**UNIVERSIDAD NACIONAL DE INGENIERIA**

**ESCUELA CENTRAL DE POSGRADO**



**PLAN DE TESIS DE DOCTORAL**

**ESTUDIO METODOLOGICO DE LAS PROYECCIÓN DE LA DEMANDA DE GAS NATURAL EN LA PRODUCCION ELÉCTRICA NACIONAL MEDIANTE MODELOS ECONOMÉTRICOS Y ANALISIS DE SERIES TEMPORALES**

**PARA OBTENER EL GRADO DE DOCTOR EN CIENCIAS CON MENCION EN ENERGETICA**

**ELABORADO POR:**

**ROBERTO ÁLVAREZ CALLE**

**ASESOR**

**DR. SALOME GONZALES CHÁVEZ**

**COASESOR EXTERNO**

**DR. ANÍBAL JESÚS SALAZAR MENDOZA**

**LIMA – PERÚ**

**2023**

# DEDICATORIA

Este trabajo va dedicado a mi Madre, Dionisia Calle Luque y a mi Padre Juan Álvarez Cahuín, que fueron las personas que impulsaron y apoyaron la realización de mis estudios.

A mi esposa por el optimismo que me transmitió para seguir adelante.

A mis hijos: Liliana, Juan y Linsdy por su comprensión y cariño.

Y a mis familiares y amigos que tuvieron siempre unas palabras de aliento.

# AGRADECIMIENTO

En primer lugar, quiero agradecer a la Facultad de Ingeniería de Petróleo, por la formación recibida durante los años de estudio.

Esta tesis, si bien ha requerido un esfuerzo y dedicación, no se hubiera finalizado sin la cooperación desinteresada de todas las personas que a continuación citare:

Por último, hago extensivo este agradecimiento profundo y sincero para mi familia que, sin su apoyo y colaboración habría sido imposible llevar a cabo este trabajo.

# INDICE GENERAL

[DEDICATORIA 1](#_Toc127641793)

[AGRADECIMIENTO 2](#_Toc127641794)

[INDICE GENERAL 3](#_Toc127641795)

[INDICE DE FIGURAS 6](#_Toc127641796)

[INDICE DE TABLAS 8](#_Toc127641797)

[RESUMEN 9](#_Toc127641798)

[ABSTRACT 10](#_Toc127641799)

[CAPITULO I: ASPECTOS GENERALES 11](#_Toc127641800)

[1.1. INTRODUCCIÓN 11](#_Toc127641801)

[1.2. ANTECEDENTES 12](#_Toc127641802)

[1.4. FORMULACIÓN DEL PROBLEMA 15](#_Toc127641803)

[1.5. PROPUESTA DE INVESTIGACIÓN 15](#_Toc127641804)

[1.5.1. Objetivo General 15](#_Toc127641805)

[1.5.2. Objetivos Específicos 15](#_Toc127641806)

[1.6. ALCANCE 16](#_Toc127641807)

[1.7. HIPÓTESIS PARA EL PRESENTE ESTUDIO 16](#_Toc127641808)

[1.7.1 Hipótesis General 16](#_Toc127641809)

[1.7.2 Hipótesis Específicas 16](#_Toc127641810)

[1.8. METODOLOGÍA DE LA INVESTIGACIÓN 17](#_Toc127641811)

[1.8.1. Procesamiento de la Información 17](#_Toc127641812)

[1.8.2. Criterios de Procesamiento, Estructuración y Análisis de la Información 17](#_Toc127641813)

[CAPITULO II: MARCO TEORICO 18](#_Toc127641814)

[2.1 MODELO ECONOMETRICO DE REGRESION UNIECUACIONAL 18](#_Toc127641815)

[2.1.1 Series de tiempo 18](#_Toc127641816)

[2.1.2. Procesos estocásticos 19](#_Toc127641820)

[2.1.3. El fenómeno de regresión espuria 22](#_Toc127641832)

[2.1.4. Prueba de estacionariedad 22](#_Toc127641833)

[2.1.5. Cointegración 25](#_Toc127641834)

[2.1.6. Modelo de Corrección de Errores (ECM) 26](#_Toc127641835)

[2.1.7. Causalidad de Grenger 27](#_Toc127641836)

[**2.2 MODELO UNIVARIANTE ARIMA** 29](#_Toc127641837)

[2.2.1. Enfoque y procesos estocásticos 29](#_Toc127641838)

[2.2.2. Modelos AR, MA y mixtos 30](#_Toc127641839)

[2.2.3. Metodología Box-Jenkins No estacional 34](#_Toc127641840)

[2.2.4. Metodología Box-Jenkins Estacional 60](#_Toc127641841)

[2.2.5. Ventajas y desventajas de la modelación ARIMA 64](#_Toc127641842)

[2.2.6. Modelos ARIMA con variables de Intervención 66](#_Toc127641843)

[2.2.7. Modelos ARMA con variables externas (ARMAX) 66](#_Toc127641844)

[CAPITULO III: METODOLOGIA, FORMULACION DE LOS MODELOS PROYECTIVOS Y ANALISIS DE LAS VARIABLES OBJETIVOS 67](#_Toc127641848)

[3.1. DATOS 67](#_Toc127641849)

[3.2. MODELO DE REGRESIÓN UNIECUACIONAL 67](#_Toc127641850)

[3.2.1. Prueba de estacionariedad 69](#_Toc127641851)

[3.2.2. Colinealidad entre variables explicativas 69](#_Toc127641852)

[3.2.3. Procedimiento de cointegracion: Engle-Granger 70](#_Toc127641853)

[3.2.4. Causalidad de Granger 71](#_Toc127641854)

[3.3. ANALISIS DE LAS VARIABLES OBJETIVOS 71](#_Toc127641855)

[3.3.1. Variable Dependiente: demanda de gas natural seco para generadores eléctricos 72](#_Toc127641860)

[3.3.2. Variable Independiente: Producto Bruto Interno 73](#_Toc127641861)

[3.3.3. Variable Independiente: Precio del gas en Boca de Pozo 75](#_Toc127641862)

[3.3.4. Variable Independiente: Precio del Petróleo 75](#_Toc127641863)

[3.4.5. Variable Independiente: Producción Total de gas natural seco de Camisea 76](#_Toc127641864)

