelaciones  $\rho_1, \rho_{11}, \rho_{12}$  y  $\rho_{13}$  son distintas de cero, donde  $\rho_{11} = \rho_{13}$ .

## 2.2.5 Estimación de los parámetros

Para demostrar el método de estimación de los parámetros, se supone que  $Z_t \sim ARMA(p, q)$ , donde  $Z_t = (X_t - \mu)$ ,  $\mu = E(X_t)$  y  $X_t$  es una serie estacionaria. El proceso de estimación de los parámetros de los modelos *SARIMA* es idéntico. En esta sección presentamos la idea original, desarrollada en Ruey S. Tsay (2005).

De la ecuación (30) sigue:

$$\epsilon_t Z_t - \phi_1 Z_{t-1} - \dots - \phi_p Z_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}. \quad (45)$$

Utilizando el supuesto que los errores  $\epsilon_t$ ,  $t = 1, \dots, T$  presentan una serie de ruido blanco, y que además,  $\epsilon_t \sim N(0, \sigma^2)$ , se obtiene la función de verosimilitud como función de los parámetros  $\tilde{\phi} = (\phi_1, \dots, \phi_p)$ ,  $\tilde{\theta} = (\theta_1, \dots, \theta_q)$ ,  $\mu$  y  $\sigma^2$  y de los valores iniciales  $Z_{inicial} = (Z_0, \dots, Z_{-(p-1)})$  y  $\epsilon_{inicial} = (\epsilon_0, \dots, \epsilon_{-(q-1)})$ , de la siguiente forma:

$$L(Z_1, \dots, Z_T; \tilde{\phi}, \tilde{\theta}, \mu, \sigma^2, Z_{inicial}, \epsilon_{inicial}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^T e^{-\frac{1}{2\sigma^2} \sum_{t=1}^T \epsilon_t^2(\tilde{\phi}, \tilde{\theta}, \mu, \sigma^2, Z_1, \dots, Z_T, Z_{inicial}, \epsilon_{inicial})} \quad (46)$$

donde  $\epsilon_t(\tilde{\phi}, \tilde{\theta}, \mu, \sigma^2, Z_1, \dots, Z_T, Z_{inicial}, \epsilon_{inicial})$  se obtienen de forma recursiva, utilizando (45). Ruey S. Tsay (2005) sugieren utilizar  $Z_{inicial} = (\bar{Z}, \dots, \bar{Z})$  y  $\epsilon_{inicial} = (0, \dots, 0)$ .

Entonces, los estimadores para  $\tilde{\phi}, \tilde{\theta}, \mu$  y  $\sigma^2$  se obtienen maximizando la función de verosimilitud definida por (46). En este caso los estimadores obtenidos son estimadores de máxima verosimilitud condicional, puesto que en el procedimiento de estimación fueron utilizados valores iniciales.

Observamos que para  $t = p + 1, \dots, T$ , la ecuación (46) no incluye los valores de  $Z_{inicial}$ . Entonces, suponiendo que  $\epsilon_p = \epsilon_{p-1} \dots = \epsilon_{p-(q-1)} = 0$ , se puede definir la función de verosimilitud de forma alternativa como

$$L(Z_1, \dots, Z_T; \tilde{\phi}, \tilde{\theta}, \mu, \sigma^2, \epsilon_{inicial}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^T e^{-\frac{1}{2\sigma^2} \sum_{t=p+1}^T \epsilon_t^2(\tilde{\phi}, \tilde{\theta}, \mu, \sigma^2, Z_1, \dots, Z_T, \epsilon_{inicial})} \quad (47)$$

Este abordaje generalmente se utiliza si la serie analizada es suficientemente larga. Las varianzas para los estimadores para  $\tilde{\phi}, \tilde{\theta}, \mu$  y  $\sigma^2$  se obtienen utilizando la teoría asintótica de estimadores de máxima verosimilitud. Además, la inferencia sobre los parámetros  $\tilde{\phi}, \tilde{\theta}, \mu$  y  $\sigma^2$  se realiza basándose en la aproximación normal de la distribución asintótica de estimadores de máxima verosimilitud (Ruey S. Tsay, 2005).

En caso que el modelo ajustado es el modelo autorregresivo,  $AR(p)$ , se utilizan las ecuaciones de Yule-Walker (23), donde las autocorrelaciones  $\rho_k, k = 1, \dots, p$  son sustituidos por sus estimaciones respectivas usando (27).

Cabe notar que, en la etapa de estimación del modelo, generalmente, se evalúan varios modelos candidatos. Existen muchos criterios para elegir el mejor modelo. Los criterios más utilizados son el criterio de información de Akaike (AIC), explicado por Ender (2015) y el criterio de información bayesiano (BIC), desarrollado por Ender (2015).

Dado un modelo  $M$ , Denotamos por  $L^*$  el valor máximo que alcanza la función de verosimilitud, bajo el modelo  $M$ , por  $P^*$  el número de los parámetros del modelo y por  $T^*$  - el número de residuos estimados. Entonces:

el criterio AIC se define como

$$AIC = -2 \log (L^*) + 2P^* \quad (48)$$

el criterio BIC se define como

$$BIC = -2 \log (L^*) + P^* \log (T^*) \quad (49)$$

Además, se recomienda (Box, Jenkins, Reinsel Ljung, 2016) comparar predicciones bajo cada uno de los modelos, con los valores observados. Entonces, la comparación de los modelos puede ser realizada basándose en los errores de predicción de cada uno de los modelos de la siguiente forma.

1. Los modelos se ajustan a toda la serie,  $Z_t, t = 1, \dots, T$ , y bajo cada modelo  $M_i, i = 1, \dots, I$ , donde  $I$  es el número de los modelos candidatos, se calculan las predicciones para  $Z_t$  basándose en las observaciones  $Z_1, \dots, Z_{t-l}, \hat{Z}_{t|t-l}^{(M_i)}, t = t_0 +$



$1, \dots, T$ . Observen, que si  $l = 1$ , las predicciones se realizan a un paso (innovaciones), es decir, si estamos en el tiempo  $t = k$ , la predicción se realizará para el tiempo  $t = k + 1$ , utilizando las realizaciones de la serie, observados en los tiempos  $t = 1, \dots, k$ . Es importante precisar que  $t_0 + 1$  es el primer punto, donde es posible obtener la predicción. Obviamente,  $t_0$  depende del modelo ajustado. Los errores de predicción se obtienen de la siguiente manera:

$e_{t|t-k}^{(M_i)} = Z_t - \hat{Z}_{t|t-k}^{(M_i)}$ . Entonces, el error medio de las predicciones se obtiene como:  $B^{(M_i)} = \frac{1}{T-t_0} \sum_{t=t_0+1}^T e_{t|t-k}^{(M_i)}$ , el error medio absoluto:  $Abs^{(M_i)} = \sum_{t=1}^T |e_{t|t-k}^{(M_i)}|$  y el error cuadrático medio:  $ECM^{(M_i)} = \sum_{t=1}^T (e_{t|t-k}^{(M_i)})^2$ .

2. Los modelos candidatos se ajustan a la serie  $Z_1, \dots, Z_{T_0}$ , donde  $T_0 < T$ , y las predicciones se realizan para los tiempos  $t = T_0 + 1, T_0 + 2, \dots, T$ , bajo cada uno de los modelos. Luego, basándose en los verdaderos valores de la serie, se calculan el error medio, el error medio absoluto y el error cuadrático medio para cada modelo candidato.

En la siguiente sección explicamos con detalle, cómo se realizan predicciones usando los modelos *SARIMA*.

### 2.2.6 Predicciones bajo los modelos *SARIMA*.

En esta sección presentamos de forma general y concisa los resultados obtenidos por Box, Jenkins, Reinsel Ljung (2016).

Supongamos que la serie  $Z_t, t = 1, \dots, T$  es una serie estacionaria, tal que  $Z_t = (1-L)^d(1-L^S)^D X_t$ , donde  $X_t \sim SARIMA(p, d, q)(P, D, Q)_S$  y que  $EZ_t = 0$  (Si  $EZ_t = \mu \neq 0$  el proceso de la predicción es idéntico). Entonces, el modelo para  $Z_t$  puede ser presentado como

$$\Phi_p(B)\Phi_p(B^S)Z_t\theta_q(B)\theta_Q(B^S)\epsilon_t \quad (50)$$

De aquí obtenemos que

$$Z_t = [\Phi_p(B)\Phi_p(B^S)]^{-1}\theta_q(B)\theta_Q(B^S)\epsilon_t = \Psi(B)\epsilon_t, \quad (51)$$

es decir,  $Z_t = \epsilon_t + \psi_1\epsilon_{t-1} + \psi_2\epsilon_{t-2} + \dots$

Para predecir el valor de la serie en el tiempo  $t + l$  basándose en las observaciones  $Z_1, Z_2, \dots, Z_t$ , es decir predicción a  $l$  pasos, notemos que

$$Z_{t+l} = \epsilon_{t+l} + \psi_1\epsilon_{t+l-1} + \psi_2\epsilon_{t+l-2} + \dots + \psi_l\epsilon_t + \psi_{l+1}\epsilon_{t-1} + \dots \quad (52)$$

Puesto que los valores futuros de los errores generalmente se predicen como ceros, sigue que la predicción a  $l$  pasos es

$$\hat{Z}_{t+l|t} = \psi_l \epsilon_t + \psi_{l+1} \epsilon_{t-1} + \dots \quad (53)$$

Es importante resaltar que el predictor (53) minimiza el error cuadrático medio de la predicción, es decir,  $E(Z_{t+l} - \hat{Z}_{t+l|t})^2$ , lo que implica que

$$\hat{Z}_{t+l|t} = E(Z_{t+l}|Z_t, Z_{t-1}, Z_{t-2}, \dots, Z_1) \quad (54)$$

donde la esperanza condicional se calcula bajo el modelo estimado. Luego, el error de la predicción a  $l$  pasos se expresa como (Box, Jenkins, Reinsel Ljung, 2016)

$$e_t(l) = Z_{t+l} - \hat{Z}_{t+l|t} = \epsilon_{t+l} + \psi_1 \epsilon_{t+l-1} + \dots + \psi_{l-1} \epsilon_{t+1} \quad (55)$$

De (55) obtenemos que

$$e_t(1) = Z_{t+1} - \hat{Z}_{t+1|t} = \epsilon_{t+1} \quad (56)$$

El resultado (56) significa que los residuos del modelo  $\epsilon_t$  se obtienen como la diferencia entre el valor de la serie observado en el tiempo  $t$ , y la predicción para ese tiempo, realizado en el tiempo  $t - 1$ , es decir, basándose en los valores  $Z_1, \dots, Z_{t-1}$ . En realidad, puesto que los parámetros de los modelos, utilizados para realizar predicciones se estiman a partir de los datos, se obtienen los residuos estimados,  $\hat{\epsilon}_t$ . Estos residuos se utilizarán en la etapa de validación del modelo para contrastar la hipótesis que la serie  $\epsilon_t$  es una serie de ruido blanco.

Para ilustrar el método de predicción, presentado arriba, supongamos que  $Z_t \sim ARMA(1,1)$ , es decir,  $Z_t = \phi Z_{t-1} + \epsilon_t - \theta \epsilon_{t-1}$ . La predicción a un paso se obtiene de la siguiente forma:

$$\hat{Z}_{t+1|t} = E(\phi Z_t + \epsilon_{t+1} - \theta \epsilon_t | Z_t, Z_{t-1}, Z_{t-2}, \dots, Z_1) = \phi Z_t - \theta \hat{\epsilon}_t = \phi Z_t - \theta(Z_t - \hat{Z}_{t|t-1}) \quad (57)$$

Observemos, que la ecuación de la predicción obtenida es una ecuación recursiva, donde la predicción a un paso, realizado en el tiempo  $t$  depende de la predicción a un paso, realizada en el tiempo  $t - 1$ . Como una predicción inicial se puede utilizar  $\hat{Z}_{1|0} = Z_1$ . Enseguida, obtenemos que  $\hat{Z}_{2|1} = \phi Z_1$ ,  $\hat{Z}_{3|2} = \phi Z_2 - \theta(Z_2 - \phi Z_1)$ ,  $\dots$

De forma parecida, la predicción a dos pasos se obtiene como:

$$\hat{Z}_{t+2|t} = E(\phi Z_{t+1} + \epsilon_{t+2} - \theta \epsilon_{t+1} | Z_t, Z_{t-1}, Z_{t-2}, \dots, Z_1) = \phi E(Z_{t+1} | Z_t, Z_{t-1}, Z_{t-2}, \dots, Z_1) = \phi Z_{t+1|t}$$

De forma general, la predicción a  $l > 2$  pasos se obtiene como:

$$\hat{Z}_{t+l|t} = E(\phi Z_{t+l} + \epsilon_{t+l} - \theta \epsilon_{t+l-1} | Z_t, Z_{t-1}, Z_{t-2}, \dots, Z_1) = \phi E(Z_{t+l-1} | Z_t, Z_{t-1}, Z_{t-2}, \dots, Z_1) = \phi^2 E(Z_{t+l-2} | Z_t, Z_{t-1}, Z_{t-2}, \dots, Z_1) = \dots = \phi^{l-1} E(Z_{t+1} | Z_t, Z_{t-1}, Z_{t-2}, \dots, Z_1) = \phi^{l-1} Z_{t+1|t}$$

Obviamente, si  $Z_t = (1 - L)X_t = X_t - X_{t-1}$ , el valor de la serie original  $X_{t+1}$  puede ser predicho como:  $\hat{X}_{t+1|t} = X_t + \hat{Z}_{t+1|t}$ . Si  $Z_t = (1 - L)^d X_t$  o  $Z_t = (1 - L)^d (1 - L^s)^D X_t$ , el proceso de la predicción es muy parecido.

### 2.2.7 Validación del modelo

La validación del modelo ajustado,  $M$  se basa principalmente en análisis de los errores, las estimaciones de los cuales se obtienen implementando (Box, Jenkins, Reinsel Ljung. 2016). Básicamente, los supuestos de los modelos *SARIMA* son:

1. La serie de los errores,  $\epsilon_t, t = 1, \dots, T$  es una serie de ruido blanco;
2. Los errores  $\epsilon_t, t = 1, \dots, T$  siguen una distribución normal, es decir,  $\epsilon_t \sim N(0, \sigma^2)$ .

Para demostrar que una serie es ruido blanco, se puede contrastar si las autocorrelaciones individuales son iguales a cero, es decir,  $H_0: \rho_k = 0$  vs  $H_1: \rho_k \neq 0, k = 1, \dots, K$ . para  $K$  suficientemente grande. La desventaja de este método es que el contraste de varias hipótesis de forma simultánea, aumenta considerablemente el error tipo I. Para contrastar la hipótesis,  $H_0: \rho_1 = \rho_2 = \dots = \rho_K = 0$  con el nivel  $\alpha = 0.05$ , se usa la prueba de Ljung-Box (Box, Jenkins, Reinsel Ljung, 2016), cuyo estadístico se define de la siguiente forma

$$Q = T(T + 2) \sum_{k=1}^K \frac{\hat{\rho}_k^2}{T-k} \quad (58)$$

Bajo  $H_0$ , el estadístico  $Q$  tiene una distribución asintótica,  $\chi^2$  con  $K - P *$  grados de libertad, donde  $P *$  es el número de los parámetros desconocidos en el modelo ajustado,  $M$ . Obviamente, caso  $H_0$  se rechace, por lo menos una de las autocorrelaciones es distinta de cero.

Para contrastar la normalidad de los errores, basándose en los errores estimados (Box, Jenkins, Reinsel Ljung, 2016) se utilizan tantas herramientas gráficas (principalmente, P-P plot y Q-Q plot) cómo pruebas paramétricas y semi paramétricas. Las pruebas de Shapiro-Wilk (Shapiro y Wilk, 1965) y de Jarque-Bera (Jarque y Bera, 1987) son comúnmente utilizadas para contrastar el supuesto de la normalidad. Cabe destacar que las pruebas clásicas, tales como la prueba de Kolmogorov-Smirnov, Darling-Anderson etc., no son adaptables en este caso puesto que los mismos suponen que los parámetros del modelo ajustado son conocidos. En el caso de los modelos *SARIMA*, los parámetros se estiman a partir de los datos, y por ende las distribuciones asintóticas de los estadísticos de las pruebas mencionadas no pueden ser utilizadas.

### 2.3 Modelos con variables exógenas

Supongamos que además de la serie analizada  $Z_t, t = 1, \dots, T$ , se observan simultáneamente los valores de una serie  $X_t, t = 1, \dots, T$ , que proporciona información relevante sobre la serie analizada. La serie  $X_t$  en este caso se denomina la serie de variables exógenas. En esta sección suponemos que la dimensión de las variables

exógenas es igual a uno (los casos generales, donde  $\dim(X_t) = p > 1$  se analizan de forma idéntica). Además, por simplicidad, en esta sección suponemos, que las partes estocásticas de los modelos siguen un modelo *ARIMA*. Obviamente, la metodología presentada puede fácilmente ser extendida a los modelos *SARIMA*.

### 2.3.1 El modelo de regresión con errores autorregresivos.

Existen varias formas de incorporar  $X_t$  en el modelo. Uno de los abordajes más simples es el modelo de regresión con errores autorregresivos:

$$Z_t = \beta_0 + \beta_1 X_t + \epsilon_t \quad (59)$$

tal que

$$\epsilon_t = \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \dots + \phi_p \epsilon_{t-p} + \eta_t, \eta_t \sim N(0, \tau^2) \quad (60)$$

donde  $\eta_t$  es una serie de ruido blanco. El supuesto importante del modelo definido es que  $Z_t$  y  $X_t$  son estacionarias.

El modelo (59) puede ser escrito como

$$Z_t = \beta_0 + \beta_1 X_t + \Phi^{-1}(L)\eta_t \quad (61)$$

donde se supone que  $\Phi^{-1}(L)$  existe, es decir, el modelo (61) es invertible.

Ahora, definamos:

$$\tilde{Z}_t = \Phi(L)Z_t, \tilde{X}_t = \Phi(L)X_t \text{ y } \tilde{\beta}_0 = \Phi(L)\beta_0 = (1 - \phi_1 - \phi_2 - \dots - \phi_p)\beta_0$$

Entonces, (61) puede ser reescrito como

$$\tilde{Z}_t = \tilde{\beta}_0 + \beta_1 \tilde{X}_t + \eta_t \quad (62)$$

Observemos, que el coeficiente  $\beta_1$  es el mismo en los modelos (59) y (61).