[3.3.6. Variable Independiente: Producción Eléctrica Total del SEIN 76](#_Toc127641865)

[3.3.7. Variable Independiente: Producción Eléctrica con gas natural seco de Camisea 77](#_Toc127641866)

[3.4. MODELO UNIVARIANTE ARIMA 78](#_Toc127641867)

[3.4.1. Origen y estructura de los datos 78](#_Toc127641868)

[3.4.2. Procedimiento 78](#_Toc127641869)

[CAPITULO IV: ANALISIS Y RESULTADOS DE LA INVESTIGACION 80](#_Toc127641870)

[4.1. PROYECCION DEL MODELO DE REGRESION UNIECUACIONAL 80](#_Toc127641871)

[4.1.1. Importación del archivo Excel al programa Eviews y definición de variables 80](#_Toc127641872)

[4.1.2. Especificación Funcional 80](#_Toc127641873)

[4.1.3. Visualización de la función de cada una de las variables 81](#_Toc127641874)

[4.1.4. Transformación de las variables 82](#_Toc127641875)

[4.1.5. Prueba de estacionariedad 84](#_Toc127641876)

[4.1.6. Determinación de colinealidad entre variables explicativas 86](#_Toc127641877)

[4.1.7. Modelo Econométrico Uniecuacional 87](#_Toc127641878)

[4.1.8. Análisis de cointegración 88](#_Toc127641879)

[4.1.9. Proyección dinámica de largo plazo de la demanda de gas para G.E. 93](#_Toc127641880)

[4.2. PROYECCION DEL MODELO UNIVARIANTE ARIMA EN RESOLUCION MENSUAL 96](#_Toc127641881)

[4.2.1. Paso 1: Prueba de estacionariedad 96](#_Toc127641882)

[4.2.2. Paso 2: Identificación del modelo ARIMA 102](#_Toc127641883)

[4.2.3. Paso 3 Estimación del modelo identificado 106](#_Toc127641884)

[4.2.4. Paso 4: Evaluación 110](#_Toc127641885)

[4.2.5.Paso 5: Proyección 112](#_Toc127641886)

[4.2.6. Comparación de los modelos Ecométrico y ARIMA 119](#_Toc127641888)

[CAPITULO V: CONCLUSIONES Y RECOMENDACIONES 121](#_Toc127641889)

[5.1 CONCLUSIONES 121](#_Toc127641890)

[5.1.1. Modelo econométrico 121](#_Toc127641891)

[5.1.2. Modelo ARIMA 121](#_Toc127641892)

[5.2. RECOMENDACIONES 122](#_Toc127641893)

[BIBLIOGRAFIA 124](#_Toc127641894)

# INDICE DE FIGURAS

[**Figura 1:** **Producción y consumo mundial de natural por regiones (bcm)** 11](#_Toc101066232)

[**Figura 2: Producción de energía eléctrica nacional enero 2021** 12](#_Toc101066233)

[**Figura 3: Demanda de GN seco para GE desde el año 2005 al 2020 en forma anual** 17](#_Toc101066234)

[**Figura 4: Proyección del PBI desde el año 2005 al 2025 en forma anual** 18](#_Toc101066235)

[**Figura 5: Proyección del Precio del gas en boca de pozo** 19](#_Toc101066236)

[**Figura 6: Proyección de la Producción Total de G.N. seco** 20](#_Toc101066237)

[**Figura 7: Proyección de la Producción Eléctrica Total del SEIN** 20](#_Toc101066238)

[**Figura 8: Proyección de la Producción Eléctrica con G.N. seco** 21](#_Toc101066239)

[**Figura 9: Representación de una serie temporal** 27](#_Toc101066240)

[**Figura 10: Descomposición de una serie temporal en sus componentes** 28](#_Toc101066241)

[**Figura 11: Proceso estocástico estacionario** 29](#_Toc101066242)

[**Figura 12: Función de autocorrelación parcial** 30](#_Toc101066243)

[**Figura 13: función de cada una de las variables** 38](#_Toc101066244)

[**Figura 14 Variables con la función logaritmo** 39](#_Toc101066245)

[**Figura 15: Variables con la función diferencia de logaritmo estacionalizadas** 40](#_Toc101066246)

[**Figura 16: Estimación de la prueba de raíz unitaria al logaritmo de las variables** 42](#_Toc101066247)

[**Figura 17: Estimación de la prueba de raíz unitaria a la diferencia de logaritmos** 44](#_Toc101066248)

[**Figura 18: Estimación de la ecuación de prueba** 45](#_Toc101066249)

[**Figura 19: Atípico del año 2020** 46](#_Toc101066250)

[**Figura 20: Ecuación de largo plazo estimada** 47](#_Toc101066251)

[**Figura 21: Ecuación base** 48](#_Toc101066252)

[**Figura 22: Ecuación de corrección de error** 49](#_Toc101066253)

[**Figura 23: Ecuación de proyección estimada** 50](#_Toc101066254)

[**Figura 24: Proyección dinámica de la demanda de gas natural seco para G.E.** 51](#_Toc101066255)

[**Figura 25: Registro mensual de la variable demanda** 54](#_Toc101066256)

[**Figura 26: Muestra la evolución de la demanda de gas natural para G.E.** 55](#_Toc101066257)

[**Figura 27: funciones de autocorrelación (FAC)** 56](#_Toc101066258)

[**Figura 28: Dispersión versus nivel de DEM por año para el cálculo de λ** 57](#_Toc101066259)

[**Figura 29: Raíz cuadrada de la demanda versus el tiempo** 58](#_Toc101066260)

[**Figura 30: Grafico de frecuencia del estadístico descriptivo d1D1** 59](#_Toc101066261)

[**Figura 31: Configuración en el SPSS para obtener la orden regular** 60](#_Toc101066262)

[**Figura 32: Autocorrelación de la parte regular** 60](#_Toc101066263)

[**Figura 33: Modelos AR y MA** 61](#_Toc101066264)

[**Figura 34: Configuración en el SPSS para la orden estacional** 62](#_Toc101066265)

[**Figura 35: Autocorrelación de la parte estacional** 62](#_Toc101066266)

[**Figura 36: Error para modelo ARIMA optimo** 68](#_Toc101066267)

[**Figura 37: Curva de demanda histórica vs pronostico a diciembre 2020** 69](#_Toc101066268)

[**Figura 38: Pronostico de la demanda eléctrica para generadores con gas natural** 70](#_Toc101066269)

# INDICE DE TABLAS

[**Tabla 1: Producción de energía eléctrica nacional según destino y fuente (GWh)** 12](#_Toc101067727)

[**Tabla 2: Identificación de modelos** 35](#_Toc101067728)

[**Tabla 3: Series de las variables que se utilizaran** 38](#_Toc101067729)

[**Tabla 4: Matriz de correlación de variables con logaritmos** 45](#_Toc101067730)

[**Tabla 5: Matriz de correlación de variables con diferencia de logaritmos** 46](#_Toc101067731)

[**Tabla 6: Estadísticos descriptivos** 59](#_Toc101067732)

[**Tabla 7: Modelos Identificados** 64](#_Toc101067733)

[**Tabla 8: Estimación de Parámetros incluyendo la constante** 65](#_Toc101067734)

[**Tabla 9: Estimación de parámetros del modelo incluyendo la constante** 66](#_Toc101067735)

[**Tabla 10: Resumen de los parámetros del modelo incluyendo atípicos** 67](#_Toc101067736)

[**Tabla 11: Estimación de parámetros** 68](#_Toc101067737)

[**Tabla 12: Autocorrelaciones, modelo optimo** 69](#_Toc101067738)

[**Tabla 13: Se muestra la demanda histórico real y la predicción mediante ARIMA** 71](#_Toc101067739)

[**Tabla 14: Comparación de MAPE** 77](#_Toc101067740)

# RESUMEN

La presente tesis presenta una alternativa para la proyección de la demanda de gas natural para generadores eléctricos en el Perú a mediano plazo del 2022 al 2026. Mediante el modelo Econométrico en resolución anual y el modelo ARIMA en resolución mensual. La disponibilidad de energía es crucial para lograr un proceso de crecimiento y desarrollo sustentable en el tiempo. Para poder administrar y generar políticas energéticas es necesario disponer de modelos confiables de proyección de la demanda de energía.

Para el modelo econométrico se relacionó la demanda sectorizada del consumo de gas natural seco de Camisea para generadores eléctricos con el PBI del Perú, la producción total de gas seco de Camisea, la producción eléctrica total del SEIN, la producción eléctrica con gas de Camisea y el precio tarifario en boca de pozo para generadores eléctricos y para el modelo ARIMA solo se relaciona la demanda de gas natural seco, para ello todas las variables participantes tienen el mismo tamaño histórico de diecisiete años, del 2005 al 2021, la metodología de cálculo lo constituye el tratamiento de cada serie temporal objetivo, según el modelo, mediante transformaciones estadístico-matemáticas apropiadas, hasta alcanzar el modelo predictivo para la demanda de gas natural para generadores eléctrico en el Perú.

Utilizando los modelos de proyección econométrico y ARIMA, se estimó el rendimiento del pronóstico de cada uno de modelos, los resultados indican que el modelo econométrico presenta un mejor ajuste que el modelo ARIMA.

**Palabras claves**: demanda de gasnatural, modelo econométrico, cointegracion, centrales termoeléctricas, sistema interconectado.

# ABSTRACT

This thesis presents an alternative for the projection of the demand for natural gas in Peru for electric generators in the medium-long term from 2020 to 2025. Through the Econometric projection models in annual resolution and the ARIMA projection model in monthly resolution. The availability of energy is crucial to achieve a sustainable growth and development process over time. In order to manage and generate energy policies, it is necessary to have reliable models for projecting energy demand.

For the econometric model, the sectorized demand for the consumption of dry natural gas from Camisea for electricity generators was related to the total production of dry gas from Camisea, the total electricity production from the SEIN, the electricity production from gas from Camisea, the tariff price at of well, the GDP of Peru and for the ARIMA model only the univariate, for this all the participating variables have the same historical size of fifteen years, from 2005 to 2021, the calculation methodology is the treatment of each objective time series, according to the model, through appropriate statistical-mathematical transformations, until reaching the predictive model for gas demand for electricity generators in Peru.

Using the econometric and ARIMA prediction models, the forecast performance of each of them was estimated, the results indicate that the econometric model has a better fit than the ARIMA model.

**Keywords:** natural gas demand, cointegration model, projection, thermoelectric plants, interconnected system.

# CAPITULO I: ASPECTOS GENERALES

## 1.1. INTRODUCCIÓN

La proyección de acontecimientos necesita un análisis minucioso de eventos pasados y las relaciones existentes entre ellos para de ahí intentar extrapolar futuros acontecimientos. Con el transcurso del tiempo es evidente el desarrollo de nuevas tecnologías que ha mejorado y ampliado en diferentes áreas la predicción de eventos, ya sea con fines científicos o económicos. Hoy en día se cuenta, en la mayoría de los sectores productivos, con bases de datos históricas que almacenan información valiosa sobre su actividad [4] [5].

Uno de los principales campos de investigación en mercados energéticos es el modelado de diferentes series que incluyen los precios de los energéticos, su demanda en diferentes escalas de tiempo y muchas otras series relacionadas, con el fin de comprender mejor los diferentes hitos históricos que explican sus fluctuaciones en el tiempo. Además de comprender el comportamiento de las series, desde un punto de vista pragmático, deben desarrollarse modelos para pronosticar su evolución [20].

Para objetivos fundamentales académicos, hacer una tesis planteando un nuevo modelo econométrico, para analizar y diagnosticar el mercado de gas natural en el Perú, utilizando técnicas basadas en la teoría económica y en la formulación estadístico matemático, y se plantee un método de naturaleza econométrico para evaluar el mercado del gas natural, complementariamente se desarrolle una evaluación mediante análisis univariante de series temporales para poder hacer comparaciones en el nivel de ajuste que tenga la metodología, respecto a los resultados del análisis proyectivo econométrico que se realice, con esto se obtendría potenciar los alcances del aprendizaje y la investigación aplicada [13].

La determinación del consumo de gas natural de Camisea para la producción de energía eléctrica para los años futuros conlleva a realizar cálculos del despacho de generación de las centrales que se encuentran conectadas al ducto del gas de Camisea. Estos cálculos se realizan a través de simulaciones de optimización con información pública disponible en el Comité de Operación Económica del Sistema Interconectado (COES), el Organismo Supervisor de la Inversión en Energía y Minería (OSINERGMIN), MEM, Perupetro y TGP. Las centrales térmicas que se alimentan con gas natural de Camisea y con mayor producción de energía son Kallpa, Fénix, Chilca I, Ventanilla y Santo Domingo de Olleros [2] [12] [17].