Para estimar los parámetros del modelo, se utiliza el procedimiento, desarrollado por Cochrane y Orcutt (1949), que consiste en los siguientes pasos.

1. Estimar los parámetros  $\beta_0$  y  $\beta_1$ , basándose en el modelo (62).
2. Calcular los errores estimados:  $\hat{\epsilon}_t = Z_t - \hat{\beta}_0 - \hat{\beta}_1 X_t$ .
3. Estimar los parámetros  $\phi_i, i = 1, \dots, p$ , es decir,  $\hat{\phi}_i$  a partir de los errores, hallados en el paso anterior, utilizando las ecuaciones de Yule-Walker, definidos en (23). En este paso es recomendable analizar las autocorrelaciones de  $\hat{\epsilon}_t$ , para confirmar que ellos presentan un patrón de comportamiento autorregresivo.
4. Utilizando los estimadores  $\hat{\phi}_i, i = 1, \dots, p$ , calcular  $\tilde{Z}_t$  y  $\tilde{X}_t$ .
5. Estimar  $\tilde{\beta}_0$ , ajustando el modelo (62) a las series  $\tilde{Z}_t$  y  $\tilde{X}_t$  y, utilizando la igualdad  $\tilde{\beta}_0 = (1 - \phi_1 - \phi_2 - \dots - \phi_p)\beta_0$ , hallar el estimador para  $\beta_0$ , es decir,  $\hat{\beta}_0$  actualizado.
6. Repetir los pasos 2 y 3 para actualizar  $\hat{\phi}_i, i = 1, \dots, p$ .

7. Continuar los pasos 4,5 y 6 hasta alcanzar la convergencia.

Luego, basándose en el modelo resultante, hallar los  $\hat{\eta}_t$ , para realizar la validación del modelo implementado, es decir, contrastar la normalidad y las autocorrelaciones de los mismos (ver sección 2.2.2).

Predicción bajo el modelo de regresión lineal con errores autorregresivos consiste en dos pasos: predicción de la parte de regresión del modelo, y predicción de la parte autorregresivo de los errores, donde se implementa el mismo abordaje, descrito en la sección (2.2.2).

### 2.3.2 El modelo de regresión con errores ARIMA.

La definición del modelo de regresión con errores ARIMA es muy parecida con la definición del modelo de regresión con errores autorregresivos:

$$Z_t = \beta_0 + \beta_1 X_t + \epsilon_t \quad (63)$$

tal que

$$(1 - L)^d (1 - \phi_1 L - \dots - \phi_p L^p) \epsilon_t = (1 - \theta_1 L - \dots - \theta_q L^q) \eta_t \quad (64)$$

donde  $\eta_t$  es una serie de ruido blanco,  $\eta_t \sim N(0, \tau^2)$ .

Observemos que denotando:  $Z'_t = (1 - L)^d Z_t$ ,  $X'_t = (1 - L)^d X_t$  y  $\epsilon'_t = (1 - L)^d \epsilon_t$ , obtenemos que

$$Z'_t = \beta_1 X'_t + \epsilon'_t \quad (65)$$

Donde

$$(1 - \phi_1 L - \dots - \phi_p L^p) \epsilon'_t = (1 - \theta_1 L - \dots - \theta_q L^q) \eta_t \quad (66)$$

es decir,  $\epsilon'_t \sim ARMA(p, q)$ . Observemos que los errores del modelo, definido para las series originales,  $Z_t$  y  $X_t$ , siguen un modelo ARIMA, es decir,  $\epsilon_t \sim ARIMA(p, d, q)$ .

De forma alternativa, el modelo puede ser escrito como  $Z'_t = \beta_1 X'_t + \frac{\theta(L)}{\phi(L)} \eta_t$ , donde  $\phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$  y  $\theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q$ .

Si  $Z'_t$  y  $X'_t$  son estacionarias, se puede implementar el algoritmo, desarrollado en la sección 2.3.1, donde en el paso 3, los parámetros del modelo ARMA se estiman, maximizando la función de verosimilitud, definida en la sección (2.2.5).

Alternativamente, todos los parámetros del modelo, definido por (59) y (63) pueden ser estimados simultáneamente. Notando que dé (59) y (63) puede ser reescrito como

$$\begin{aligned} \eta_t = & (Z'_t - \beta_1 X'_t) - \phi_1 (Z'_{t-1} - \beta_1 X'_{t-1}) - \dots - \phi_p (Z'_{t-p} - \beta_1 X'_{t-p}) + \\ & + \theta_1 \eta_{t-1} + \dots + \theta_q \eta_{t-q} \end{aligned} \quad (67)$$

y que  $\eta_t$  es una serie de ruido blanco, se puede implementar el abordaje, presentado en la sección (2.2.3) para definir la función de verosimilitud, como una función de los parámetros desconocidos de ambos modelos.

Validación del modelo y predicción se realizan de la misma manera como en el caso de los errores autorregresivos.

### **2.3.3 Modelos SARIMAX.**

La percepción de los modelos *ARIMAX* es muy simple: en la ecuación (30) que define el modelo *ARMA*, se agrega una (o varias) variables exógenas con sus coeficientes respectivos:

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + \beta X_t + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q} \quad (68)$$

Por simplicidad utilizaremos una variable exógena. Evidentemente, el caso de varias variables exógenas es idéntico. El modelo (68) supone que ambas series,  $Z_t$  y  $X_t$  son estacionarias. Como en los casos anteriores si  $Z_t$  y  $X_t$  no son estacionarias, se aplican transformaciones apropiadas.

El modelo (68) puede ser presentado como

$$Z_t = \frac{\beta X_t}{\phi(B)} + \frac{\theta(B)}{\phi(B)} \epsilon_t \quad (69)$$

Estimación de los parámetros desconocidos, predicción de valores futuros y validación del modelo se realiza de igual manera como en el caso de los modelos *ARMA*.

Es importante resaltar que el modelo definido por (69) es un modelo *ARMAX*, puesto que  $Z_t$  y  $X_t$  son estacionarias. Si  $Z_t$  y  $X_t$  se obtienen utilizando transformaciones de diferencias regulares de las series originales  $Z'_t$  y  $X'_t$ , el modelo correspondiente en este caso es *ARIMAX*. Como ya fue mencionado, incorporación en dicho modelo autocorrelaciones y diferencias estacionales resultará en un modelo *SARIMAX*.

### **2.3.4 Modelos de Función de Transferencia.**

En esta sección describimos de forma general los modelos de función de transferencia, que incorporan el efecto de los tiempos pasados de  $X_t$ , es decir, para predecir el valor de la serie  $Z_t$ , además de las realizaciones de la misma serie en los tiempos anteriores, se utilizan las realizaciones pasadas de la serie  $X_t$ . Los modelos presentados en las secciones 2.3.1-2.3.3 constituyen casos particulares de los modelos de función de transferencia. Según la terminología de modelos dinámicos la serie  $X_t$  se llama la serie de entrada y la serie  $Z_t$  se llama la serie de salida, y el conjunto de ambas formas

un sistema dinámico. Un choque en la serie de entrada no causa un impacto inmediato en la serie de salida, sino existe un retraso hasta que la serie de salida sufre el impacto. Box, Jenkins y Reinsel (1994) introducen una clase modelos parsimoniosos de función de transferencia de la siguiente forma:

$$\delta(L)Z_t = \omega(L)X_{t-b} \quad (70)$$

o, de forma alternativa, como

$$Z_t = v(L)X_t \quad (71)$$

donde  $v(L) = L^b \delta^{-1}(L)\omega(L)$  es la función de transferencia y  $b$  es el retraso de la respuesta del sistema al choque en la serie de entrada.

Puesto que adicionalmente, el sistema puede sufrir choques aleatorios, se agrega un ruido  $N_t$ , lo que produce el modelo combinado de función de transferencia y ruido.

$$\delta(L)Z_t = \omega(L)X_{t-b} + N_t \quad (72)$$

donde  $N_t \sim ARMA(p, q)$ , es decir  $\Phi(L)N_t = \theta(L)\epsilon_t$ , donde  $\epsilon_t$  es una serie de ruido blanco. Se supone que  $N_t$  y  $X_t$  son independientes.

Cabe notar, que el modelo definido en (72) implica que el proceso que genera  $(X_t, Z_t)$  es estacionario. En caso contrario es necesario aplicar las transformaciones apropiadas.

Cabe resaltar, que generalmente, si  $X_t$  y  $Z_t$  son estacionarios, esto no implica que  $(X_t, Z_t)$  es estacionario. Sin embargo, en la literatura sobre los modelos de función de transferencia, se supone que si  $X_t$  y  $Z_t$  son estacionarios y fueron obtenidos como resultado de transformación de diferencia de otras series,  $(X_t, Z_t)$  también es estacionaria (Box, Jenkins y Reinsel, 1994).

#### 2.3.4.1 Correlaciones cruzadas

Ahora, vamos a definir de forma concisa los pasos para ajustar un modelo de función de transferencia, definido en la sección anterior. El primer paso es la identificación del modelo posible, utilizando la estructura de las autocorrelaciones y de correlaciones cruzadas. Las autocorrelaciones cruzadas se utilizan para identificar el valor de retraso  $b$ .

Primero, definamos los coeficientes de autocovarianzas cruzadas,  $\gamma_{xz}(j)$  y  $\gamma_{zx}(j)$ :

$$\gamma_{xz}(j) = E[(X_t - \mu_x)(Z_{t+j} - \mu_z)] \quad j = 0, 1, 2, \dots \quad (73)$$

y

$$\gamma_{zx}(j) = E[(Z_t - \mu_z)(X_{t+j} - \mu_x)] \quad j = 0, 1, 2, \dots \quad (74)$$

donde  $\mu_x = E(X_t)$  y  $\mu_z = E(Z_t)$ ,  $\forall t = 1, \dots, T$ .

Cabe resaltar que  $\gamma_{xz}(j) = \gamma_{zx}(-j)$ , y que generalmente,  $\gamma_{xz}(j) \neq \gamma_{zx}(j)$ .

Entonces, la función de autocovarianzas cruzadas es  $\gamma_{xz}(j)$ ,  $j = 0, \pm 1, \pm 2, \dots$

De forma semejante, se define la función de autocorrelaciones cruzadas,

$$\rho_{xz}(j) = \frac{\gamma_{xz}(j)}{\sigma_x \sigma_z}, \quad j = 0, \pm 1, \pm 2, \dots \quad (75)$$

donde  $\sigma_x^2$  y  $\sigma_z^2$  son las varianzas de  $X_t$  y  $Z_t$ , respectivamente.

Los estimadores para  $\gamma_{xz}(j)$ ,  $j = 0, \pm 1, \pm 2, \dots$  se obtienen de la siguiente forma

$$\hat{\gamma}_{xz}(j) = \frac{1}{n-k} \sum_{t=1}^{n-k} (x_t - \bar{x})(z_{t+k} - \bar{z}) \quad j = 0, 1, 2, \dots \quad (76)$$

y

$$\hat{\gamma}_{xz}(-j) = \frac{1}{n-k} \sum_{t=k+1}^n (x_t - \bar{x})(z_{t-k} - \bar{z}) \quad j = 0, -1, -2, \dots \quad (77)$$

donde  $n$  es el número de pares de las realizaciones  $(x_1, z_1), \dots, (x_n, z_n)$  observadas.

Las estimaciones para las autocorrelaciones cruzadas se obtienen como,

$$\hat{\rho}_{xz}(\pm j) = \frac{\hat{\gamma}_{xz}(\pm j)}{\hat{\sigma}_x \hat{\sigma}_z}, \quad j = 0, \pm 1, \pm 2, \dots \quad (78)$$

donde  $\hat{\sigma}_z = \hat{\gamma}_{zz}(0)$  y  $\hat{\sigma}_x = \hat{\gamma}_{xx}(0)$ ,  $\gamma_{xx}(j) = E[(X_t - \mu_x)(X_{t-j} - \mu_x)]$ ,  $\gamma_{zz}(j) = E[(Z_t - \mu_z)(Z_{t-j} - \mu_z)]$ , y  $\hat{\gamma}_{xx}$  y  $\hat{\gamma}_{zz}$  pueden ser estimados, utilizando la fórmula (4) en la sección 2.1.1.

Para contrastar las hipótesis  $H_0: \rho_{xz} = 0$  contra la alternativa  $H_0: \rho_{xz} \neq 0$ , se utiliza el resultado desarrollado en Bartlett (1955), donde se demuestra la normalidad asintótica de los estimadores para  $\rho_{xz}(\pm j)$ , y aproximación de la varianza de dichos estimadores.

### 2.3.4.2 Identificación del modelo

Primero, observemos, que  $X_t = \frac{\theta_X(L)}{\phi_X(L)} W_t$ ,  $N_t = \frac{\theta_N(L)}{\phi_N(L)} v_t$  y  $v(L) = L^b \delta^{-1}(L) \omega(L)$ , donde  $W_t, v_t$  son series de ruidos blancos, y  $\theta_X(L)$ ,  $\theta_N(L)$  y  $\phi_X(L)$ ,  $\phi_N(L)$  son los polinomios que corresponden a las partes de medias móviles y autorregresivas de las series  $X_t$  y  $N_t$ , respectivamente. Entonces, definiendo

$$\tilde{Z}_t = \frac{\phi_X(L)}{\theta_X(L)} Z_t, \quad (79)$$

obtenemos:

$$\tilde{Z}_t = v(L) W_t + \frac{\phi_X(L)}{\theta_X(L)} N_t \quad (80)$$

El paso definido por (80) se llama preblanqueo, puesto que en dicha ecuación la serie de entrada es una serie de ruido blanco. Este paso es significativo en la identificación del modelo, puesto que las autocorrelaciones cruzadas se obtienen de forma muy simple



cuando una de las series es ruido blanco. Observemos, que  $v(L)$  puede ser escrito de la siguiente forma:

$$v(L) = \sum_{i=0}^{\infty} v_i L^i \quad (81)$$

Entonces, el algoritmo de identificación del modelo es el siguiente

1. Ajustar los modelos *ARIMA* a las series  $X_t$ , utilizando la metodología de Box y Jenkins, descrita en la Sección 2.2.3. En este paso se obtienen los estimadores de los coeficientes de los polinomios  $\Phi_X(L)$  y  $\theta_X(L)$ . Notemos, que los órdenes de dichos polinomios se obtienen en el paso de identificación y estimación del modelo. Además, empleando las fórmulas (52 y 53), se obtienen las predicciones para  $W_t$ , es decir  $\widehat{W}_t$ , que serán utilizados en el tercer paso del presente algoritmo.
2. Realizar el paso de preblanqueo, definido por la ecuación (80).
3. Calcular las autocorrelaciones cruzadas entre  $\tilde{Z}_t$  y  $W_t$ :

$$\gamma_{\tilde{z},w}(j) = E(\sum_{i=0}^{\infty} v_j W_{t+j-i} W_t) = \sigma_w^2 v_j \quad (82)$$

donde  $\sigma_w^2$  es la varianza de la serie  $W_t$ , que puede ser estimada en el primer paso del presente algoritmo. Cabe notar que las estimaciones de las autocorrelaciones cruzadas se producen, basándose en las series  $\widehat{W}_t$ , y  $\hat{\tilde{Z}}_t$ , donde  $\widehat{W}_t$  se obtiene en el primer paso y  $\hat{\tilde{Z}}_t$  se deriva de (54), reemplazando los coeficientes de los polinomios  $\Phi_X(L)$  y  $\theta_X(L)$  con sus estimadores respectivos. Obviamente, los coeficientes  $v_j$  y el valor del retraso,  $b$  pueden ser estimados comparando las autocorrelaciones cruzadas en (76 y 77) con las estimaciones correspondientes.

4. Ajustar el modelo *ARIMA* a la serie  $\widehat{N}_t = Z_t - \hat{v}(L)X_t$ , donde  $\hat{v}(L)$  se obtiene reemplazando los coeficientes  $v_j$ ,  $j = 0, \dots, \infty$  por sus estimadores respectivos, obtenidos en el paso 3. Se utiliza la metodología de Box, Jenkins, Reinsel, Ljung (2016).

Es importante resaltar, que realización de predicciones futuros a  $l$  pasos es igual que en el caso univariado. Cabe notar que el método presentado es más utilizado caso las series de entrada y salida son univariados. En los casos más complejos, donde dichas series son vectores, generalmente se utilizan modelos de espacio de estado.

## **2.4 Modelo Vectorial Autorregresivo (VAR)**

### **2.4.1 Definición de un modelo VAR**

Sims (1980) introdujo el modelo Vectorial Autorregresivo (VAR, por sus siglas en inglés, Vector Autorregressive) en la literatura económica para mejorar las restricciones que se imponían en los modelos univariados. Un modelo VAR, que es una amplitud natural del modelo autorregresivo univariado, utiliza un conjunto de ecuaciones simultáneas para explicar el comportamiento dinámico de las series económicas y financieras, así como para predecir dichas variables. El modelo VAR ofrece pronósticos con frecuencia superiores a los de modelos univariados, por lo que es uno de los modelos más populares, adaptables y fáciles de usar para el análisis de series temporales multivariadas. Además, cuando existe certeza de sincronía entre un grupo de variables y que sus conexiones se transmiten a lo largo de un número predeterminado de observaciones anteriores, el modelo VAR es extremadamente útil. Este es el caso de la recaudación de ingresos corrientes (ingresos) y la ejecución de gastos (gastos). (ver Figura 11 y 12).

Cuando se quiere modelar y explicar las interacciones y los movimientos conjuntos entre un grupo de variables de series temporales, se utiliza el análisis multivariado de series temporales. Dado una serie de tiempo multivariada  $y_1, \dots, y_T$ , de dimensión  $k$ , es decir,  $y_t = (y_{1t}, \dots, y_{kt})'$ , decimos que la serie es originada por un proceso VAR( $p$ ) si satisface la ecuación

$$y_t = c + A_1 y_{t-1} + \dots + A_p y_{t-p} + e_t, \quad (83)$$

donde  $A_i$  para  $i = 1, 2, \dots, p$ , son matrices de coeficientes, cada una de orden  $k \times k$ ,  $c$  es un vector de constantes de orden  $k \times 1$  y  $e_t$  es un vector de orden  $k \times 1$  que representa un término de error no observable. Es usual asumir que  $e_t$  es un ruido blanco invariante en el tiempo con media cero y matriz de covarianza definida positiva  $E(e_t e_t') = \Sigma_e$ , es decir,  $e_t$  son vectores aleatorios con  $e_t \sim (0, \Sigma_e)$ .

El proceso VAR dado definido en (1.1) es estable si

$$\det(I_k - A_1 z - \dots - A_p z^p) \neq 0 \quad \text{para } |z| \leq 1 \quad (84)$$

La ecuación  $\det(I_k - A_1 z - \dots - A_p z^p) = 0$  es llamada ecuación característica del modelo. Entonces, de la ecuación (1.2), un modelo VAR es estable si las raíces de la ecuación característica están fuera del círculo unitario (para más detalle sobre estabilidad de un modelo VAR ver Lutkepohl (2005) [cap.2].)

Un modelo VAR con dos variables y  $p$  retardos es escrito de la siguiente expresión:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} a_{11}^1 & a_{12}^1 \\ a_{21}^1 & a_{22}^1 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \dots + \begin{bmatrix} a_{11}^p & a_{12}^p \\ a_{21}^p & a_{22}^p \end{bmatrix} \begin{bmatrix} y_{1,t-p} \\ y_{2,t-p} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix} \quad (85)$$

### 2.4.2 Proceso estacionario en covarianza

Un proceso estocástico es estacionario en covarianza o de segundo orden si su primer momento es constante y sus covarianzas son invariantes en el tiempo, es decir, el proceso  $\{y_t\}$  es estacionario si

$$E(y_t) = \mu \text{ para todo } t, E[(y_t - \mu)(y_{t-k} - \mu)'] = \Gamma(k) = \Gamma(-k)' \text{ para todo } t \text{ y } k \quad (86)$$

En el contexto de series de tiempo un proceso estacionario de segundo orden será llamado simplemente de proceso estacionario. Notemos que todo proceso VAR estable es estacionario (ver Lutkepohl (2005) [Cap.2, pág. 25]).

### 2.4.3 Autocorrelaciones de un proceso VAR

Sabemos que una serie temporal multivariada  $\{y_t\}$  es estacionaria en covarianza y estable si todas sus coordenadas (series temporales) son estacionarias y ergódicas (ver Lutkepohl (2005) [Cap. 2, pág. 2]). Entonces, el vector de medias es igual a

$$E(y_t) = \mu = (\mu_1, \dots, \mu_n)' \quad (87)$$

$$Var(y_t) = \Gamma_0 = E(y_t - \mu)(y_t - \mu)' \quad (88)$$

y la matriz de covarianza es igual a

$$Var(y_t) = \Gamma_0 = \begin{bmatrix} Var(y_{1t}) & Cov(y_{1t}, y_{2t}) & \cdots & Cov(y_{1t}, y_{nt}) \\ Cov(y_{2t}, y_{1t}) & Var(y_{2t}) & \cdots & Cov(y_{2t}, y_{nt}) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(y_{nt}, y_{1t}) & Cov(y_{nt}, y_{2t}) & \cdots & Var(y_{nt}) \end{bmatrix} \quad (89)$$

Finalmente, de lo anterior, la matriz de correlaciones del vector  $\{y_t\}$  es definida por

$$Corr(y_t) = R_0 = D^{-1}\Gamma_0 D^{-1} \quad (90)$$

donde  $D$  es una matriz diagonal con  $j$ -ésima elemento de la diagonal dado por  $\gamma_{jj}^0 = \sqrt{Var(y_{jt})}$

Los parámetros  $\mu$ ,  $\Gamma_0$  y  $R_0$  son estimados a partir de los datos usando los siguientes estimadores muestrales

$$\underline{y} = \frac{1}{T} \sum_{t=1}^T y_t \rightarrow E y = \mu \quad (91)$$

$$\hat{\Gamma}_0 = \frac{1}{T} \sum_{t=1}^T (y_t - \underline{y})(y_t - \underline{y})' \rightarrow Var(y_t) = \Gamma_0 \quad (92)$$

$$\hat{R}_0 = \hat{D}^{-1} \hat{\Gamma}_0 \hat{D}^{-1} \rightarrow Corr(y_t) = R_0 \quad (93)$$

donde  $\hat{D}$  es la matriz diagonal formado por las desviaciones muestrales en la diagonal. Los estimadores muestrales  $\underline{y}$ ,  $\hat{\Gamma}_0$  y  $\hat{R}_0$  son consistentes (ver Lutkepohl (2005) [Cap. 5, pág. 213]).

Las condiciones de estacionariedad son importantes para este trabajo, por tal motivo es necesario validar nuestro modelo VAR para las variables que estudiamos. De modo general, para ajustar un modelo VAR se debe de cumplir que las variables introducidas en el vector sean estacionarias, y para la identificación del modelo VAR se debe seleccionar el rezago o retardo óptimo para el modelo. Luego se hace la respectiva estimación de los parámetros, y finalmente, para el diagnóstico del modelo se deben analizar los residuos, para garantizar que hay homocedasticidad y ausencia de autocorrelación serial en los residuos.

A continuación, explicamos de forma concisa cada uno de los pasos usuales que se siguen cuando se ajusta un modelo VAR. Para más detalles se puede revisar ver Lutkepohl (2005), Enders (2014) y Hamilton (1994).

#### **2.4.4 Análisis de raíz unitaria**

Una raíz unitaria es una propiedad que pueden presentar las series de tiempo no estacionarias. Esta característica puede provocar dudas en la inferencia estadística de los modelos de series de tiempo, generando estimadores inconsistentes, y en los modelos vectoriales puede generar relaciones espurias entre las variables (para más detalles y ejemplos ver Hamilton (1994) y Enders (2014)).

Decimos que una serie de tiempo  $\{y_t\}$  tiene una raíz unitaria si el módulo de alguna raíz de la ecuación característica del proceso VAR es igual a 1, es decir,  $w$  es una raíz unitaria de  $\{y_t\}$  si es raíz de la ecuación  $\det(I_k - A_1z - \dots - A_pz^p) = 0$  y  $|w| = 1$ . Por lo tanto, un proceso con raíz unitaria no es estacionario. Si las demás raíces de la ecuación característica se encuentran fuera del círculo unitario, es decir, tienen un módulo mayor que uno, entonces la primera diferencia del proceso es estacionaria. Este tipo de procesos son denominados procesos integrados de orden 1, y son denotados por  $\{y_t\} \sim I(1)$ .

Testear la ausencia de raíces unitarias es una forma indirecta de analizar la estacionariedad de una serie de tiempo, pero recordemos que el principal objetivo del test de raíz unitaria es identificar la naturaleza de la tendencia de una serie. Para testear raíces unitarias generalmente se utiliza la prueba de Dickey-Fuller Aumentada (ADF por sus siglas en inglés), que pasamos a describir a continuación:

El test ADF surge cuando los residuos del test de DF presentan correlación serial. En ese caso, necesitamos agregar suficiente número de rezagos como regresores adicionales para poder transformar el residuo en ruido blanco. Para ello, el test ADF añade rezagos de la diferencia de la variable dependiente:

$$\Delta y_t = a_0 + a_2 t + \gamma y_{t-1} + \sum_{i=1}^p \beta_i \Delta y_{t-i} + e_t \quad (94)$$

La llamada prueba de raíz unitaria de Dickey-Fuller Aumentada es la prueba  $t$  para la hipótesis nula  $H_0 : \gamma = 0$ . Note que si no se rechaza  $H_0$  no necesariamente la conclusión es la presencia de raíz unitaria. Para obtener una mejor conclusión sobre el modelo Dickey y Fuller introducen dos nuevas pruebas, llamadas pruebas  $\phi$ . En la Tabla 1 se hace un resumen las pruebas de ADF para el modelo (94):

**Tabla 2**

*Test de Dickey y Fuller para el modelo (94)*

Modelo	Hipótesis: $H_0$	Test Estadístico
$\Delta y_t = a_0 + a_2 t + \gamma y_{t-1} + \sum_{i=2}^p \beta_i \Delta y_{t-i+1} + e_t$	$\gamma = 0$	$\tau_T$
	$\gamma = a_2 = 0$	$\phi_3$
	$a_0 = \gamma = a_2 = 0$	$\phi_2$
$\Delta y_t = a_0 + \gamma y_{t-1} + \sum_{i=2}^p \beta_i \Delta y_{t-i+1} + e_t$	$\gamma = 0$	$\tau_\mu$
	$a_0 = \gamma = 0$	$\phi_1$
$\Delta y_t = \gamma y_{t-1} + \sum_{i=2}^p \beta_i \Delta y_{t-i+1} + e_t$	$\gamma = 0$	$\tau$

Usando la Tabla 1 de forma adecuada podemos construir un procedimiento para identificar raíz unitaria y estacionariedad en una serie de tiempo (ver Enders (2014) [Cap. 4]).

## 2.4.5 Estimación de un modelo VAR

En la Sección 2.4.1 asumimos que tenemos una serie de tiempo multivariada  $\{y_t\}_1^T$  de dimensión  $k$ , con  $y_t = (y_{1t}, \dots, y_{kt})'$ , que se sabe que es generada por un proceso VAR( $p$ ), dado por (83). En la práctica, rara vez se sabe si la serie cumple las condiciones que requiere el modelo VAR( $p$ ), y sobre todo se desconoce el orden  $p$  del VAR. En la práctica, generalmente el orden suele ser desconocido. Por lo tanto, se deben utilizar herramientas estadísticas para comprobar de qué modelo VAR proviene de la serie de datos. En esta sección, discutimos de forma resumida algunas de estas herramientas.

Si se desconoce el verdadero orden del modelo VAR y sólo se conoce un límite superior, digamos  $M$ . Una posibilidad para verificar si ciertas matrices de coeficientes pueden ser cero, es establecer una prueba de significancia. Para nuestro problema particular de determinar el orden VAR correcto, podemos configurar una secuencia de pruebas. Primero se prueba  $H_0: A_M = 0$ . Si no se puede rechazar esta hipótesis nula, probamos  $H_0: A_{M-1} = 0$  y así sucesivamente hasta que podamos rechazar una hipótesis nula. Este procedimiento nos da una forma simple de estimar el orden de modelo.

Debido a que sólo necesitamos testear restricciones para los coeficientes de un modelo VAR, podemos usar el estadístico de Wald (ver Lutkepohl (2005) [Cap. 3]). Para arrojar algo más de luz sobre este tipo de estadística, puede ser instructivo considerar el principio de prueba de razón de verosimilitud. Se basa en comparar los máximos del logaritmo de la función de verosimilitud en el espacio de parámetros restringido y no restringido. Específicamente, la estadística de razón de verosimilitud es dado por

$$\lambda_{LR} = 2[\ln l(\hat{\delta}) - \ln l(\hat{\delta}_r)] \quad (95)$$

donde  $\hat{\delta}$  es el estimador de máxima verosimilitud sin restricciones para el vector de parámetros  $\delta$ , obtenido al maximizar la función de verosimilitud en todo el espacio de parámetros, y  $\hat{\delta}_r$  es el estimador de máxima verosimilitud restringido que se obtiene al maximizar la función de verosimilitud en la parte del espacio de parámetros donde se tiene las restricciones de interés. Teóricamente se puede demostrar que  $\lambda_{LR}$  tiene una distribución asintótica  $\chi_l^2$ , es decir,

$$\lambda_{LR} \rightarrow \chi_l^2 \quad (96)$$

donde  $l$  es igual al aumento en el número de retardos, multiplicado por el número de variables, elevado al cuadrado. La ecuación (96) es un teorema típico de los modelos de regresión lineal. (ver Lutkepohl (2005) [Cap. 3, Proposición 3.5] para su demostración).

Cuando se decide el número de retardos, también se debe tener demasiado en cuenta la exclusión de la autocorrelación serial en los residuos. El estadístico anterior no considera este aspecto principal y, por lo tanto, no debe usarse por sí solo. Por último, pero no menos importante, una buena estrategia es comenzar con menos retrasos y examinar las funciones de autocorrelación de los desechos junto con estadísticos del tipo Ljung-Box para comparar la probable existencia de autocorrelación. Esto requeriría una mayor cantidad de retrasos, lo que aumentaría la cantidad de parámetros estimados.

Además, debemos tener en cuenta que el número de retrasos no puede ser ni muy pequeño ni muy grande. Si el número de retrasos es muy pequeño, es posible que no se pueda capturar completamente la dinámica del sistema que se está analizando, mientras que, si es demasiado grande, se corre el riesgo de perder grados de libertad y tener que estimar un número de parámetros muy grande.

El análisis de los criterios de información, que son correcciones sobre el valor muestral de la función logaritmo de la verosimilitud, es una estrategia diferente para determinar el orden del modelo VAR. Los más notables son (a) AIC, el Criterio de Información de Akaike, (b) SC, el Criterio de Información de Schwarz, (c) HQ, el Criterio de Información de Hannan Quinn y (d) FPE Predicción Final del Error.

$$AIC = -2\frac{l}{T} + 2\frac{n}{T} \quad (97)$$

$$SBC = -2\frac{l}{T} + n\frac{\ln T}{T} \quad (98)$$

$$HQ = -2\frac{l}{T} + 2\frac{k\ln(\ln T)}{T} \quad (99)$$

$$FPE(k) = \frac{n+k}{n-k} \hat{\sigma}^2 \quad (100)$$

donde el logaritmo de la función de verosimilitud es  $l$ , el número de variables exógenas es  $d$ , el orden del VAR es  $p$  y el número de variables es  $k$ , y  $n = k(d + pk)$  el número de parámetros estimados en el modelo VAR. Según Lutkepohl (2005) [Cap. 4, Prop. 4.2 y Prop. 4.3] los criterios de información HQ y SC otorga estimaciones coherentes del verdadero orden del retardo del modelo, mientras que el AIC y FPE sobre estiman el orden del retardo del modelo con probabilidad positiva.

#### 2.4.6 Análisis de los Residuos del Modelo VAR

El análisis de los residuos es un paso importante en el análisis de series temporales, puesto que los modelos teóricos tienen ciertos supuestos sobre los residuos, como son normalidad, ausencia de correlación y homocedasticidad.

**a. Prueba de Normalidad.** Si los errores del modelo VAR no se distribuyen como una distribución normal indican que no es una buena muestra del proceso generador de los datos. Con ese fin, se utilizará la conocida prueba de Jarque & Bera (1980) para evaluar normalidad. La hipótesis del test de Jarque Bera es:

$H_0$ : Los residuos tienen una distribución normal

$H_1$ : Los residuos no tienen una distribución normal

**b. Prueba de Autocorrelación.** La prueba de la autocorrelación serial tiene un objetivo particularmente importante para verificar la existencia de correlación en los residuos hasta un orden determinado. Permite aclarar la especificación del modelo en términos del número óptimo de rezago. La prueba usada se denomina prueba de Multiplicadores de Lagrange (LM) o prueba de Breusch Godfrey. Si  $u_t$  representa los residuos, entonces suponemos que  $u_t = D_1u_{t-1} + D_2u_{t-2} + \dots + D_hu_{t-h} + v_t$ , donde  $v_t$  es un ruido blanco. La prueba de Breusch Godfrey plantea la Hipótesis  $H_0: D_1 = \dots = D_h = 0$ . El estadístico de la prueba será denotado por  $\lambda_{LM}$  (para más detalles sobre la forma del estadístico ver Lutkepohl (2005) [Cap. 4, Prop. 4.8]). El estadístico  $\lambda_{LM}$  tiene una distribución asintótica igual a  $\chi_{hk^2}^2$ , donde  $k$  el número de variables en la serie.

## 2.4.7 Análisis de Cointegración

El concepto estadístico de cointegración es requerido para dar sentido a los modelos de regresión y VAR con series  $I(1)$ . Un vector  $y_t = (y_{1t}, y_{2t}, \dots, y_{nt})$  de variables de naturaleza  $I(1)$  se dicen cointegradas si existe una combinación lineal de las mismas, definida por un vector  $\alpha = (\alpha_1, \dots, \alpha_n)$ , tal que  $\alpha'z$  es una variable aleatoria  $I(0)$ , es decir,

$$\alpha' y_t = \alpha_1 y_{1t} + \dots + \alpha_n y_{nt} \sim I(0) \quad (101)$$

Desde un panorama económico, la combinación lineal  $\alpha'z_t$  se puede relacionar con el concepto de equilibrio a largo plazo, y las diferencias entre ellas son estables aun cuando cada serie en particular abarque una tendencia estocástica y sea por lo tanto no estacionaria. Como se menciona en Granger (1981), "Al menos a un nivel perfeccionado de la teoría económica, existe la convicción que ciertos pares de variables económicas no deberían divergir mucho una de otra durante mucho tiempo, al menos a largo plazo". Entonces, el concepto de cointegración admite especificar modelos que captan parte de la relación a largo plazo.



Para contrastar cointegración, la prueba Johansen es actualmente la más utilizada. Dado el modelo VAR de orden  $p$ , para  $n$  variables:

$$y_t = c + A_1 y_{t-1} + \dots + A_p y_{t-p} + e_t \quad (102)$$

donde  $y_t = (y_{1t}, y_{2t}, \dots, y_{kt})$  y  $e_t$  es el vector de las innovaciones.