## 1.2. ANTECEDENTES

Zarate I. et al (2015) en su paper “Análisis económico del consumo de gas natural para generación de energía eléctrica en México: 2005-2014” presentan los resultados obtenidos de un análisis econométrico realizado al consumo de gas natural para la generación de energía eléctrica en México con información anual de 2005 a 2014 [21].

Entre los resultados que se encontraron, es que el consumo citado en el país es inelástico (-0.0461); en la actualidad el combustóleo y el diésel son las fuentes alternas directas al gas natural para generación de energía eléctrica en México, no son sustitutos reales. El diésel no puede competir en costos por kW/h generado al gas natural y el combustóleo lo hace, pero genera altas emisiones de contaminantes.

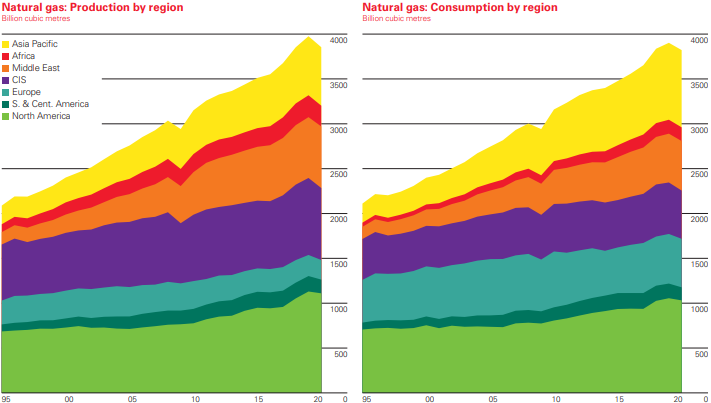
En un estudio realizado por Gil Salvador (2007), “proyección de la demanda de gas para mediano y largo plazo” presenta los resultados de proyecciones de consumo de gas natural en la Argentina a mediano (3 a 5 años) y largo plazo (6 a 25 años). En particular se presenta la fundamentación de los métodos utilizados para proyectar los consumos de gas residenciales, comerciales, GNC, industriales y de centrales eléctricas [8].

Por otra parte, Gonzales G. et al (2017) realizaron un estudio “Estimation of Residential Natural Gas Consumición in Medellín-Antioquia” presentan una alternativa para pronosticar el consumo de gas natural en Medellín (Antioquia), la segunda ciudad más grande de Colombia.

El conjunto de datos abarcó el primer trimestre de 2003 hasta el segundo trimestre de 2014, utilizando las predicciones de los modelos ARIMA y Support Vector Machine, se estimó el rendimiento del pronóstico. Los resultados indican que el modelo ARIMA presenta un mejor ajuste que el modelo SVM en términos de RSME y MAE.

Gil S. & Deferrari J. (1999) en su paper” Modelo de Predicción del Consumo de Gas Natural de la República Argentina”, el modelo es de carácter general y fue exitosamente probado en el Gran Buenos Aires y las ciudades de Córdoba, Mendoza, La Plata, Bahía Blanca y Neuquén. El mismo tiene la capacidad de predecir los consumos ininterrumpibles con una incerteza menor del 10%, en 90% de los días del año, además posee una gran robustez en la predicción del consumo ya que se basa parcialmente tanto en los pronósticos térmicos como en los escenarios preexistentes.

**1.3. SITUACIÓN ACTUAL**

El consumo mundial de gas natural decreció en 2% o 81 billones de metros cúbicos (bcm) en 2020, similar a la caída en 2009 durante la crisis financiera. El consumo cayo en muchas regiones, a excepción de China donde la demanda creció en 6.9%. en contraste a la demanda de gas en norte América y Europa donde cayó en 2.6% y 2.5% respectivamente. La producción cayó en 123 bcm (-3.3%), la más grande caída vista en Rusia (-41 bcm) y los US (-15 bcm), como se muestra en la Figura 1 [19].

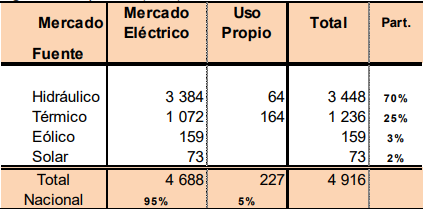
**Figura 1:** **Producción y consumo mundial de natural por regiones (bcm)**

**Fuente**: PB Statistical Review of Word Energy 2021

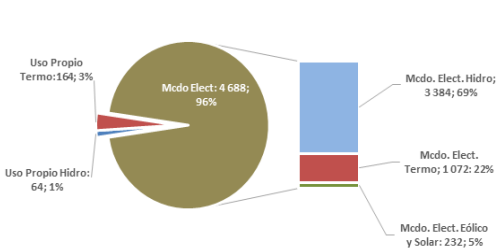
En el Perú la producción total de energía eléctrica a nivel nacional registrada en el mes de enero 2021, incluyendo los Sistemas Aislados y SEIN, fue 4,916 GWh ver Tabla 1 y Figura 2, valor 1,1% menor al mes de enero del año pasado. Las unidades que participan del Sistema Eléctrico Interconectado Nacional – SEIN generaron 4,6 GWh, y fue 1% menor que en similar mes de 2020 [15].

Asimismo, la producción en la generación no convencional, las centrales solares produjeron 74 GWh, 26% mayor a enero 2020, y las unidades eólicas generaron 159 GWh, superando en 41% lo generado en enero del año anterior.

**Tabla 1: Producción de energía eléctrica nacional enero 2021 según destino y fuente (GWh)**



**Fuente:** **Principales indicadores del sector eléctrico a nivel nacional [15]**

******

**Figura 2: Producción de energía eléctrica nacional enero 2021. Total: 4916 GWh**

**Fuente:** **Principales indicadores del sector eléctrico a nivel nacional**

El modelado y análisis de las relaciones entre las variables económicas financieras en diferentes mercados de energía (electricidad, petróleo, gas, carbón, biocombustibles, etc.) son dos de los principales ejes de investigación de la economía energética:

Las metodologías utilizadas provienen de diversos campos del conocimiento que incluyen, entre otros, la economía, la econometría, la investigación de operaciones, la estadística y las finanzas. Gran parte del análisis está centrado en descubrir y entender la relación entre diferentes variables, así como su dinámica, a partir de la información histórica [20].