Sabemos que el modelo VAR( $p$ ) es estable si todas las raíces de  $\det(I_k - A_1 z - \dots - A_p z^p) = 0$  están fuera del círculo unitario. Si hay raíces sobre el círculo unitario entonces algunas o todas las variables en  $y_t$  son  $I(1)$  y ellas pueden estar también cointegradas (Lutkepohl, 2005). Si  $y_t$  está cointegrada entonces la representación VAR no es la representación más adecuada para nuestro análisis porque las relaciones de cointegración no son explícitamente aparentes. Entonces, necesitamos reescribir el modelo de tal forma que las relaciones de cointegración se hagan más aparentes. Para esto, el modelo VAR( $p$ ) anterior se puede escribir de la forma

$$\Delta y_t = \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + e_t \quad (103)$$

donde  $\Pi = \sum_{i=1}^p A_i - I_k$ ,  $\Gamma_i = -\sum_{j=i+1}^p A_j$ . El término  $\Pi y_{t-1}$  se conoce como *término de corrección del error*, y tiene un papel importante en el análisis de cointegración. La función anterior es conocido como *Modelo Vectorial de Corrección del Error* (VECM por sus abreviaturas en inglés). Del modelo corregido observamos que las relaciones de cointegración están contenidas en el término  $\Pi y_{t-1}$ . Denotemos por  $\Phi(z) = I_k - A_1 z - \dots - A_p z^p$ . Consideremos los siguientes tres casos:

1.  $Rango(\Pi) = 0$ : Esto implica  $\Pi = 0$  y las variables que componen el vector  $y_t$  no están cointegradas. El modelo VECM se reduce a un modelo VAR en primeras diferencias:

$$\Delta y_t = \Gamma_1 \Delta y_{t-1} + \Gamma_2 \Delta y_{t-2} + \dots + \Gamma_{p-1} \Delta y_{t-p+1} + e_t \quad (104)$$

2.  $Rango(\Pi) = n$ . Esto implica  $|\Phi(1)| \neq 0$  e  $y_t$  no contienen raíces unitarias, es decir,  $y_t$  es  $I(0)$ : El modelo VECM no es informativo, y examinamos directamente un modelo VAR en

$$\Delta y_t = \Gamma_1 \Delta y_{t-1} + \Gamma_2 \Delta y_{t-2} + \dots + \Gamma_p \Delta y_{t-p} + e_t \quad (105)$$

3. Si  $0 < Rango(\Pi) = k < n$ : Entonces, podemos escribir  $\Pi = \alpha \beta'$  donde  $\alpha$  y  $\beta$  son matrices  $n \times k$  con  $Rango(\alpha) = Rango(\beta) = k$ . El modelo VECM se convierte en:

$$\Delta y_t = \alpha \beta' y_{t-1} + \Gamma_1 \Delta y_{t-1} + \Gamma_2 \Delta y_{t-2} + \dots + \Gamma_{p-1} \Delta y_{t-p+1} + e_t \quad (106)$$

El vector  $y_t$  y sus variables están cointegradas, y  $k$  vectores de cointegración son linealmente independientes. Estos vectores son proporcionados por:  $w = \beta' y_t$  y tiene  $n - k$  raíces unitarias, que generan  $n - k$  tendencias estocásticas comunes a las variables que

componen  $y_t$ . Como resultado, un método para contrastar la cointegración es examinar el rango de la matriz  $\Pi$  de multiplicadores a largo plazo.

El enfoque de Johansen consiste en estimar el modelo VAR en función de la máxima verosimilitud y analizar el rango de la matriz  $\Pi$ . Johansen propone dos contrastes: el contraste del máximo autovalor  $\lambda_{max}$  y el contraste de la traza  $\lambda_{tr}$ .

Los contrastes se realizan paso a paso, de manera secuencial. La hipótesis nula para el contraste de traza es la siguiente:

$$H_0: \text{Rango}(\Pi) \leq k, \quad H_1: \text{Rango}(\Pi) > k \quad (107)$$

o, lo que es equivalente a la siguiente hipótesis nula:  $H_0$ : El vector  $y_t$  tienen un máximo de  $k$  relaciones de cointegración, frente a la hipótesis alternativa:  $H_1$ : El vector  $y_t$  tienen más de  $k$  relaciones de cointegración. El estadístico de prueba de la traza para contrastar  $H_0$  es:

$$\lambda_{tr}(k) = -(T - p) \sum_{i=k+1}^n \ln(1 - \hat{\lambda}_i) \quad (108)$$

donde  $\hat{\lambda}_i$  son los autovalores generalizados estimados para una definida matriz que aparece en el proceso de estimación por Máxima Verosimilitud. El leyente puede buscar detalles en Lutkepohl (2005) [Cap. 8, Sec. 8.2.1]. El estadístico del autovalor máximo es un poco más complejo de describir, y no es el objetivo de la tesis describir técnicamente dicha metodología. Usaremos ambos estadísticos para analizar nuestras variables.

#### 2.4.8 Análisis de la Función Impulso Respuesta

En muchos trabajos aplicados, frecuentemente el interés principal de ajustar un modelo VAR es conocer cómo reacciona o fluctúa una variable (respuesta) ante fluctuaciones de otra variable del sistema (impulso), en un sistema donde interactúan muchas variables. Este tipo de análisis ayuda a entender un poco la relación causal que puede existir entre dos variables, y ver como es el impacto del shock de una variable sobre otra. El análisis de *impulso-respuesta* está basada en la representación de Wold de un modelo VAR( $p$ ). Esta representación nos dice que todo modelo VAR( $p$ ) se puede escribir como una media móvil de orden infinito, es decir, si  $y_t \sim \text{VAR}(p)$  entonces

$$y_t = \mu + \sum_{k=0}^{\infty} \Psi_k \epsilon_{t-k} \quad (109)$$

Para efectos didácticos, veamos el caso VAR(1) bivariado. Un modelo VAR de dos variables tiene la forma

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} a_{10} \\ a_{20} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \quad (110)$$

Usando la definición de forma recursiva tenemos

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} e_{1t-i} \\ e_{2t-i} \end{bmatrix} \quad (111)$$

La ecuación anterior expresa las variables  $x_t$  e  $y_t$  en términos de las sucesiones  $\{e_{1t}\}$  y  $\{e_{2t}\}$ . Note que  $\{e_{1t}\}$  y  $\{e_{2t}\}$  no son independientes, por lo que es conveniente escribir la ecuación anterior en términos de dos ruidos independientes  $\{\epsilon_{1t}\}$  y  $\{\epsilon_{2t}\}$ , que corresponden a las innovaciones del proceso SVAR (para más detalle ver Enders (2014) [Cap. 7]). Sabemos que  $e_{1t} = B^{-1}\epsilon_t$ , es decir,

$$\begin{bmatrix} e_{1t-i} \\ e_{2t-i} \end{bmatrix} = \frac{1}{1-b_{12}b_{21}} \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} e_{1t-i} \\ e_{2t-i} \end{bmatrix} \quad (112)$$

Por lo tanto

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} + \frac{1}{1-b_{12}b_{21}} \sum_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} e_{1t-i} \\ e_{2t-i} \end{bmatrix} \quad (113)$$

Denotemos por

$$\phi_i = \begin{bmatrix} \phi_{11}(i) & \phi_{12}(i) \\ \phi_{21}(i) & \phi_{22}(i) \end{bmatrix} = \frac{1}{1-b_{12}b_{21}} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} e_{1t-i} \\ e_{2t-i} \end{bmatrix} \quad (114)$$

Entonces tenemos la representación en función de choques ortogonales

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} \phi_{11}(i) & \phi_{12}(i) \\ \phi_{21}(i) & \phi_{22}(i) \end{bmatrix} \begin{bmatrix} e_{1t-i} \\ e_{2t-i} \end{bmatrix} \quad (115)$$

Las cuatro funciones  $\phi_{11}(i)$ ,  $\phi_{21}(i)$ ,  $\phi_{12}(i)$ ,  $\phi_{22}(i)$  son denominadas en la literatura Funciones Impulso-Respuesta. Las entradas de las matrices  $\phi_i$  interpretan el impacto que tienen las dos variables del modelo  $x_t$  e  $y_t$  ante una perturbación en las innovaciones  $\epsilon_{1t}$  y  $\epsilon_{2t}$ . Graficar la función impulso-respuesta es una forma práctica para visualizar la respuesta de las series  $\{y_t\}$  y  $\{x_t\}$  ante impulsos o choques de otras variables.

#### 2.4.9 Predicción puntual e intervalo

El proceso de predicción puntual y por intervalo para los modelos VAR es similar a la metodología descrita para los modelos *ARIMA* y *SARIMA*. En esta sección damos un resumen de dicho procedimiento.

Para predecir el valor de la serie VAR en el tiempo  $t+l$  basándose en las observaciones  $y_1, y_2, \dots, y_t$ , es decir predicción a  $l$  pasos, notemos que

$$y_{t+l} = u_{t+l} + \psi_1 u_{t+l-1} + \psi_2 u_{t+l-2} + \dots + \psi_l u_t + \psi_{l+1} u_{t-1} + \dots \quad (116)$$

Puesto que los valores futuros de los errores generalmente se predicen como ceros, sigue que la predicción a  $l$  pasos es

$$\hat{y}_{t+l|t} = \psi_l u_t + \psi_{l+1} u_{t-1} + \dots \quad (117)$$

Es importante resaltar que el predictor (117) minimiza el error cuadrático medio de la predicción, es decir,  $E(y_{t+l} - \hat{y}_{t+l|t})^2$ , lo que implica que

$$\hat{y}_{t+l|t} = E(y_{t+l}|y_t, y_{t-1}, y_{t-2}, \dots) \quad (118)$$

donde la esperanza condicional se calcula bajo el modelo estimado. Luego, el error de la predicción a  $l$  pasos se expresa como (ver Lutkepohl, 2005)

$$e_t(l) = Z_{t+l} - \hat{Z}_{t+l|t} = u_{t+l} + \psi_1 u_{t+l-1} + \dots + \psi_{l-1} u_{t+1} \sim N(0, \Sigma_u(l)) \quad (119)$$

Este resultado implica que el error de predicción de las componentes individuales es normal y satisface

$$\frac{y_{k,t+l} - y_{k,t}(l)}{\sigma_k(l)} \sim N(0,1) \quad (120)$$

donde  $y_{k,t}(l)$  es la  $k$ -ésima componente de  $y_t(l)$  y  $\sigma_k(l)$  es la raíz cuadrada de la  $k$ -ésima diagonal de la matriz  $\Sigma_u(l)$ . Por lo tanto, un intervalo de confianza de  $(1 - \alpha)100\%$  para la predicción a  $l$  pasos en el futuro de la  $k$ -ésima componente es

$$y_{k,t}(l) \pm z_{\alpha/2} \sigma_k(l) \quad (121)$$

#### 2.4.10 Descomposición de la varianza del error de predicción.

Si se pueden identificar las innovaciones que realmente impulsan el sistema, se encuentra disponible una herramienta para interpretar los modelos VAR. Supongamos una identificación recursiva esquema está disponible para que la representación MA

$$y_t = \mu + \sum_{i=0}^{\infty} \Phi_i P P^{-1} u_{t-i} = \mu + \sum_{i=0}^{\infty} \theta_i w_{t-i} \quad (122)$$

Donde  $\theta_i = \Phi_i P$  y  $w_t = P^{-1} u_t$  es ruido blanco con matriz de covarianza  $\Sigma_w = P^{-1} \Sigma_u (P^{-1})'$  debido a que los errores de ruido blanco tienen componentes no correlacionados, son a menudo llamados residuos ortogonales o innovaciones.

En el contexto de la representación  $y_t = \mu + \sum_{i=0}^{\infty} \theta_i w_{t-i}$  con  $\Sigma_w = I_K$ , el error del pronóstico óptimo de paso  $h$  es

$$y_{t+h} - y_t(h) = \sum_{i=0}^{h-1} \Phi_i u_{t+h-i} = \sum_{i=0}^{h-1} \Phi_i P P^{-1} u_{t+h-i} = \sum_{i=0}^{h-1} \theta_i w_{t+h-i} \quad (123)$$

Denotando el elemento  $mn$ -ésimo de  $\theta_i$  por  $\theta_{mn,i}$  como antes, el error de pronóstico de paso  $h$  de la  $j$ -ésima componente de  $y_t$  es

$$y_{j,t+h} - y_{j,t}(h) = \sum_{i=0}^{h-1} (\theta_{j1,i} w_{1,t+h-i} + \dots + \theta_{jk,i} w_{K,t+h-i}) = \sum_{k=0}^K (\theta_{jk,0} w_{k,t+h} + \dots + \theta_{jk,h-1} w_{k,t+1}) \quad (124)$$

Por lo tanto, el error de pronóstico del componente j-ésima consiste potencialmente en todas las innovaciones  $w_{1t}, \dots, w_{Kt}$ . Por supuesto, algunos de los  $\theta_{mn,i}$  pueden ser cero, de modo que algunos componentes pueden no aparecer en la ecuación anterior. Porque los  $w_{k,t}$  no están correlacionados y tienen varianzas unitarias, el MSE de  $y_{j,t}(h)$  es

$$E \left( y_{j,t+h} - y_{j,t}(h) \right)^2 = \sum_{k=0}^K (\theta_{jk,0}^2 + \dots + \theta_{jk,h-1}^2) \quad (125)$$

Por lo tanto,

$$\theta_{jk,0}^2 + \theta_{jk,1}^2 + \dots + \theta_{jk,h-1}^2 = \sum_{i=0}^{h-1} (e_j' \theta_i e_k)^2 \quad (126)$$

A veces se interpreta como la contribución de las innovaciones en la variable k a la varianza del error de pronóstico o MSE del pronóstico de paso h de la variable j. Aquí  $e_k$  es la k-ésima columna de  $I_K$ . Dividiendo la ecuación anterior por

$$MSE[y_{j,t}(h)] = \sum_{i=0}^{h-1} \sum_{k=1}^K \theta_{jk,i}^2 \quad (127)$$

da

$$w_{jk,h} = \sum_{i=0}^{h-1} (e_j' \theta_i e_k)^2 / MSE [y_{j,t}(h)] \quad (128)$$

que es la proporción de la varianza del error de pronóstico del paso h de la variable j, contabilizada por innovaciones  $w_{k,t}$ . Si  $w_{k,t}$  se puede asociar con la variable k,  $w_{jk,h}$  representa la proporción de la varianza del error de pronóstico de paso h explicada por innovaciones en la variable k. Por tanto, la varianza del error de pronóstico se descompone en componentes explicados por las innovaciones en las diferentes variables del sistema. De la ecuación el error del pronóstico óptimo de paso h, se ve que la matriz MSE de pronóstico es

$$\Sigma_y(h) = MSE[y_t(h)] = \sum_{i=0}^{h-1} \theta_i \theta_i' = \sum_{i=0}^{h-1} \Phi_i \Sigma_u \Phi_i' \quad (129)$$

Los elementos diagonales de esta matriz son los *MSE* de las variables  $y_{jt}$  que puede usarse en la ecuación de la proporción de la varianza del error de pronóstico, (ver Lutkepohl (2005) [Cap. 2]).

## Capítulo III: Marco Metodológico

### 3.1 Diseño, tipo de la Investigación

#### 3.1.1 Tipo de investigación

La investigación es de tipo aplicada, pues tiene como objetivo aplicar teorías existentes a fin de resolver el problema de encontrar la mejor aproximación de un proceso económico.

#### 3.1.2 Nivel de investigación

El nivel de investigación del presente trabajo es predictivo y descriptivo, ya que predice y describe un modelo que expresa el comportamiento de la variable ingreso y gasto. Hernández et. al. (2014).

#### 3.1.3 Diseño de investigación

La presente investigación tiene un enfoque cuantitativo; el diseño es no experimental longitudinal; no experimental, pues se realiza sin manipulación deliberada de las variables y longitudinal pues según Hernández et.al. (2004) “Recaba datos en diferentes puntos en el tiempo para realizar inferencia acerca de la evolución del fenómeno” ya que en el estudio se analiza y evalúa los mismos datos (meses) por un periodo prolongado de tiempo del 2010 al 2018.

### 3.2. Método de la investigación

El método aplicado es de G.E.P. Box & G.C. Tiao (1975), basado en la metodología Box - Jenkins y el Modelo Vectorial Autorregresivo Sim's (1980), el primero es un método de predicción que se fundamenta en el análisis de las propiedades estocásticas de las series temporales de los ingresos y gastos, por la fuente de financiamiento, RDR ,de recursos directamente recaudados de la Universidad Nacional de Ingeniería por mes; y el segundo es una prolongación natural del modelo autorregresivo univariado, donde se incorporan un sistema de ecuaciones simultaneas para describir y predecir el comportamiento dinámico de las series recaudación de ingresos y ejecución de gastos de la UNI.

Pueden incluirse también como variable explicativa una variable, en valor contemporáneo que en este caso será los ingresos, que se considera exógena respecto a las variables que integran el modelo, es decir al gasto.

### **3.2.1. Población**

La recaudación de ingresos y ejecución de gastos mensuales, de la Universidad Nacional de Ingeniería.

### **3.2.2 Muestra**

Para el estudio de la recaudación de ingresos y ejecución de gastos de la Universidad Nacional de Ingeniería, se emplean los ingresos y gastos mensuales para el periodo enero de 2010 a diciembre del 2018

## **3.3 Definiciones Operativas**

### **3.3.1. Variables de Estudio**

Caracterización de las Variables

Los Recursos Directamente Recaudados; se entiende como los Ingresos generados por los centros de costos de la UNI, que no constituyen recursos ordinarios. Tienen su inicio en el esfuerzo propio de recaudación y proceden de los servicios que brindan; entre los cuales se pueden citar como: las tasas, la venta de bienes, la prestación de servicios y las rentas de la propiedad.

Entre las principales entidades generadoras de estos Ingresos Corrientes tenemos a la Dirección de Admisión, Centro de Infraestructura y Proyectos (CENIP), Dirección de Responsabilidad Social y desarrollo Sostenible, Unidad de Centro Pre-Universitario, Centro Peruano Japonés de Investigaciones Sísmicas y Mitigación de Desastres (CISMID) y la Unidad de Laboratorio Nacional de Hidráulica.

Asimismo, las Facultades de la Universidad Nacional de Ingeniería, conforman parte de la generación de ingresos y esto es atribuible a la calidad profesional de nuestros académicos, la confianza de la sociedad y del estado en la Universidad Nacional de Ingeniería, una apropiada enseñanza en relación a una práctica real y la magnitud de la inversión que permite su desarrollar integral.

La recaudación de Ingresos, recursos financieros que se obtienen de manera regular o periódico y la ejecución de Gastos de todas las Dependencias y Facultades que constituyen la Universidad Nacional de Ingeniería determinaron las variables de estudio del presente trabajo.

- $I_t$ : Ingresos de la UNI (En Soles Constantes) (Base dic. 2017 = 100)  
Por la Fuente de Financiamiento: Recursos Directamente Recaudados
- $G_t$ : Gastos de la UNI (En Soles Constantes) (Base dic. 2017 = 100)  
Por la Fuente de Financiamiento: Recursos Directamente Recaudados
- $t$ : Mes (desde enero del año 2010 a diciembre del año 2018) ( $t = 1, 2, \dots, 108$ )

### **3.3.2 Operacionalización de variables**

Los ingresos y gastos recopilados provienen de data secundaria y se encuentran en soles corrientes mensualizados, desde enero del año 2010 hasta diciembre del 2018, que luego de desarrollar un proceso deflactación se convierte en soles constantes.

## **3.4 Formulación de Hipótesis de Investigación**

### **3.4.1. Hipótesis General**

El modelo SARIMA, SARIMAX o VAR produce la mejor predicción de la ejecución de gastos por la fuente de financiamiento de recursos directamente recaudados, de la Universidad Nacional de Ingeniería en el periodo 2010 al 2018.

### **3.4.2. Hipótesis Específicas**

El modelo SARIMA predice la ejecución de gastos, por la fuente de financiamiento de recursos directamente recaudados, de la Universidad Nacional de Ingeniería en el periodo 2010 al 2018

El modelo SARIMAX predice la ejecución de gastos, por la fuente de financiamiento de recursos directamente recaudados, de la Universidad Nacional de Ingeniería en el periodo 2010 al 2018.

El Modelo VAR predice la ejecución de gastos, por la fuente de financiamiento de recursos directamente recaudados, de la Universidad Nacional de Ingeniería en el periodo 2010 al 2018.



### **3.5 Técnicas e instrumentos de recolección de datos**

La técnica para Córdova (2012) es un conjunto de pasos organizados con el fin de recolectar datos para conocer o medir una variable y los instrumentos son medios materiales que se emplean para recoger y almacenar la información.

La técnica de recolección de datos empleada en la presente investigación es el Registro histórico de los ingresos y gasto de la Universidad Nacional de Ingeniería y el instrumento empleado es la base de datos de la Unidad de Tesorería y Contabilidad de la Universidad Nacional de Ingeniería.

Cabe señalar que se usan las teorías de los modelos SARIMA, SARIMAX y VAR a fin de construir los modelos que permitan realizar los pronósticos respectivos.

### **3.6 Tratamiento estadístico de la información**

Diversos autores siempre desarrollan un trabajo previo con la serie de datos monetarios; esto es expresar en datos constantes, mediante el uso de deflatores.

En el Perú, el Instituto Nacional de Estadísticas e informática (INEI), mide la inflación mediante la variación porcentual del Índice de Precios al Consumidor (IPC), y está basado en la metodología de Laspeyres. El índice de Laspeyres puede explicarse como la relación entre el valor que se consigue al comprar a precios actuales la misma cantidad de artículos de bienes y servicios del periodo base, y el valor de esos mismos artículos en el periodo base.

En la presente investigación se emplea cifras conciliadas mensuales de la recaudación de ingresos (I) en soles corrientes y ejecución de gastos (G) en soles corrientes, desde enero de 2010 hasta diciembre del 2018, de la Universidad Nacional de Ingeniería. Los cuales pasan por un proceso de deflatación empleando los índices de precios al consumidor para eliminar el impacto de los precios; lográndose dos series históricas en soles constantes.

#### **3.6.1. Metodología Box -Jenkins y Box Tiao**

La serie histórica de la ejecución de gastos en soles constantes se iniciará con los supuestos de un modelo estocástico SARIMA. Sin embargo, la suposición básica esencial en el análisis de las series de tiempo es que los factores que han influido en el pasado y

en el presente en los modelos de la actividad económica de la institución seguirán haciéndolo más o menos en la misma forma en el futuro.

Se inicia a evaluar la estacionariedad de la serie de gasto utilizando, la prueba de Dickey Fuller Aumentada si los resultados indican la existencia de una tendencia estocástica y tendencia determinística. Para quitar los efectos de estacionalidad realizamos una transformación de diferencia estacional del orden 12. Aplicando la prueba de Dickey-Fuller a la serie transformada.

Ahora, vamos a analizar los correlogramas de la FAC y FACP de la serie de los gastos después de aplicación de la transformación de diferencia de orden 12, para identificar los posibles modelos para ajustar. Se deben de indagar los posibles valores tanto de la parte regular del modelo ( $p$  y  $q$ ) como de la parte estacional ( $P$  y  $Q$ ). buscando el modelo generador ajustado ARIMA o SARIMA y se desarrolla los sobreajustes del mejor modelo ajustado.

Se ejecuta la validación de los supuestos de los residuos que deben de comportarse como ruido blanco, es decir los residuos tienen una media cero, varianza constante, proceden de una población normal y que no tengan autocorrelación entre sus residuos. Si el modelo elegido no cumple con los requisitos de ruido blanco, se debe de elegir otro modelo de entre los candidatos o replantearse el modelo.

En la etapa de la adición de la variable exógena, debe de cumplir que tiene la misma transformación y aplicación de diferencias estacionarias y estacional con la variable dependiente, luego se usa a correlación cruzada para determinar con que rezagos entran estas variables en el proceso ARIMA y/o SARIMA identificado hasta ese momento.

El modelo seleccionado y ajustado ARIMAX y/o SARIMAX se le valida con los supuestos de ruido blanco. Luego se debe de evaluar la capacidad predictiva con las estadísticas tales el error cuadrático medio.

En conclusión, con la metodología, para el análisis de series temporales, formulado por Box and Jenkins se obtiene el modelo SARIMA al cual se incluye una variable exógena (ingreso), obteniendo como resultado un modelo SARIMAX, formulado por Box – Tiao.

### **3.6.2. Metodología del Modelo Vectorial Autorregresivo (VAR)**

Se procede a obtener el Modelo Vectorial Autorregresivo para lo cual se establecen como vector dependiente a los gastos mensuales de la UNI y como vector independiente a los ingresos mensuales.

Se evalúa si las series G y I tienen raíz unitaria o es no estacionaria con la prueba de Dickey Fuller Aumentada; si se cumple, aplicar diferencia estacionaria y/o estacional a las series con el mismo nivel. Una vez obtenido que las nuevas series no tiene raíz unitaria o es estacionaria se eligen en parejas de igual diferencia.

Se utilizan los criterios de información de Akaike (AIC), Schwarz (SC), Hanna Quinn (HQ) y el predictor final de error (FPE) para estimar el número de rezagos de las series estacionarias de transformación y diferencias del mismo nivel.

Después de estimar el modelo VAR, se realiza una prueba de la estabilidad del modelo y se prueban los desechos del modelo. Se debe verificar si los desechos tienen una distribución normal, no tienen autocorrelación hasta el retardo de orden h y son homocedásticos. Si se incumple una de las condiciones para clasificarlo como ruido blanco de desechos, se reconsidera la cantidad de residuos.

El análisis de cointegración se utiliza para determinar si dos variables de nivel similar están cointegradas entre sí. Se implementa un mecanismo de corrección de errores al modelo VAR en primeras diferencias si se encuentra al menos una relación de cointegración entre las variables.

El análisis de función Impulso-Respuesta se lleva a cabo después de cumplir con todos los requisitos para otorgar validez al modelo VAR estimado. El objetivo de este análisis es replicar el comportamiento o respuesta de una variable que se ve afectada por los choques o innovaciones provocados por otra variable. Para el análisis de la descomposición de la varianza a través de una simulación, determina el porcentaje de variabilidad o la importancia condicional de una variable sobre otras en el modelo, así como su propia variabilidad.

Finalmente, para comparar el modelo VAR con los modelos SARIMA y SARIMAX, se calcula el ECM del modelo.

Para el desarrollo de los diferentes análisis econométricos se utilizó instrumentos informáticos disponibles en el mercado.

## Capítulo IV: Resultados y Discusión

En este capítulo se presentan los resultados para el cumplimiento de los objetivos, que es el modelamiento y la predicción de las series temporales de la Recaudación de los Ingresos Corrientes (I) y la Ejecución del gasto (G) de la Universidad Nacional de ingeniería, por la fuente de financiamiento de Recursos Directamente recaudados. Nuestro objetivo principal es sugerir un modelo para la predicción de los gastos. Para ese fin ajustamos y comparamos las predicciones, basándose en los siguientes modelos:

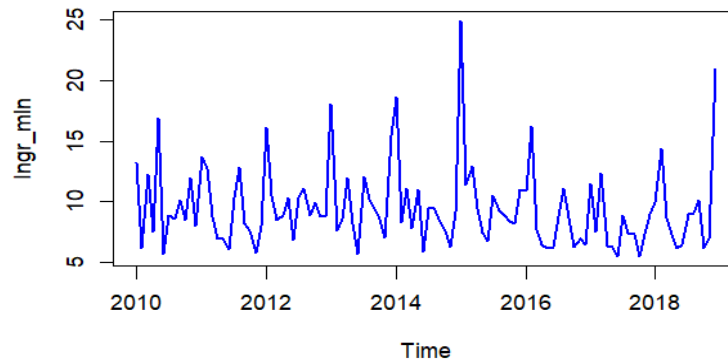
1. El modelo SARIMA, ajustado a la serie G.
2. El modelo SARIMAX, ajustado a las series G e I. Observemos, que, para ajustar este modelo, la serie I se utiliza como una serie de variables exógenas.
3. El modelo VAR, ajustado a la serie bidimensional, cuyas componentes son G y I.

Cabe notar que al contrario del modelo SARIMA, los modelos SARIMAX y VAR utilizan la información adicional, y por ende, si la correlación entre las series G e I son sustancialmente altas, las predicciones bajo el modelo SARIMA tendrán errores más grandes. Por otro lado, el modelo SARIMA al contrario del modelo VAR, toma en cuenta los efectos y autocorrelaciones estacionales. Por lo tanto, si las autocorrelaciones estacionales son altas, el modelo VAR puede producir peores predicciones, comparando con los dos otros modelos. Finalmente, el modelo SARIMAX es el modelo que incorpora tanto la información de la serie I, como las autocorrelaciones estacionales, sin embargo, ese modelo es más complejo, y por lo tanto su ajuste puede ser peor, comparando con los modelos VAR y SARIMA, lo que potencialmente puede afectar la bondad de las predicciones.

Como ya fue mencionado, los datos que se utilizan son las series de ingreso y gasto que empiezan en el mes de enero del 2010 y termina en diciembre del 2018, sumándose a 108 observaciones por cada serie, con un mismo año base a diciembre del 2017, como se muestra gráficamente en la Figura 11 y 12.

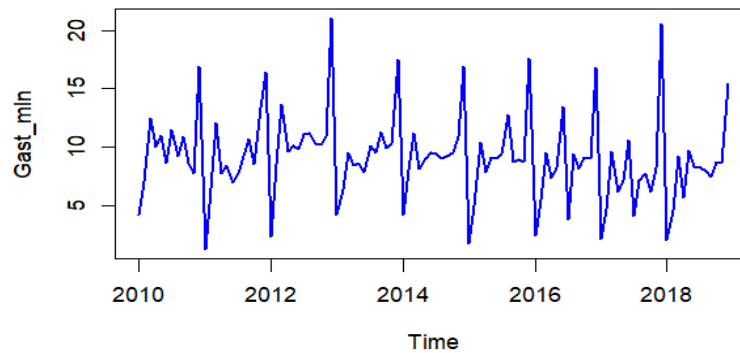
**Figura 11**

*Serie histórica de la Recaudación de Ingresos Corrientes (En millones de soles constantes)*



**Figura 12**

*Serie histórica de la Ejecución del gasto (En millones de soles constantes)*



#### **4.1 Modelo SARIMA para predicción de gastos futuros.**

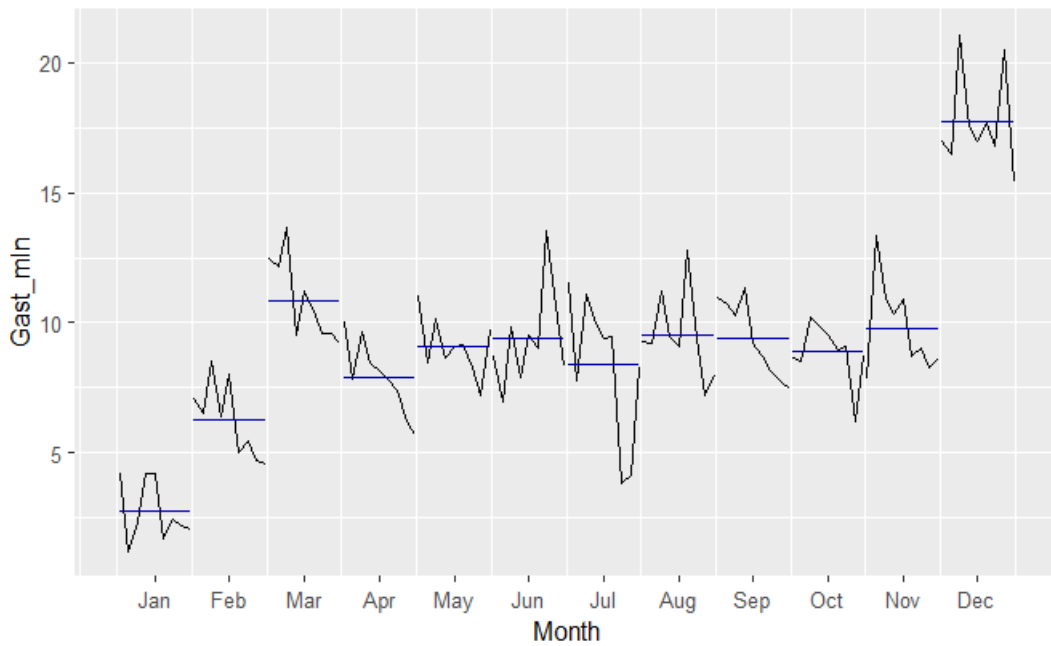
En esta sección analizamos la serie de los gastos, utilizando la metodología desarrollada por Box, Jenkins, Reinsel y Ljung (2016). El objetivo es estudiar el patrón del comportamiento de la serie a lo largo de los años 2010-2018 para poder utilizar este patrón para la realización de predicciones futuras.

Como ya fue explicado en el Capítulo 2, el primer paso del análisis es identificar el modelo (los modelos) que pueden, potencialmente, ajustarse bien a los datos. Para realizar esa tarea, vamos a analizar la estacionariedad de la serie y los correlogramas.

En la Figura 12 podemos observar que la serie no tiene tendencia a lo largo del tiempo. Por otro lado, se observan los picos que corresponden a los meses de diciembre y valles en los enero, lo que implica falta de estacionariedad del modelo (Figura 12).

**Figura 13.**

*Gastos observados en distinto meses*



Además, aplicando la prueba de Dickey-Fuller (vea la sección 2.4.4) para contrastar la hipótesis  $H_0: \gamma = 0$  contra  $H_1: \gamma < 0$ , se han obtenido el siguiente resultado (Tabla 3):

**Tabla 3**

*Resultados de aplicación de la prueba de Dickey-Fuller a la serie original de los gastos.*

Augmented Dickey-Fuller Test  
Alternative: stationary

Type 1: no drift no trend

	lag	ADF	p. value
[1, ]	0	-2.838	0.010
[2, ]	1	-1.557	0.119
[3, ]	2	-0.783	0.398
[4, ]	3	-0.608	0.461
[5, ]	4	-0.537	0.487

Type 2: with drift no trend

	lag	ADF	p. value
[1, ]	0	-12.40	0.01
[2, ]	1	-10.06	0.01
[3, ]	2	-6.30	0.01
[4, ]	3	-5.34	0.01
[5, ]	4	-4.57	0.01

Type 3: with drift and trend

	lag	ADF	p. value
[1, ]	0	-12.66	0.01
[2, ]	1	-10.68	0.01
[3, ]	2	-6.87	0.01
[4, ]	3	-5.98	0.01
[5, ]	4	-5.20	0.01

---

Note: in fact, p. value = 0.01 means p.value <= 0.01

Los resultados de la prueba “Type 1” indican la existencia de una tendencia estocástica. Por otro lado, análisis de los resultados de las pruebas “Type 2” y “Type 3” indica la existencia de una tendencia determinística (observemos que después de adicionar el factor de la tendencia determinística, es decir,  $a_2t$ , se rechaza  $H_0: \gamma = 0$ ). Sin embargo, la gráfica que presenta el comportamiento de la serie a lo largo del tiempo no demuestra ninguna tendencia. Dichas contradicciones ocurren por causa de la presencia de los efectos estacionales de la serie, lo que implica que la serie no es estacionaria. Para quitar los efectos de estacionalidad realizamos una transformación de diferencia estacional del orden 12. Aplicando la prueba de Dickey-Fuller a la serie transformada, obtenemos los siguientes resultados, presentados en la Tabla 4

**Tabla 4**

*Resultados de la prueba de Dickey-Fuller aplicada a la serie de diferencias de orden 12.*

```

Augmented Dickey-Fuller Test
Alternative: stationary

Type 1: no drift no trend
      lag  ADF      p. value
[1, ]  0  -8.29     0.01
[2, ]  1  -5.74     0.01
[3, ]  2  -4.50     0.01
[4, ]  3  -3.76     0.01
Type 2: with drift no trend
      lag  ADF      p. value
[1, ]  0  -8.32     0.01
[2, ]  1  -5.78     0.01
[3, ]  2  -4.54     0.01
[4, ]  3  -3.79     0.01
Type 3: with drift and trend
      lag  ADF      p. value
[1, ]  0  -8.30     0.0100
[2, ]  1  -5.78     0.0100
[3, ]  2  -4.56     0.0100
[4, ]  3  -3.83     0.0207
---
Note: in fact, p. value = 0.01 means p.value <= 0.01

```

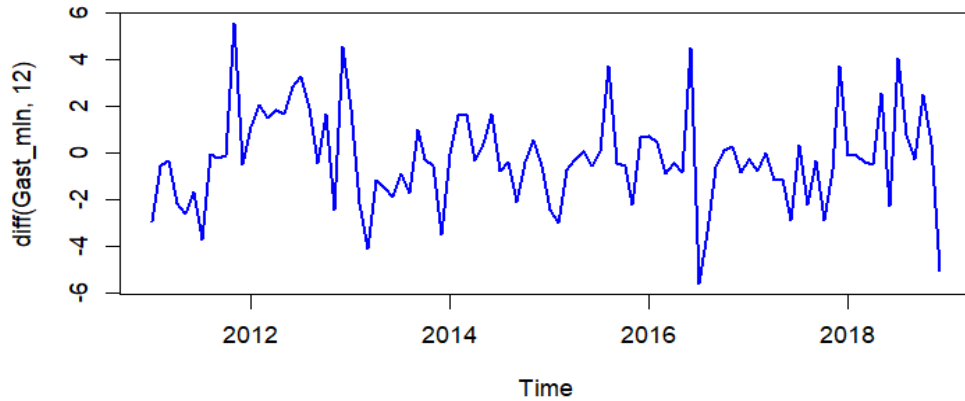
Los resultados demuestran que la serie no posee tendencia ni de origen determinístico ni de origen estocástico. La Tabla 4 presenta el comportamiento de la serie de los gastos. después de aplicación de la transformación de diferencia de orden 12.