## 1.4. FORMULACIÓN DEL PROBLEMA

Son muy pocos los estudios que tratan de pronosticar la demanda de gas natural a través de modelos cuantitativos, de tal manera que la programación de la producción de gas natural atienda a las verdaderas necesidades del mercado nacional.

En este sentido, El modelo que se determine para el análisis de la variable objetivo, dependerá de las características intrínsecas temporales de cada una de las variables explicativas elegidas en principio, lo cual mediante el análisis de especificación funcional y el cumplimiento de las condicionantes estadísticas y econométricas, llevarán a la selección final de una expresión multivariable uniecuacional o biecuacional con las variables definitivas [7]. En cuando a antecedentes investigativos en torno al presente estudio, a nivel del Perú no se tiene referencia.

Para poder administrar y generar políticas energéticas es necesario disponer de modelos confiables de proyección de la demanda de energía. Estos modelos deben ser capaces de predecir los consumos en el corto plazo (pocos días) y también a largo plazo (algunos años). Por otra parte, es necesario predecir los picos de demanda a corto plazo producidos en general por condicionantes climáticos [8].

## 1.5. PROPUESTA DE INVESTIGACIÓN

### 1.5.1. Objetivo General

Comparar las dos metodologías de cálculo de proyección de demanda de gas natural atendiendo a la exactitud de cálculo de proyección de demanda de gas natural en el Perú

### 1.5.2. Objetivos Específicos

1) Determinar el nivel de exactitud de la metodología de cálculo de proyección de demanda de gas natural para generadores eléctricos de tipo Econométrico con la proyección de demanda de gas natural seco para generadores eléctricos real.

2) Determinar el nivel de exactitud de la metodología de cálculo de proyección de demanda de gas natural seco para generadores eléctricos del modelo tipo ARIMA con la proyección de demanda de gas natural seco para generadores eléctricos real.

## 1.6. ALCANCE

En este trabajo de investigación, el aporte constituye el estudio metodológico de la proyección de la demanda de gas natural en la generación eléctrica interconectada del Perú, utilizando técnicas econométricas y análisis de series temporales, con la finalidad de proyectar su comportamiento en el mediano a largo plazo con un nivel de confianza más aproximado estadísticamente.

El mediano a largo plazo, se acota a un nivel de cinco años dado el tamaño muestral de cada una de las variables participantes, las mismas que se han ordenado desde enero del 2005 a diciembre del 2021.

## 1.7. HIPÓTESIS PARA EL PRESENTE ESTUDIO

**1.7.1 Hipótesis General**

La metodología de cálculo de proyección de demanda de gas natural seco de Camisea para generadores eléctricos, permitirá el nivel de exactitud de proyección de demanda de gas natural en el Perú.

**1.7.2 Hipótesis Específicas**

1) La metodología de cálculo de proyección de demanda de gas natural seco de Camisea para generadores eléctricos del modelo tipo Econométrico permitirá una mayor exactitud en el cálculo de proyección de demanda de gas natural seco de Camisea para generadores eléctricos en el Perú.

2) La metodología de cálculo de proyección de demanda de gas natural seco para generadores eléctricos del modelo tipo ARIMA permitirá una mayor exactitud en el cálculo de proyección de demanda de gas natural seco para generadores eléctricos en el Perú.

**Variables Independientes**

* PGT: Producción total de gas natural seco de Camisea (MMPC)
* PET: Producción eléctrica total del SEIN (GWh)
* PEG: Producción eléctrica del gas natural seco de Camisea (GWh)
* PBI: Producto bruto interno del Perú (MMS)
* TAR: Tarifa de gas natural seco en boca de pozo para generadores eléctricos (US$/MMBTU)

**Variable Dependiente**

* DEM: Demanda de gas natural seco de Camisea para generadores eléctricos (MMPC)

## 1.8. METODOLOGÍA DE LA INVESTIGACIÓN

La presente tesis se realizó con la metodología, inductivo y deductivo que tiene como objetivo desarrollar la metodología econométrica de cálculo de proyección de demanda de gas natural seco para generadores eléctricos atendiendo a la exactitud de cálculo de proyección de demanda de gas natural seco para generadores eléctricos en el Perú.

**1.8.1. Procesamiento de la Información**

Durante la etapa del proceso de proyección de la demanda, se ha realizado la recopilación de información histórica de la demanda y de variables macroeconómicas del Perú.

Las actividades de procesamiento y análisis de la información histórica fueron las siguientes: 1) Recopilación de información; 2) Validación de la información recopilada; 3) Revisión y consolidación de la Información; y 4) Elaboración de los cuadros resúmenes.

**1.8.2. Criterios de Procesamiento, Estructuración y Análisis de la Información**

Los criterios seguidos en el procesamiento, estructuración y análisis de la información

histórica fueron los siguientes:

* La fuente de la información procesada está basada fundamentalmente en los registros estadísticos mensuales y anual proporcionados por el COES, OSINERGMIN, Perupetro, BCRP, por lo que durante las proyecciones se ha mantenido el patrón de comportamiento de consumo del año 2005 al 2021.

Demanda de gas para generadores eléctricos por empresas: definidas como el consumo de combustibles gaseosos (en millones de metros cúbicos) de las que se dispone de información en la estadística de operación del COES, desde 2005 hasta diciembre del 2021 en forma mensual y anual.

**CAPITULO II: MARCO TEORICO**

Para abordar el análisis general que se pretende realizar en este estudio es necesario recurrir a diferentes herramientas y modelos econométricos que ayudaran a visualizar las distintas características de las series en estudio y la relación existente entre ellas. En base a ello se definirán algunos términos que se utilizarán durante el desarrollo del estudio, así como sus utilidades y propósitos de su uso.