**Figura 14**

*La serie de los gastos. después de aplicación de la transformación de diferencia de orden*

*12*

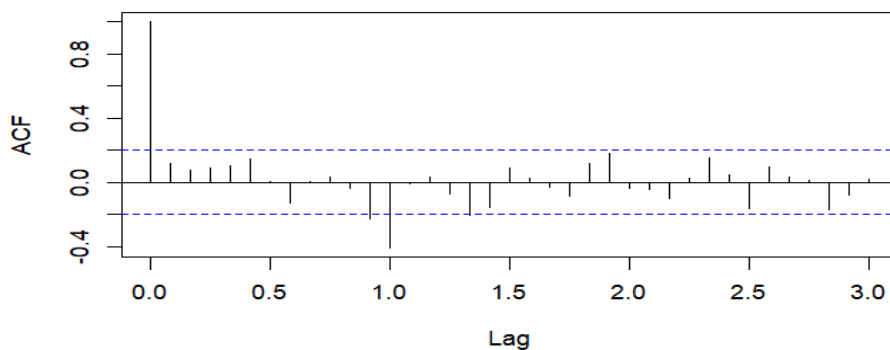


De la Figura 14 se puede concluir que la serie de las diferencias de orden 12 no presenta tendencia, ni un patrón de estacionalidad, que se repite cada cierto periodo de tiempo. Además, dicha serie no presenta cambios notables en la dispersión de los datos a lo largo del tiempo. Es decir, no hay presencia de periodos, donde la varianza crezca o disminuya. Sin embargo, puesto que en cada tiempo  $t$ , se observa sólo una realización, no existen métodos para demostrar que la varianza es constante a lo largo del tiempo, y por lo tanto, la conclusión al respecto se basa, generalmente en visualización del comportamiento de la serie a lo largo del tiempo.

Ahora, vamos a analizar los correlogramas de la serie de los gastos después de aplicación de la transformación de diferencia de orden 12, para identificar los posibles modelos para ajustar (Figuras 15 y 16).

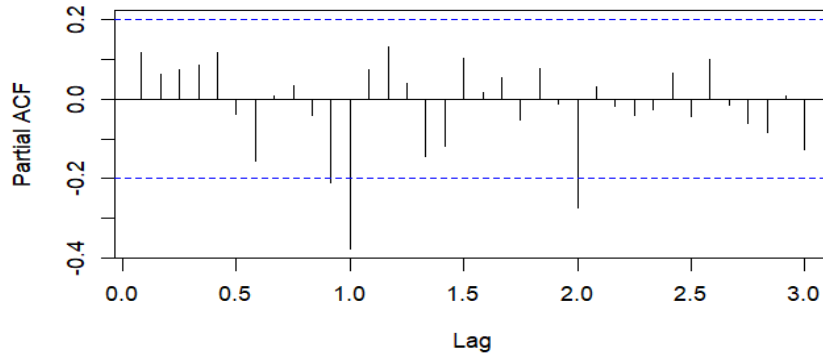
**Figura 15**

*Autocorrelaciones de la serie de los gastos después de aplicación de la transformación de diferencia de orden 12.*



**Figura 16**

*Autocorrelaciones parciales de la serie de los gastos después de aplicación de la transformación de diferencia de orden 12.*



Los resultados, presentados en las Figuras 15 y 16, donde las únicas autocorrelaciones significativas son  $\rho_{12}$ ,  $\phi_{12,12}$  y  $\phi_{24,24}$ , evidencian que el modelo más adecuado para ajustarse a la serie original de los gastos es el modelo  $SARIMA(0,0,0)(0,1,1)_{12}$ . La siguiente tabla resume el ajuste del modelo indicado junto con modelos parecidos.

**Tabla 5**

*Resultados de ajuste de modelos SARIMA a la serie original de los gastos: estimadores de los coeficientes de los modelos con sus p-valores respectivos (en las parentesis)*

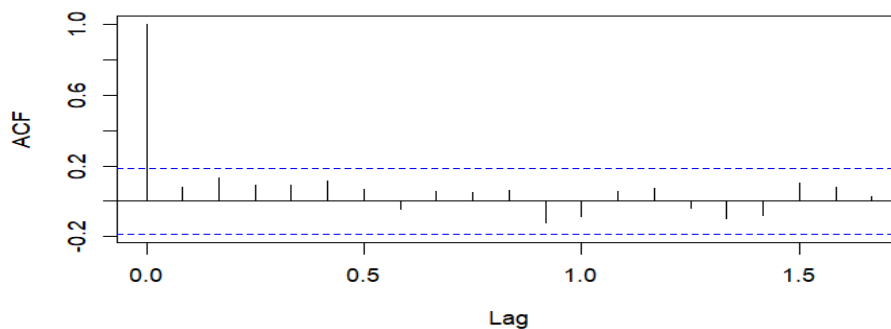
Modelo SARIMA	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\theta}_{12}$	$\hat{\theta}_{24}$	$\hat{\phi}_{12}$	$\hat{\phi}_{24}$	AIC
$(0,0,0)(0,1,1)_{12}$	---	---	---	---	-0.575 (0.000)	---	---	---	388.70
$(0,0,0)(1,1,0)_{12}$	---	---	---	---	---	---	-0.490 (0.000)	---	393.58
$(0,0,0)(1,1,1)_{12}$	---	---	---	---	-0.472 (0.007)	---	-0.142 (0.477)	---	390.23
$(0,0,0)(0,1,2)_{12}$	---	---	---	---	-0.622 (0.000)	0.101 (0.447)	---	---	390.13
$(0,0,0)(2,1,0)_{12}$	---	---	---	---	---	---	-0.614 (0.000)	-0.285 (0.013)	389.86
$(0,0,1)(0,1,1)_{12}$	0.101 (0.263)	---	---	---	-0.595 (0.000)	---	---	---	389.45
$(1,0,0)(0,1,1)_{12}$	---	---	0.138 (0.187)	---	-0.595 (0.000)	---	---	---	388.97
$(0,1,0)(0,1,1)_{12}$	---	---	---	---	-0.646 (0.000)	---	---	---	435.24
$(0,1,1)(0,1,1)_{12}$	-0.898 (0.000)	---	---	---	-0.726 (0.000)	---	---	---	382.00
$(0,1,2)(0,1,1)_{12}$	-0.912 (0.000)	0.027 (0.408)	---	---	-0.729 (0.000)	---	---	---	383.94
$(1,1,0)(0,1,1)_{12}$	---	---	-0.541 (0.000)	---	-0.699 (0.000)	---	---	---	404.80

Modelos candidatos:  $M_1: SARIMA(0,0,0)(0,1,1)_{12}$ ,  $M_2: SARIMA(0,0,0)(1,1,0)_{12}$ ,  $M_3: SARIMA(0,1,1)(0,1,1)_{12}$ . Observemos que los modelos candidatos seleccionados no incluyen coeficientes no significativos.

El siguiente paso es evaluar el comportamiento de los errores estimados. Como fue explicado en la Sección 2.2.5, se supone que los errores presentan una serie de ruido blanco, que, además, siguen una distribución normal. Para contrastar estos supuestos, se utilizan los errores estimados del modelo ajustado (ver la ecuación 56). El primer supuesto puede ser contrastado, analizando las autocorrelaciones de los errores estimados (Figuras 17 al 19) y aplicando la prueba de Ljung-Box si los datos se distribuyen de forma independiente; es decir no hay autocorrelación en el modelo (ver ecuación 58 y Tabla 6). El supuesto de la normalidad fue evaluado basándose en las gráficas Q-Q plot (Figuras 20 al 22) y pruebas de Shapiro-Wilk y Jarque-Bera (Tabla 7).

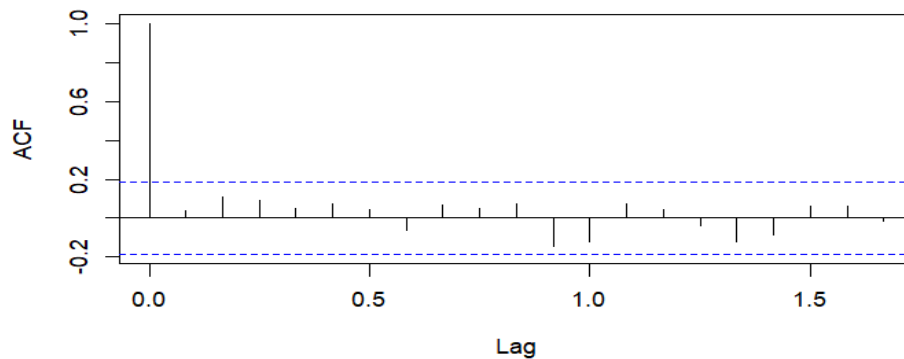
**Figura 17**

*Autocorrelaciones de los residuos del modelo  $M_1$*



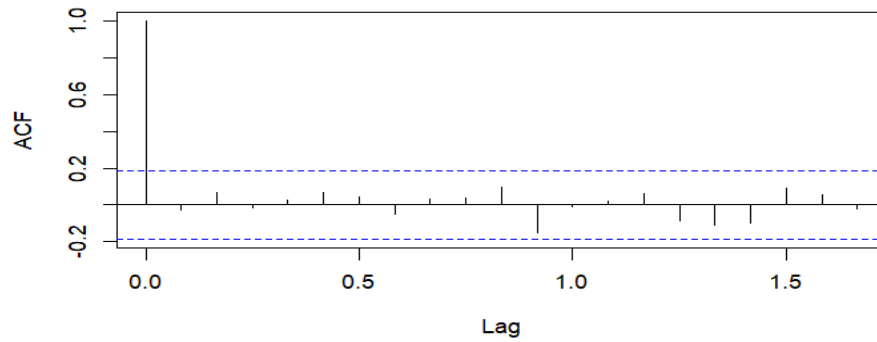
**Figura 18**

*Autocorrelaciones de los residuos del modelo  $M_2$*



**Figura 19**

Autocorrelaciones de los residuos del modelo  $M_3$



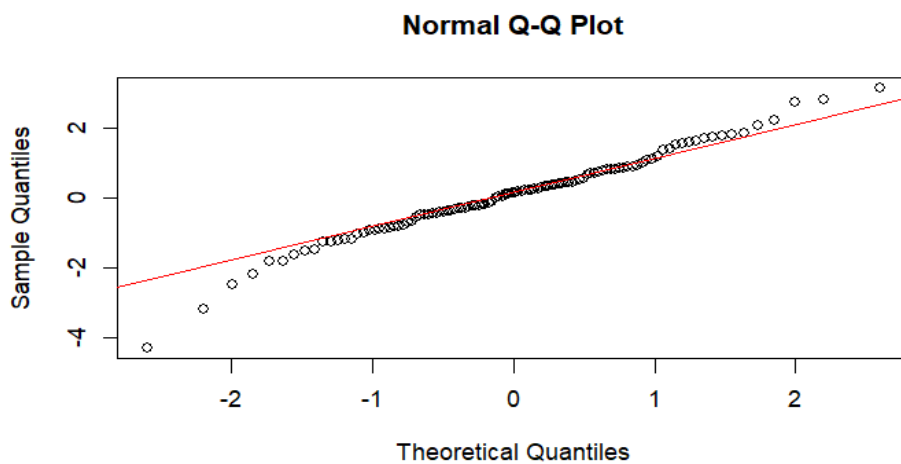
**Tabla 6**

Resultados de aplicación de la prueba de Ljung-Box a los residuos de los modelos  $M_1$ ,  $M_2$  y  $M_3$

Modelo	$K = 12$		$K = 24$		$K = 36$	
	$Q$	p-valor	$Q$	p-valor	$Q$	p-valor
$M_1$	10.08	0.6097	19.21	0.7406	24.91	0.9179
$M_2$	8.90	0.7112	20.83	0.6489	27.10	0.8578
$M_3$	5.50	0.9391	13.87	0.9496	21.28	0.9755

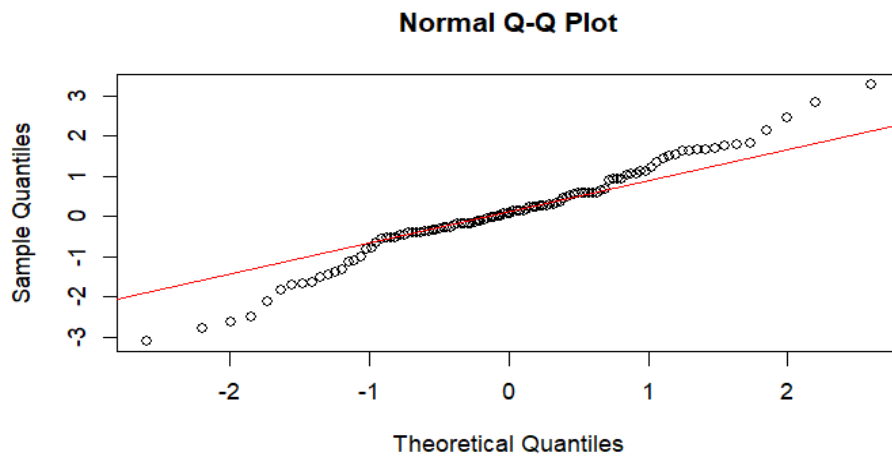
**Figura 20**

Q-Q plot de los residuos del modelo  $M_1$



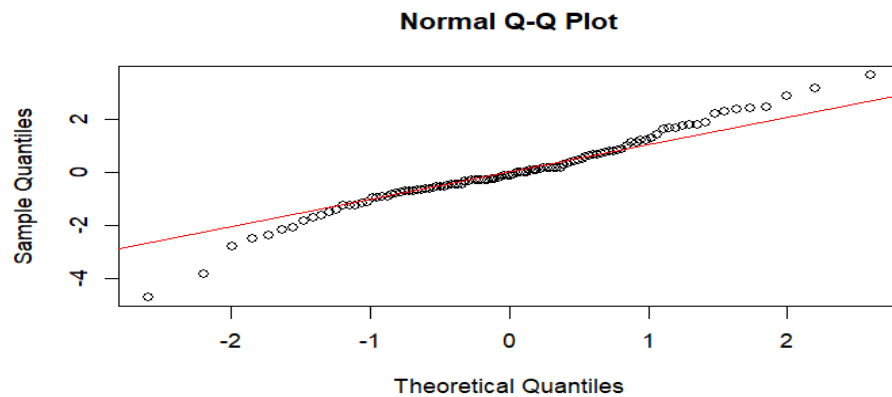
**Figura 21**

Q-Q plot de los residuos del modelo  $M_2$



**Figura 22**

Q-Q plot de los residuos del modelo  $M_3$



**Tabla 7**

Resultados de aplicación de las pruebas de Shapiro-Wilk y Jarque-Bera a los residuos de los modelos  $M_1$ ,  $M_2$  y  $M_3$ .

Modelo	Shapiro-Wilk		Jarque-Bera	
	Estadístico	p-valor	Estadístico	p-valor
$M_1$	0.9797	0.0983	3.7413	0.1540
$M_2$	0.9792	0.0885	2.5719	0.2764
$M_3$	0.9735	0.0297	8.3978	0.0150

Los resultados obtenidos demuestran que los residuos de los tres modelos candidatos no presentan autocorrelaciones, y, por ende, en los tres casos, no se rechaza la hipótesis que los residuos presentan un comportamiento de ruido blanco. Sin embargo, de los Q-Q plots se puede observar que la distribución de los mismos tiene colas más

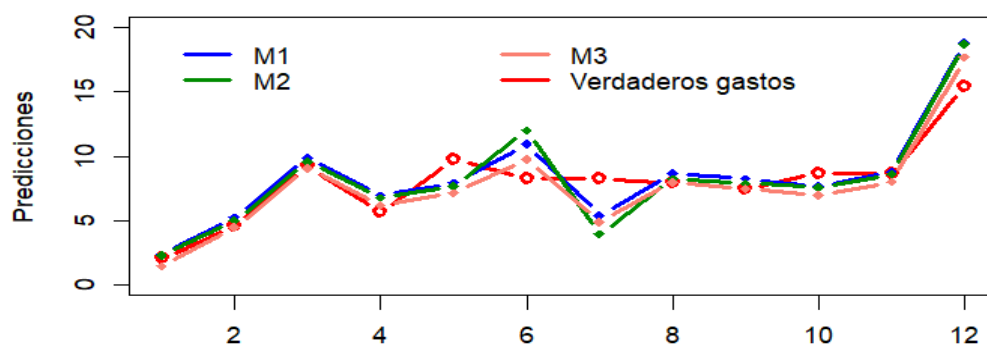
pesadas comparando con una distribución normal para todos los modelos. Por otro lado, para los modelos  $M_1$  y  $M_2$  no se rechaza la hipótesis nula que establece que los residuos siguen una distribución normal. Concluyendo, se puede constatar que los tres modelos pueden ser utilizados para realizar predicciones futuras. Sin embargo, una cierta precaución es necesaria para utilizar el modelo  $M_3$  puesto que los residuos del mismo no tienen una distribución normal. Además, en el caso del modelo tres, se realiza una transformación de diferencia regular, que según el análisis preliminar (Figuras 12 y 13) no es necesaria.

Otro abordaje para evaluar el desempeño de los modelos es la comparación de los mismos en términos de predicción de valores futuros. A fin de realizar ese paso, ajustamos los modelos  $M_1$ ,  $M_2$  y  $M_3$  a los datos hasta el año 2017, estimamos los parámetros desconocidos correspondientes, y luego utilizamos estos modelos para predecir los valores (los gastos) en el año 2018 (ver Figura 23).

Comparación entre las predicciones y verdaderos valores, observados en 2018 nos permite obtener el error medio, el error medio absoluto y el error cuadrático medio de las predicciones, definidos en la sección 2.2.6. Tabla 8 resume los errores de predicción bajo cada uno de los modelos candidatos.

**Figura 23**

*Comparación de los verdaderos gastos, realizados en el año 2018 con las predicciones bajo los modelos  $M_1$ ,  $M_2$  y  $M_3$*



**Tabla 8**

Resumen de los errores de predicción bajo los modelos  $M_1$ ,  $M_2$  y  $M_3$

Modelo	$B^{(M)}$	$Abs^{(M)}$	$ECM^{(M)}$
$M_1$	-0.417	1.362	2.904
$M_2$	0.205	1.442	4.146
$M_3$	0.395	1.118	2.406

Los resultados conseguidos indican que no hay diferencia muy considerable entre los modelos. Sin embargo, tomando en cuenta que para el modelo  $M_3$  se rechaza la hipótesis de la normalidad de los errores, los modelos  $M_1$  o  $M_2$  podrían ser los modelos seleccionados para predecir los valores de los gastos futuros.

#### 4.2 Modelo SARIMAX para predicción de gastos futuros.

En esta sección vamos a ampliar los modelos  $M_1$  y  $M_2$  adicionando como la variable exógena los ingresos  $X_t$  (sección 2.3.3, ecuación 68). Cabe notar, que en esta sección utilizamos las series de las diferencias de orden 12 de los ingresos y de los gastos, es decir,  $(1 - L^{12})X_t$  y  $(1 - L^{12})Z_t$  para ajustar los modelos. Tabla 9 resume los resultados de ajuste de los modelos *SARIMAX*,  $\tilde{M}_1$  y  $\tilde{M}_2$ , basados en los modelos  $M_1$  y  $M_2$ , respectivamente.

**Tabla 9**

Resultados de ajuste de modelos *SARIMAX* a la serie original de los gastos,  $Z_t$  usando los ingresos como variables exógenas,  $X_t$ : estimadores de los coeficientes de los modelos con sus *p*-valores respectivos (en los paréntesis)

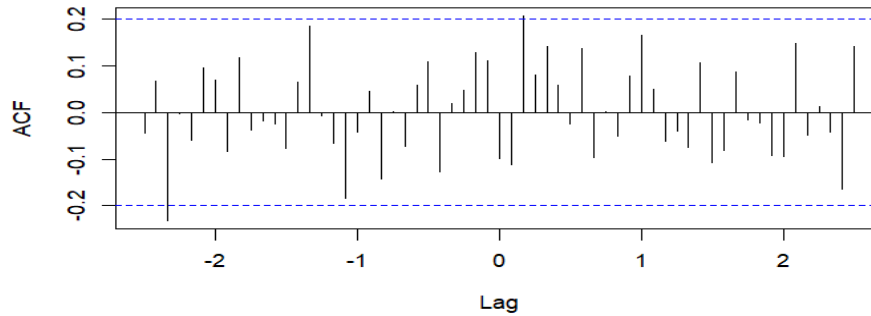
Modelo SARIMAX	$\hat{\theta}_1$	$\hat{\theta}_{12}$	$\hat{\Phi}_{12}$	$\hat{\beta}$	AIC
$\tilde{M}_1$	---	-0.581 (0.000)	---	0.0062 (0.923)	388.41
$\tilde{M}_2$	---	---	-0.489 (0.000)	0.0074 (0.902)	391.58

Observemos, que aparentemente, los ingresos en el tiempo  $t$ , no proporcionan información valiosa para predecir los gastos del mismo tiempo ya que la ejecución del gasto tiene un proceso de ejecución por etapas después de la recaudación de ingresos. Sin embargo, el modelo *SARIMAX* definido en la sección 2.3.3, puede ser modificado,

reemplazando en la ecuación (70)  $X_t$  por  $X_{t-b}$ , para un valor  $b$  adecuado, que puede ser obtenido a partir de la gráfica de correlaciones cruzadas (Figura 24).

**Figura 24**

*Ingresos y gastos: correlaciones cruzadas*



Basándose en los resultados, presentados en la Figura 24, se puede utilizar  $X_{t+2}$  para explicar los gastos en el tiempo  $t$ , es decir,  $Z_t$ . Cuando hay una continuidad del convenio o del servicio que presta la institución a terceros los ingresos futuros explican los gastos. Sin embargo, cabe destacar que según la Figura 24, la correlación entre  $(1 - L^{12})Z_t$  y  $(1 - L^{12})X_{t+2}$  no son fuertes.

La Tabla 10, resume los resultados de ajuste de los modelos *SARIMAX* modificados, basados en los modelos  $M_1$ ,  $M_2$  y  $M_3$ .

**Tabla 10**

*Resultados de ajuste de modelos SARIMAX a la serie original de los gastos,  $Z_t$  usando los ingresos como variables exógenas,  $X_{t+2}$ : estimadores de los coeficientes de los modelos con sus p-valores respectivos (en los paréntesis)*

Modelo SARIMAX	$\hat{\theta}_1$	$\hat{\theta}_{12}$	$\hat{\Phi}_{12}$	$\hat{\beta}$	AIC
$\tilde{M}_1$ Modificado	---	-0.487 (0.000)	---	0.122 (0.052)	374.44
$\tilde{M}_2$ Modificado	---	---	-0.435 (0.000)	0.110 (0.062)	376.81

Los resultados presentados en Tabla 10 indican que el modelo  $\tilde{M}_1$  modificado tiene AIC más bajo, y, además, bajo este modelo el coeficiente  $\beta$  salió más significativo, lo que puede interpretarse que dicho modelo es mejor para realizar las predicciones futuras. Examinando los residuos de ambos modelos, fueron obtenidos resultados muy parecidos con los obtenidos para los modelos  $M_1$  y  $M_2$ . Es decir, en ambos casos, no se rechaza la



hipótesis que los residuos tienen un comportamiento de ruido blanco. Respecto a la normalidad, de igual manera, los Q-Q plots son muy parecidos con los Q-Q plots obtenidos para los residuos de los modelos  $M_1$  y  $M_2$ . Aplicación de las pruebas de Shapiro-Wilk y Jarque-Bera resultaron en p valores de 0.162 y 0.119, y 0.184 y 0.122, respectivamente.

Para evaluar la eficiencia del modelo  $\tilde{M}_1$  modificado, en términos de los errores de predicción, fueron calculadas las predicciones bajo este modelo para los meses de enero-octubre del año 2018. Observemos que para predecir los gastos para los meses de noviembre y diciembre bajo  $\tilde{M}_1$  modificado, necesitamos saber los ingresos en los meses de enero y febrero del año 2019. La Tabla 11 resume los errores de predicción bajo los modelos  $M_1$ ,  $M_2$ ,  $M_3$  y  $\tilde{M}_1$  modificado, para los meses de enero- octubre del año 2018.

**Tabla 11**

*Resumen de los errores de predicción bajo los modelos  $M_1$ ,  $M_2$ ,  $M_3$  y  $\tilde{M}_1$  modificado, en los meses de enero-octubre del año 2018*

<b>Modelo</b>	$B^{(M)}$	$Abs^{(M)}$	$ECM^{(M)}$
$M_1$	-0.141	1.275	2.305
$M_2$	0.079	1.401	3.908
$M_3$	0.638	1.054	2.342
$\tilde{M}_1$ modificado	0.225	1.208	3.627

Los resultados obtenidos demuestran que en términos de predicciones el modelo  $\tilde{M}_1$  modificado no es mejor comparando con los modelos  $M_1$ ,  $M_2$  y  $M_3$ . Esto se debe a que el efecto de los ingresos no es muy pronunciado, y por ende, la inclusión de esa variable como una variable exógena en el modelo no contribuye mucha información al mismo.

#### **4.3 Modelo VAR para predicción de gastos futuros.**

Continuando con el análisis de las variables para la predicción del gasto, en esta sección ajustamos un modelo VAR a las variables. Esta relación es de central interés en la gestión pública puesto que puede ayudar a explicar el gasto en el sector público ante un shock en el ingreso. Recordemos que la teoría económica sugiere que el gasto depende del ingreso. Por lo tanto, nosotros consideramos ecuaciones simultáneas para el ingreso y el gasto. De forma similar como fue explicado en el Capítulo 2, el primer paso del análisis es identificar el modelo (los modelos) VAR que pueden, potencialmente, ajustarse bien a los datos. Para realizar esa tarea, vimos que ambas series no son estacionarias. Al analizar

la existencia de raíz unitaria en las series se encontraron algunas contradicciones de la serie en los tests. Dichas contradicciones ocurren por causa de la presencia de los efectos estacionales de la serie, lo que implica que la serie no es estacionaria (raíz unitaria asociada a la componente estacional). Para quitar los efectos de estacionalidad realizamos una transformación de diferencia estacional del orden 12 y trasladamos el ingreso dos unidades de tiempo, es decir, consideraremos las variables  $(1 - L^{12})G_t$  y  $(1 - L^{12})I_{t+2}$  (ver correlogramas y correlaciones cruzadas en la sección anterior para su justificación).

Después ajustamos un modelo VAR a las series resultantes. Es necesario tratar las componentes estacionales puesto que puede distorsionar la relación entre las variables de interés (Lutkepohl y Kratzig, 2004). Según los resultados del análisis de raíz unitaria, mostrada en la Tabla 3, tenemos que la aparente presencia de raíz unitaria en las series es generada por una componente estacional altamente persistente. Otro análisis importante al ajustar modelos VAR es el análisis de cointegración, pero en este caso, como las series con las que trabajaremos son estacionarias (las series transformadas por una diferencia de orden 12), el análisis de cointegración nos da que la hipótesis nula  $H_0: r \leq r_0$  siempre se rechaza para  $r_0 = 0, 1$  y para los niveles de significancia del 5% y 10%, es decir, el rango de la matriz que contiene las relaciones de cointegración es dos. Estos resultados son mostrados en la Tabla 12. En dicha tabla observamos que para un número pequeño de rezagos para la diferencia colocados en el modelo el modelo la hipótesis nula siempre se rechaza, pero eso cambia cuando colocamos 12 rezagos, donde se observa una aparente relación de cointegración, la hipótesis nula no es rechazada. Esta conclusión es consecuencia por la alta persistencia en la componente estacional. Entonces, para nuestro caso, el MCE no es informativo, y es mejor analizar directamente un modelo VAR o SVAR para nuestras variables.

**Tabla 12**

*Prueba de cointegración de Johansen para las variables  $(1 - L^{12})G_t$  y  $(1 - L^{12})I_{t+2}$*

Variables	Nº de rezagos para las diferencias	$H_0: r = r_0$	Prueba estadística	Valor crítico 10%	Valor crítico 5%
$I, G$	1	$r_0 = 0$	66.24	17.85	19.96
		$r_0 \leq 1$	28.56	7.52	9.24
	12	$r_0 = 0$	34.05	17.85	19.96
		$r_0 \leq 1$	5.19	7.52	9.24

El siguiente paso en nuestro análisis es buscar un modelo adecuado para nuestro sistema bidimensional  $((1 - L^{12})G_t, (1 - L^{12})I_{t+2})$ . Como el modelo VAR es un modelo autorregresivo, tenemos que seleccionar el número adecuado de rezagos. Como la prueba de cointegración nos sugiere que el rango de la matriz de largo plazo es dos, entonces ajustamos directamente un modelo VAR para nuestras variables. Según Lutkepohl (2005) los criterios de información HQ y SC proporcionan estimaciones consistentes del verdadero orden del retardo del modelo, mientras que el AIC y FPE sobre estiman el orden del retardo del modelo con probabilidad positiva. Entonces, para seleccionar un modelo adecuado usamos los criterios de información HQ y SC. Los resultados son mostrados en la Tabla 13.

**Tabla 13**

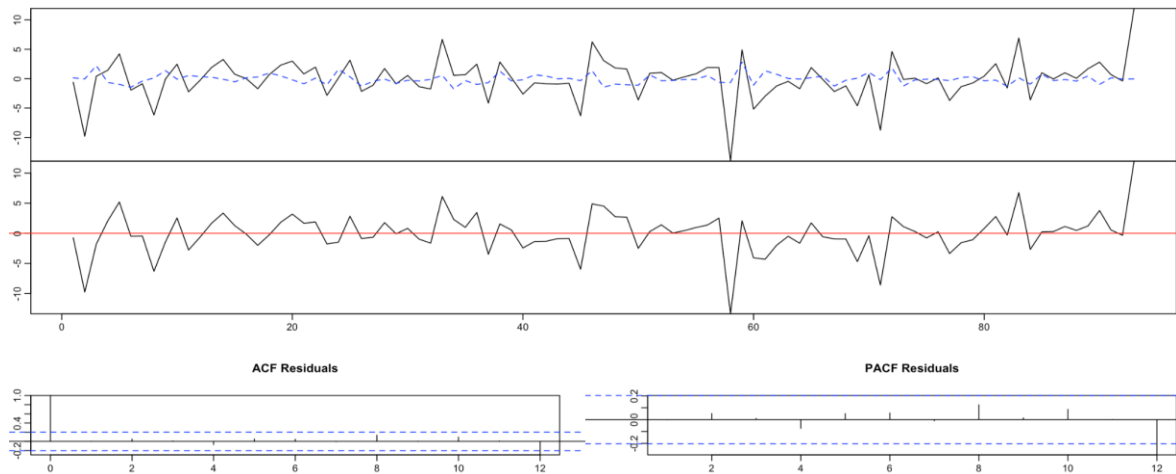
*Criterios de información AIC, HQ, SC y FPE*

	<i>AIC(n)</i>	<i>HQ(n)</i>	<i>SC(n)</i>	<i>FPE(n)</i>
1	3.714997	3.785699	3.891098	41.061136
2	3.787770	3.905607	4.081273	44.171220
3	3.845829	4.010800	4.256732	46.836447
4	3.874734	4.086840	4.403038	48.255506
5	3.905736	4.164976	4.551441	49.848556
6	3.968336	4.274711	4.731442	53.182554
7	4.016146	4.369656	4.896653	55.950959
8	4.038792	4.439437	5.036700	57.454260
9	4.084213	4.531992	5.199522	60.422106
10	4.103055	4.597969	5.335765	61.952924
11	4.058264	4.600312	5.408374	59.689625
12	3.732049	4.321232	5.199560	43.470892

Según la Tabla 13: y el criterio de información HQ y SC tenemos que tanto HQ como SC sugieren un modelo  $VAR(1)$ . Los gráficos para el ajuste y los residuos son mostrados en las Figuras 25 y 26. Estos gráficos sugieren fuertemente que los residuos del modelo no tienen distribución normal.

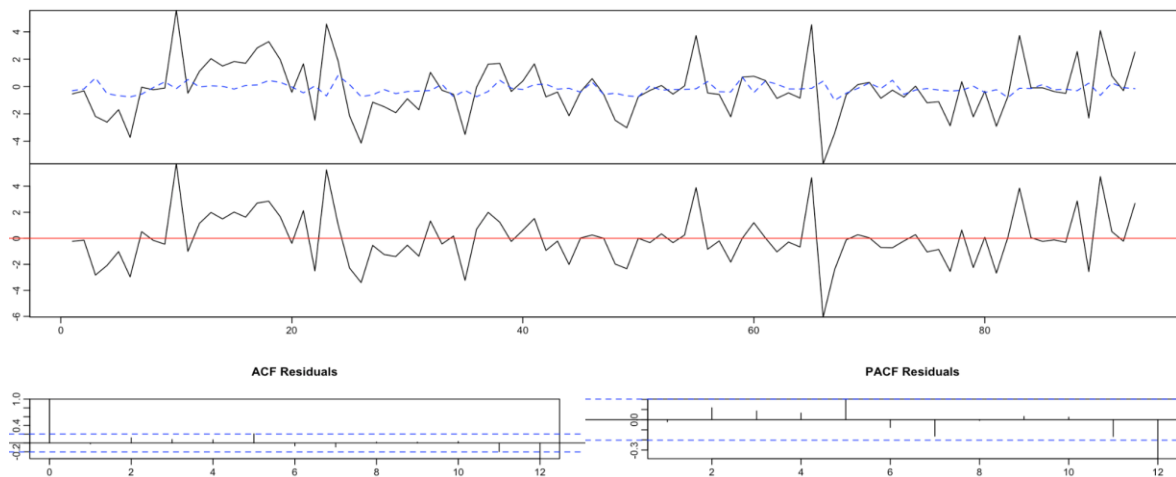
**Figura 25**

*Ajuste y residuos para I, usando un modelo VAR(1). La línea azul es el ajuste del modelo.*



**Figura 26**

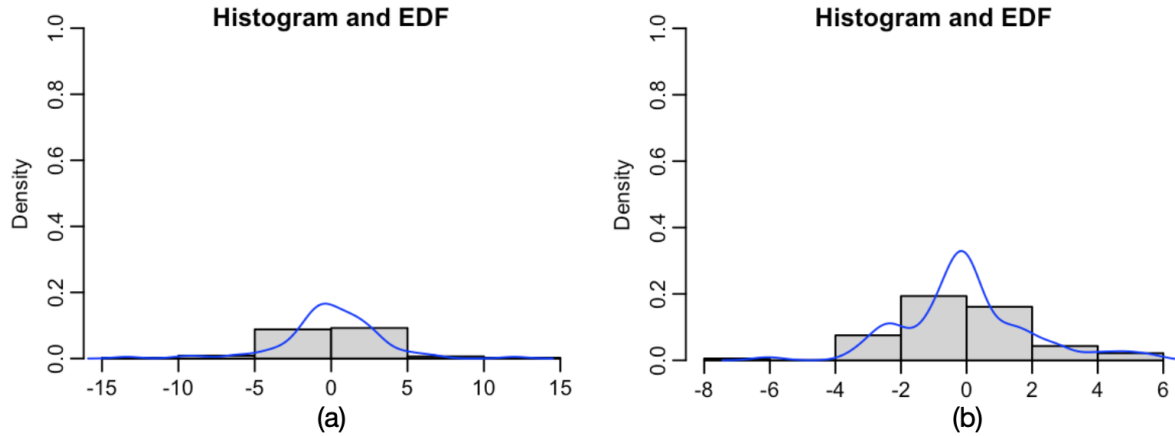
*Ajuste y residuos para G, usando un modelo VAR(1). La línea azul es el ajuste del modelo.*



En las Figuras 25 y 26, observamos que el ajuste no es bueno, y que la componente estacional aún es persistente. Esto también se puede ver claramente en el PACF de ambas series. En el caso de la prueba de normalidad multivariada, el test de Jarque-Bera multivariado y univariado nos da p-valores menores que  $2.2 \times 10^{-16}$ , sugiriendo que los residuos multivariados, así como de las marginales, sigue una distribución normal como se observa en la Figura 27.