En términos generales, hay cuatro enfoques para la predicción económica basados en series de tiempo:

(1) Los modelos de regresión uniecuacionales, (2) los modelos de regresión de ecuaciones simultáneas, (3) los modelos autorregresivos integrados de media móvil (ARIMA) y (4) los modelos de vectores autorregresivos (VAR). En este estudio solo se utilizará el modelo de regresión uniecuacional y el modelo autorregresivo integrados de media móvil (ARIMA)

Garcia\_Luis\_I\_MC\_Economia\_2017 ModeloRegrer.Uniecuacional-TeoriaRegresionUniecuacional)(UTFSMChile)

**2.1 MODELO ECONOMETRICO DE REGRESION UNIECUACIONAL**

En los modelos de regresión uniecuacional, una variable llamada dependiente, es expresada como una función lineal de una o más variables, llamadas explicativas. En modelos de este tipo se supone implicitante que, si existen relaciones causales entre las variables dependientes y explicativas, éstas van en una dirección solamente: de las variables explicativas hacia la variable dependiente (Gujarati & Porter, 2004).

**2.1.1 Series de tiempo**

Una serie de tiempo es un conjunto de datos, valores u observaciones realizados a lo largo de tiempo (Fernández, 1998). En otras palabras, consiste en una sucesión estadística de los valores de una variable en el tiempo, donde se tiene que la variable endógena (Y) es explicada por el tiempo (t), de este modo se tiene una serie de la forma:

; t = 1, 2, . . . si se habla de variables discretas o

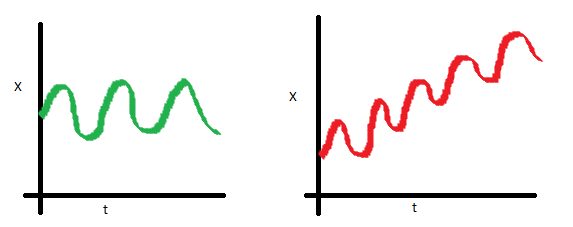
Y(t); t = 1, 2, . . . sí son continuas

**2.1.2. Procesos estocásticos**

Según Gujarati & Porter (2004), un proceso estocástico o aleatorio es una colección de variables aleatorias ordenadas en el tiempo. La teoría clásica considera que una serie de tiempo posee cuatro características que la definen y componen:

* Tendencia: existe cuando una serie histórica tiende a aumentar o a disminuir sus valores medios con el tiempo.
* Estacionalidad: existe cuando una serie de tiempo fluctúa de acuerdo con un factor temporal que suele depender de un determinado período.
* Ciclicidad: Comportamientos interanuales de la serie.
* Aleatoriedad: Movimientos esporádicos de las series de tiempo

Como se puede apreciar en la Figura 3, la serie temporal del lado izquierdo presenta una media constante, mientras que la de la derecha presenta tendencia, con una media que se incrementa con el paso del tiempo.

****

**Figura 3: Serie estacionaria y serie no estacionaria**

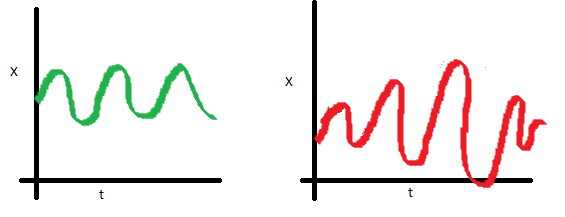
**Fuente:** web Analytics Vidhya

**a. Proceso estocástico estacionario**

Se define un proceso estocástico estacionario si se cumple que su media y su varianza son constantes en el tiempo, además de que el valor de la covarianza entre periodos distintos dependa solamente del rezago k. se expresa matemáticamente de la siguiente forma:

* Media constante:
* Varianza constante:
* Covarianza en función del rezago k:

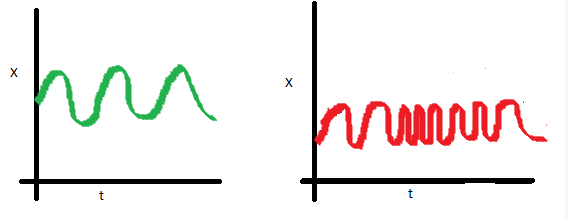
Como observamos, en la Figura 4, la gráfica del lado izquierdo es estacionaria, mientras que la del lado derecho presenta una variabilidad (varianza) que aumenta con el nivel de la serie, lo que se conoce como heterocedasticidad.

****

**Figura 4: Serie estacionaria y serie no estacionaria**

**Fuente:** web Analytics Vidhya

En la Figura 5, se observa una serie estacionaria en el lado izquierdo, y en el derecho una serie con autocovarianza que está en función del tiempo.

****

**Figura 5: Serie estacionaria y serie no estacionaria**

**Fuente: web Analytics Vidhya**

**b. Proceso estocástico no estacionario**

Por el contrario, un proceso estocástico no estacionario como es el caso de la caminata aleatoria o "random walk". Existen dos tipos; sin deriva y con deriva. En el primero se supone que el error es aleatorio o también llamado "ruido blanco", en cambio, en el segundo se incorpora un parámetro de variación . Se dice que el precio del petróleo tanto como los tipos de cambios siguen un modelo de caminata aleatoria (MCA) ya que no se puede predecir con exactitud su valor en el tiempo, lo mismo sucede con la mayoría de los precios de las series financieras.

* Caminata aleatoria sin deriva (Sin termino constante o de intercepto)
* Caminata aleatoria con deriva (con termino constante)

En resumen, el MCA, con o sin deriva, es un proceso estocástico no estacionario (Gujarati & Porter, 2010). Para ambos modelos se puede incorporar un análisis de primeras diferencias donde se obtiene que, a pesar de que ambos modelos no son estacionarios, su primera diferencia sí lo es.

* Primera diferencia de caminata aleatoria sin deriva
* Primera diferencia de caminata aleatoria con deriva

**c. Procesos estocásticos integrados**

El modelo de caminata aleatoria es un caso particular de una clase más general de procesos estocásticos conocidos como procesos integrados. El MCA sin deriva es no estacionario, pero su serie de primeras diferencias es estacionaria. Por lo que el MCA sin deriva se llama proceso integrado de orden *1* y se denota como ***I***(1). Un proceso integrado se refiere al orden de diferenciación donde la serie temporal resulta por lo menos débilmente estacionaria, entonces si la variable debe diferenciarse *d* veces para lograr estacionariedad, la serie es integrada de orden . Cuando es estacionario significa que la serie es integrada de orden uno o su primera diferencia es integrada de orden cero . Si la serie en su nivel () resulta estacionaria, sin diferenciarla, se denomina integrada de orden cero (Rosales et al., 2010).