**Figura 27**

Gráfica de los residuos del modelo VAR(1) ajustado para las variables (a) Ingreso y (b) Gasto



Para el caso del test de heterocedasticidad, este nos da un p-valor de 0.8566, sugiriendo la ausencia de heterocedasticidad en los residuos resultantes del ajuste. En el caso de la prueba de correlación serial, el test de Breusch-Godfrey nos da un p-valor de 0.1929, sugiriendo que la ausencia de correlación serial para los residuos después de ajustar un VAR(1).

Usando la técnica de ajuste, descrito en la sección 2.4, tenemos que el modelo final estimado es escrito como

$$\begin{bmatrix} \Delta^{12}I_{t+2} \\ \Delta^{12}G_t \end{bmatrix} = \begin{bmatrix} -0.07374 \\ -0.14454 \end{bmatrix} + \begin{bmatrix} -0.24008 & 0.21546 \\ -0.08402 & 0.15525 \end{bmatrix} \begin{bmatrix} \Delta^{12}I_{t+1} \\ \Delta^{12}G_{t-1} \end{bmatrix} + \begin{bmatrix} \hat{\varepsilon}_{1t} \\ \hat{\varepsilon}_{2t} \end{bmatrix}$$

donde  $\hat{\varepsilon}_t \sim N(0, \hat{\Sigma})$  y

$$\hat{\Sigma} = \begin{bmatrix} 11.313 & 1.157 \\ 1.157 & 3.949 \end{bmatrix}$$

Como queremos incorporar efectos contemporáneos, a partir del modelo VAR ajustaremos un modelo  $By_t = A_0 + A_1y_{t-1} + e_t$ . La matriz B incorpora el efecto instantáneo de la serie. Para esto solamente necesitamos estimar la matriz B. Para estimar la matriz B aplicamos la descomposición de Choleski a la matriz de covarianza de los residuos e introducimos algunas restricciones de contemporaneidad al modelo (Enders, 2014). Para el caso de ingreso y gasto, los test de causalidad y la teoría económica sugieren que la matriz B es de la forma,

$$B = \begin{bmatrix} 1 & 0 \\ * & 1 \end{bmatrix}$$

La serie ingreso es la fuerza dominante de la realización de los objetivos institucionales, mientras que la serie gasto representa el interés de los convenios, la prestación de servicios a terceros y además con la ejecución del gasto con los saldos

obtenidos, complementan el cierre de brechas en el presupuesto por resultados de la institución público en este caso la UNI.

Al introducir la matriz B en la descomposición de Cholesky a la matriz de covarianza de los residuos obtenemos la siguiente estimación,

$$\hat{B} = \begin{bmatrix} 1 & 0 \\ -0.9 & 1 \end{bmatrix}$$

Finalmente, para comparar el modelo VAR con los modelos SARIMA y SARIMAX, calcular el ECM del modelo. En este caso, el ECM para todo el modelo es de 2.263 (para ambas variables), mientras que el ECM para la variable gasto, que es la principal variable que queremos explicar, es de 3.822. Ver Tabla 14.

**Tabla 14**

*Sesgo, error absoluto y error cuadrático medio para el modelo ajustado VAR(1)*

<b>Modelo</b>	$B^{(M)}$	$Abs^{(M)}$	$ECM^{(M)}$
VAR(1)	-2.296e-16	1.386	3.822

Para evaluar la eficiencia del modelo VAR, en términos de los errores de predicción, fueron calculadas las predicciones bajo este modelo para los meses de enero-octubre del año 2018. Observemos que, para predecir los gastos para los meses de noviembre y diciembre, necesitamos saber los ingresos en los meses de enero y febrero del año 2019. La Tabla 15 resume los errores de predicción bajo el modelo VAR.

**Tabla 15**

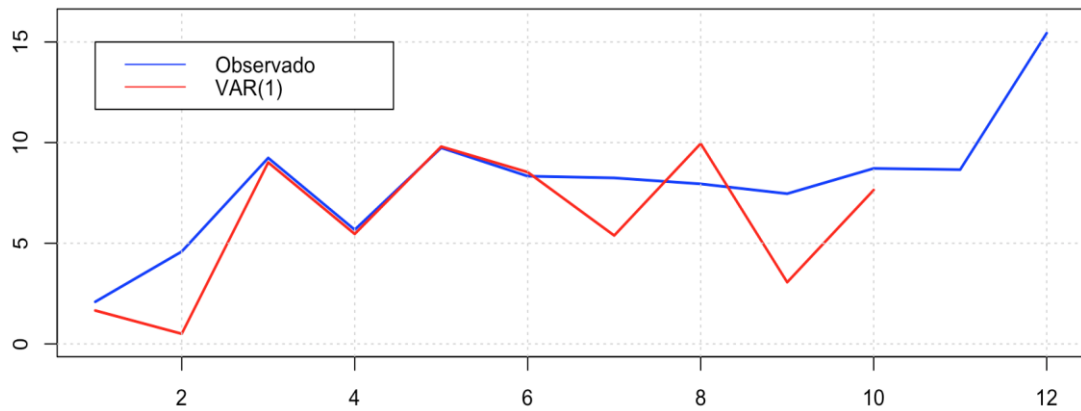
*Sesgo, error absoluto y error cuadrático medio de predicción para los meses de enero-octubre del año 2018, para el modelo ajustado VAR(1)*

<b>Modelo</b>	$B^{(M)}$	$Abs^{(M)}$	$ECM^{(M)}$
VAR(1)	1.103	1.556	4.972

En la Figura 28 se ajusta el modelo VAR(1) a los datos hasta el año 2017, estimamos los parámetros desconocidos correspondientes, y luego se utiliza este modelo para predecir los valores (los gastos) en el año 2018.

**Figura 28**

*Gasto observado vs estimación usando modelo VAR*



Comparando la Tabla 8 y 11 del modelo SARIMA y SARIMAX, respectivamente, con la Tabla 15 del modelo VAR, observamos que los errores de predicción del modelo VAR no mejoran, pudiendo concluir que el modelo SARIMA es el que menor error de predicción produce.

## Capítulo V: Conclusiones y Recomendaciones

### 5.1. Conclusiones

En este trabajo de investigación se han aplicado tres modelos; los Modelo SARIMA y SARIMAX, basado en la Metodología de Box-Jenkins y Box-Tiao y el Modelo Vectorial Autorregresivo (VAR) para hallar la capacidad predictiva del modelo gasto, por la fuente de financiamiento de recursos directamente recaudados, de la Universidad Nacional de Ingeniería. Luego de la aplicación de las metodologías a los datos de estudio y de acuerdo a los resultados obtenidos en el Capítulo IV se han obtenido las siguientes conclusiones:

#### 5.1.1. Conclusión General

En la estimación y la capacidad predictiva el Modelo SARIMA dio un mejor ajuste en los resultados, en comparación con el modelo SARIMAX y VAR por lo que se concluye que el modelo funciona bien tanto a nivel de estimación como a nivel de predicción.

#### 5.1.2. Conclusión Especifica 1

Los Modelos candidatos elegidos  $M_1: SARIMA(0,0,0)(0,1,1)_{12}$ ,  $M_2: SARIMA(0,0,0)(1,1,0)_{12}$ ,  $M_3: SARIMA(0,1,1)(0,1,1)_{12}$ , no incluyen coeficientes no significativos. El criterio de información de Akaike es de 388.70, 393.58 y 382 respectivamente.

Por otro lado, para los modelos  $M_1$  y  $M_2$  no se rechaza la hipótesis nula que establece que los residuos siguen una distribución normal, no presentan autocorrelaciones, y, por ende, en los dos casos, no se rechaza la hipótesis que los residuos presentan un comportamiento de ruido blanco.

El error cuadrático medio para el modelo  $M_1$  para la variable gasto, que es la principal variable que queremos explicar, es de 2.904.

#### 5.1.3. Conclusión Especifica 2

El modelo SARIMAX elegido obtenido  $\tilde{M}_1$  modificado basado en el modelo  $M_1$  los coeficientes salieron más significativos, lo que puede interpretarse que dicho modelo es mejor para realizar las predicciones futuras. El criterio de información de Akaike es de 374.44. También se obtiene que no se rechaza la hipótesis que los residuos tienen un comportamiento de ruido blanco. Respecto a la normalidad, de igual manera, los Q-Q plots son muy parecidos con los Q-Q plots obtenidos para los residuos de los modelos  $M_1$ . Aplicación de las pruebas de Shapiro-Wilk y Jarque-Bera resultaron en p valores de 0.162 y 0.119, respectivamente.



Para evaluar la eficiencia del modelo  $\tilde{M}_1$  modificado, en términos de los errores de predicción, fueron calculadas las predicciones bajo este modelo para los meses de enero-octubre del año 2018. El error cuadrático medio para el modelo  $\tilde{M}_1$  modificado para la variable gasto, que es la principal variable que queremos explicar, es de 3.627.

#### **5.1.4. Conclusión Especifica 3**

Para el Modelo Vectorial Autorregresivo (VAR) estimado se ha tratado de seleccionar la especificación más parsimoniosa (con menos rezagos) y de esa manera se ha dado cumplimiento a la prueba de estabilidad del modelo. Esta relación es de central interés en la gestión pública puesto que puede ayudar a explicar el gasto en el sector público ante un shock en el ingreso. Recordemos que la teoría económica sugiere que el gasto depende del ingreso. Según los resultados del análisis de raíz unitaria, tenemos que la aparente presencia de raíz unitaria en las series es generada por una componente estacional altamente persistente. En el caso de la prueba de normalidad multivariada, el test de Jarque-Bera multivariado y univariado nos da p-valores menores que  $2.2 \times 10^{-16}$ , sugiriendo que los residuos multivariados, así como de las marginales, siguen una distribución normal.

Para el caso del test de heterocedasticidad, este nos da un p-valor de 0.8566, sugiriendo la ausencia de heterocedasticidad en los residuos resultantes del ajuste. En el caso de la prueba de correlación serial, el test de Breusch-Godfrey nos da un p-valor de 0.1929, sugiriendo que la ausencia de correlación sería para los residuos después de ajustar un  $VAR(1)$ . El error cuadrático medio para el modelo ajustado para la variable gasto, que es la principal variable que queremos explicar, es de 4.972.

## 5.2. Recomendaciones

Luego de haber realizado el presente trabajo de investigación con sus resultados obtenidos, se considera las siguientes recomendaciones:

- \* La presente investigación demostró que la metodología de Box and Jenkins aplicando Modelos SARIMA dio mejor resultado porque permitió hallar una estimación más ajustada y un menor error de predicción, esto significó que nos acercamos más a la realidad de la política de la institución y del mercado, por lo que se sugiere desarrollar esta metodología para obtener modelos que predigan con menor error y además esta información sirva para la programación y formulación del presupuesto institucional del siguiente año.
- \* A través del Modelo SARIMA; lo más relevante está en el resultado obtenido para la predicción para el siguiente año. Con la experiencia laboral dentro de la institución facilita la interpretación de estos datos. Se sugiere que la Oficina de Planeamiento y Presupuesto de la institución maneje esta metodología.
- \* Al incluir en el modelo VAR variables exógenas aumenta significativamente el número de parámetros a estimar, además por ser muy complejos estos modelos pueden presentar problemas de convergencia; por lo que para estudios futuros se recomienda Modelos de Espacio de Estado.

## Referencias Bibliográficas

- Anderson O.T, *Identifying Box-Jenkins Processes, Computer Applications*, vol. 2, 1974, p. 275-283.
- Bartlett, M.S (1955). *Stochastic Processes*, Cambridge University Press, Cambridge.
- Box, G.E., Jenkins, G.M. (1970) *Time Series Analysis: Forecasting and Control*, Holden Day.
- Box, G.E. – Tiao, G.C. (1975) *Intervention analysis with applications to economic and environmental problems*, Journal of the American Statistical Association, Vol. 70, pp. 70-79.
- Box, G.E., Jenkins, G.M. (1976) *Time Series Analysis: Forecasting and Control*, Revised Edition, Holden Day.
- Box, G.E., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M. (1994) *Time Series Analysis: Forecasting and Control*, 3<sup>rd</sup> Ed., Prentice-Hall, Inc., New Jersey.
- Box, G.E., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M. (2016) *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, Hoboken.
- Caballero Díaz, F.F. (2011). Selección de Modelos mediante Criterios de Información en Análisis Factorial. Aspectos teóricos y Computacionales, [Tesis Doctoral, Universidad de Granada, España].
- Chong Fuentes, M. A. (2016). Proyección de series de tiempo para el consumo de la energía eléctrica a clientes residenciales en Ecuador [Tesis de Maestro, Universidad de Concepción. Chile].
- Cochrane, D. and Orcutt, G. (1949). Application of least squares to relations hips containing autocorrelated error terms, Journal of the American Statistical Association, Vol. 44, pp. 32-61.
- Córdova Baldeón, I. (2012). *Proyecto de investigación cuantitativa*. Lima: Editorial San Marcos.
- Cueto Naredo, R. (2015). Causalidad entre Ingresos y Gastos: Evidencia Empírica para España, [Trabajo fin de Master en Economía: Instrumento del Análisis Económico, Universidad de Oviedo, España].
- Enders, W. (2014). *Applied Econometric Time Series*. (4th ed). Wiley.
- Granger, C.W.J. (1981). Some Properties of Time Series Data and Their Use in Econometric Model Specification. Journal of Econometrics, 16, 121-130.
- Greene, W. (2012). *Econometric Analysis*. (7ma ed). Pearson.
- Gujarati, D. (2010). *Econometría*. (5ta ed). McGraw-Hill.
- Hamilton, J. (1994). *Time Series Analysis*. New Jersey: Princeton University Press.
- Hernández, R., Fernández, C. y Baptista, P. (2014). *Metodología de la investigación*. México: Mc Graw Hill.

- Jarque, Carlos M. and Bera, Anil K. (1980). *Efficient tests for normality, homoscedasticity and serial independence of regression residuals*. *Economics Letters* 6 (3): 255-259
- Karatzas y Shreve (2012). *Brownian Motion and Stochastic Calculus*. Springer-Verlag.
- Lastra Chacón, J.P. (2001). *Resumen Aplicativo sobre Econometría de Series de Tiempo*. Instituto Nacional de Estadística e Informática (INEI).
- Londoño W. (2005). *Modelos de Ecuaciones Múltiples Modelos VAR y Cointegración*. [Tesis de Maestro, Universidad EAFIT. Colombia].
- Lütkepohl, H. and Kratzig, M. (2004). *Applied Time Series Econometrics*. Cambridge University Press
- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer-Verlag Berlin.
- Mayurí Sánchez, J.L. (2015). *La Inversión en Infraestructura Pública y el Crecimiento Económico en el Perú, periodo 1950-2013*. [Tesis de Titulo, Universidad Nacional Agraria de la Molina. Perú].
- Montenegro, A. (2018). *Econometría Intermedia y Básica*. E-Books & Papers for Statisticians.
- Montgomery, D. (2015). *Introduction to Time Series Analysis and Forecasting*. (2th ed). Wiley.
- Novalés A. (2000). *Econometría*. McGraw-Hill.
- Ortega Huayta, E. (2004). *Proyección de las Importaciones Peruanas*. [Tesis de Licenciatura, Universidad Nacional de Ingeniería. Perú].
- Peña Sánchez de Rivera, D. (2010). *Análisis de Series Temporales*. Alianza Editorial S.A.
- Pérez López, C. (2011). *Econometría Avanzada*. Ibergarceta Publicaciones, S.L.
- Pindyck, Robert S. (2001). *Econometría: Modelos y Pronósticos*. McGraw-Hill.
- Ruey S. Tsay (2005). *Analysis of Financial Time Series*. Wiley & Sons, Publication
- Sims Christopher A. (Jan., 1980). *Macroeconomics and Reality*. *Econometrica*, Vol. 48, N° 1. Pp. 1-48
- Uriel, E. y Peiró, A. (2000). *Introducción al Análisis de Series Temporales*. Editorial AC.
- Tsay, Ruey S. (2005). *Analysis of Financial Time Series*. Wiley & Sons, Publication

### Fuentes Electrónicas

- Ministerio de Economía y Finanzas (2022). *Directiva N° 0002-2022-EF/50.01 – Resolución Directoral N° 0005-2022-EF/50.01*.  
 URL: <https://www.gob.pe/institucion/mef/normas-legales/2758928-0005-2022-ef-50-01>

BCRP (2023). Banco Central de Reserva del Perú: Gerencia Central de Estudios Económicos. Índice de precios Lima Metropolitana (índice 2009 = 100).

URL: <https://estadisticas.bcrp.gob.pe/estadisticas/series/mensuales/indice-de-precios-indice-2009-100>

MEF (2023). Ministerio de Economía y Finanzas: Portal de transparencia Económica. Seguimiento de la Ejecución Presupuestal (Consulta amigable). Consultar el Ingreso Presupuestal (Actualización Mensual). Consulta Amigable de Ingresos (Mensual). Presupuesto y Ejecución de Ingresos.

URL: <https://apps5.mineco.gob.pe/transparenciaingresos/mensual/>

Consultar el Gasto Presupuestal (Actualización Mensual)

Consulta Amigable (Mensual). Consulta de Ejecución del Gasto

URL: <https://apps5.mineco.gob.pe/transparencia/mensual/>