**2.1.3. El fenómeno de regresión espuria**

Existe una regresión espuria cuando en un modelo de series de tiempo no estacionaria, los resultados indican que existe una relación significativa cuando en realidad no existe. Supongamos dos modelos de caminata aleatoria, como en el experimento de Granger y Newbold (1974):

Además

Dado que y están generadas de forma independientes se esperaría que no existiera ninguna relación. Sin embargo, cerca de un 75% de las regresiones de sobre presentan contrastes “*t*” significativos a un nivel de significancia del 5%.

Lo anterior resume el fenómeno de regresión espuria o regresión sin sentido descubierto por Yule (1926), quien mostro además que la correlación (espuria) puede persistir en las series de tiempo no estacionarias, aunque la muestra sea muy grande.

También otra definición de correlación espuria es que dos variables que se mueven a lo largo del tiempo, aun cuando tengan tendencia creciente (I (I)), están cointegradas si existe una combinación lineal de ellas que sea estacionaria (I (0)). Es así como un análisis de cointegracion muestra la relación de largo plazo entre las variables en estudio y si una combinación estacionaria entre estas. (Granger, 1981; Engle y Granger, 1987).

**2.1.4. Prueba de estacionariedad**

**Prueba grafica**

Antes de realizar una prueba formal, es recomendable graficar la serie de tiempo. Si la serie muestra tendencia entonces sugiere una variación en la media, lo cual puede indicar la no estacionariedad.

**Función de autocorrelación (FAC) y correlograma**

La FAC en el rezago k, denotada por , se define como la covarianza en el rezago k entre la varianza, de forma que:

Como la covarianza y varianza se miden en las mismas unidades, es un numero sin unidad de medida, o puro. Se encuentra entre -1 y +1, igual que cualquier coeficiente de correlación. Si graficamos respecto de k, la grafica se conoce como correlograma poblacional. Sin embargo, solo se puede calcular la función de autocorrelación muestral . Se calcula primero la covarianza muestral en el rezago *k*, la varianza muestral definidas como.

Donde *n* es el tamaño de la muestra y es la media muestral. Por lo tanto, la función de autocorrelación muestral en el rezago *k* es:

La gráfica de () con respecto a (*k*), es por lo tanto el correlograma muestral.

Para un proceso de ruido blanco, las autocorrelaciones en distintos rezagos se ubican alrededor de cero. En un correlograma de una serie de tiempo no estacionaria, el coeficiente de autocorrelación comienza a un nivel cercano a uno y permanece fuerte conforme los rezagos se incrementan.

Una regla para elegir la longitud del rezago consiste en calcular la FAC hasta un tercio o una cuarta parte de la longitud de la serie de tiempo. Otra es comenzar con rezagos grandes y reducirlos mediante el criterio de información de Akaike o de Schwarz (Gujarati y Porter, 2010).

**Prueba de raíz unitaria**

Una forma de determinar si una serie de tiempo es estacionaria es aplicando la prueba de raíz unitaria, la que consiste en determinar si la primera diferencia de una serie es o no ruido blanco. Para ello se toma como punto de partida una serie de la forma:

Donde −1 ≤ *ρ* ≤ 1 y si *ρ* = 1 se obtiene un modelo de caminata aleatoria sin variaciones, el cual no es estacionario. Si se calcula la primera diferencia de la serie mencionada anteriormente, tenemos:

Con esta última expresión podemos establecer que si = 0 existe raíz unitaria y por lo tanto la serie no es estacionaria. Al contrario, si < 0 se dice que la serie es estacionaria.

Para probar sí es cero o no se utiliza la prueba de Dickey-Fuller, la cual se basa en que el coeficiente se ajusta a una distribución (tau) conocida simplemente como la tabla Dickey-Fuller. Por otro lado, si no es cero, se puede utilizar la prueba t de Student.

**Prueba Dickey-Fuller Aumentada (DFA)}**

Esta prueba desarrollada por Dickey-Fuller se aplica cuando el termino de error  esta afectado por correlación serial y se lleva a cabo a través de la siguiente especificación:

Donde es un término de error puro de ruido blanco y donde , etc. El número de términos de diferencia rezagados (m) que se deben incluir se determina de manera empírica, con la idea de incluir los términos suficientes para que en la ecuación no este serialmente relacionado y sea posible obtener una estimación insesgada de . En la DFA se sigue probando =0 y sigue la misma distribución asintótica que el estadístico DF, por lo que se pueden utilizar los mismos valores críticos (Gujarati y Porter, 2010).

**Prueba de raíz unitaria Phillips-Perrón (PP)**

Un supuesto importante de la prueba DF es que los términos de error están idéntica e independientemente distribuidos. La prueba de DFA ajusta la prueba DF a fin de tener cuidado de una posible correlación serial en los términos de error al agregar los términos *d* diferencias rezagados de las variables dependientes. Phillips y Perrón utilizan métodos estadísticos no paramétricos para evitar la correlación serial en los términos de error, sin añadir términos de diferencias rezagados. La distribución asintótica de la prueba PP es la misma que la prueba DFA (Gujarati y Porter, 2010).

**2.1.5. Cointegración**

El análisis de cointegración es esencial cuando se tiene una combinación de variables que presentan una similitud en el orden de integración, en especial, cuando las series de tiempo son ***I***(1). Si lo anterior se cumple y una combinación lineal de estas variables es estacionaria, es decir ***I***(0), entonces se dice que las variables están cointegradas y existe una regresión cointegrante. En términos económicos, dos variables están cointegradas si existe una relación de largo plazo, o de equilibrio, entre ambas (Gujarati y Porter).

La cointegración, fue enunciado por (Granger, 1981) y (Granger y Weiss, 1983) para resolver el problema de correlación espuria. Existen varios métodos para probar la cointegración, entre ellos el procedimiento Engle-Granger, el procedimiento de cointegración de Johansen y el método de cointegracion del modelo Autorregresivo de Rezagos Distribuidos (ARDL) con el enfoque de la prueba de los limites desarrollado por Pesaran, Shin y Smith (2001).

Consideremos dos series de tiempo (, las que se someten de manera individual a un análisis de raíz unitaria, determinando que ambas son ***I***(1) y por tanto no estacionarias. La regresión de sobre queda expresada por:

A continuación, si expresamos la ecuación anterior de otra manera tenemos:

Al realizar un nuevo análisis de raíz unitaria, cabe la posibilidad de descubrir que la combinación lineal es ***I***(0), resultando en una serie estacionaria. Por ende, se podría decir que una regresión de sobre sería cointegrada. En otras palabras, la combinación lineal

elimina el efecto de las tendencias estocásticas en las dos series.

Se define como Regresión Cointegrante aquella combinación lineal de series no estacionarias que generan una serie de errores estacionarios. Para este caso, a *β1* se le denomina parámetro cointegrante. Para determinar la cointegración se debe determinar la estacionariedad de la serie de errores. Este análisis se puede realizar mediante las siguientes pruebas:

* Prueba de Engle-Granger
* Prueba de Engle-Granger Aumentada
* Prueba Durbin-Watson Regresión Cointegración

**2.1.6. Modelo de Corrección de Errores (ECM)**

Un concepto que se relaciona con cointegración es el modelo de corrección de errores (ECM), popularizado por Engle y Granger (1987), este modelo corrige el desequilibrio entre las variables, es decir, permite evaluar la relación de corto plazo, así como la relación de cointegración (de largo plazo). Un importante teorema de representación de Granger, afirma que, si dos variables Y y X están cointegradas, la relación entre las dos se expresa como un modelo de corrección de errores (ECM).

El ECM es un término que se presenta como una función integrada por variables rezagadas de largo plazo, de una ecuación que se compone por variables diferenciadas. A medida que el modelo alcanza el equilibrio, y el estado estacionario, los cambios en las variables tienden a cero, y solo permanece esta relación a largo plazo (Bahmani-Oskooee y Hegerty, 2007). Este mecanismo permite que una ecuación incorpore tanto los datos diferenciados como los originales, por lo que se incorpora la información que se haya perdido debido a la diferenciación.

Modelo de largo plazo:

Vamos a incorporar al modelo, el termino corrección de error, los cuales son residuos de la regresión de largo plazo, pero rezagado en un periodo.

Introduciendo

Estimación del modelo de corrección de error

Consideraciones:

**.** ETC = = Termino corrección de error

= Es el coeficiente estimado del término corrección de error, donde -1 < < 0

**.** Los valores fuera del rango son resultados explosivos, necesita reestimar el modelo

**.** El coeficiente determina la rapidez del ajuste, hacia un equilibrio de largo plazo.

**.** La desviación desde el equilibrio de largo plazo es corregido gradualmente por el ECT a través la serie de ajuste de corto plazo parcial.

**2.1.7. Causalidad de Grenger**

No necesariamente el hecho de que dos variables estén relacionadas entre sí implica que existe causalidad entre ellas ya que su relación puede ser generada solo por mera coincidencia. De este modo es que nace el test de causalidad de Grenger, para establecer si existe alguna variable cuyos cambios puedan anteceder al comportamiento de la otra. Este test se enfoca a la aplicación variables que sean series de tiempo estacionarias ( y ) para evitar el riesgo de obtener relaciones espurias.

Sean dos series de tiempo y . Los modelos que se deben generar para determinar si existe o no causalidad son los siguientes:

A partir de estos modelos se pueden obtener 4 posibles escenarios:1. Si , implica que existe causalidad unidireccional de Y hacia X (Y =>X).

2. Si , implica que existe causalidad unidireccional de X hacia X (X => Y).

3. Si , implica que existe causalidad bidireccional o retroalimentación, es decir que existe causalidad en ambos sentidos.

4. Si , implica que no existe causalidad en ningún sentido.

SiA7Teoria-desbloaqueado (ModeloArima —TesisMetodoArima) (SiDemanda gas España)

**2.2 MODELO UNIVARIANTE ARIMA**

A finales del siglo XX diversos investigadores comenzaron a desarrollar y aplicar nuevas propuestas para la modelación de series de tiempo, las cuales lograron probar su eficacia, ya que son capaces de tratar con cualquier patrón de datos y producir pronósticos precisos en base a la descripción de patrones históricos en los datos (Aguirre 1994).

Específicamente los modelos desarrollados en las últimas décadas son los llamados autorregresivos (AR), de medias móviles (MA), integrados (I), así como las posibles combinaciones entre estos (ARMA y ARIMA).

La principal diferencia entre estos modelos y los clásicos es el enfoque estocástico que se le da a las series de tiempo, en vez de tratarlas de forma determinística. Bajo este enfoque se concibe a la serie de tiempo como un conjunto de valores de tipo aleatorio, generados a partir de un proceso totalmente desconocido, es decir, se concibe a la serie como un proceso estocástico.

Así mismo, derivado del desconocimiento del proceso generador de los datos, el objetivo de este enfoque es tratar de identificar el modelo probabilístico que represente las características principales del comportamiento de la serie (Aguirre 1994).

El desarrollo y aplicación de este tipo de modelos de series de tiempo en un principio estuvieron gravemente limitados, fundamentalmente debido a razones de cómputo. Actualmente con la amplia disponibilidad de las computadoras, el uso de los modelos ARIMA se ha hecho posible.

**2.2.1. Enfoque y procesos estocásticos**

En general las series de interés llevan asociados fenómenos aleatorios. Por este motivo el estudio de su comportamiento pasado sólo permite acercarse a la estructura o modelo probabilístico para la predicción del futuro. La teoría de los procesos estocásticos se centra en el estudio y modelización de sistemas que evolucionan a lo largo del tiempo, o del espacio, de acuerdo a unas leyes no determinísticas, esto es, de carácter aleatorio y probabilístico.

Para aclarar esta idea definimos un proceso estocástico como una familia de variables aleatorias asociadas a un conjunto índice de números reales, de tal forma que a cada elemento del conjunto le corresponda una y sólo una variable aleatoria que evoluciona con el tiempo:

Variables aleatorias {}

Con t = 1,2, 3, …, n

Bajo este enfoque, una serie de tiempo se define como una observación particular sobre el estado dinámico de una variable de un proceso de naturaleza aleatoria. Simultáneamente se presupone que los procesos de series de tiempo siguen un comportamiento aleatorio y debido a la propia dinámica de dicho proceso, tales impulsos pueden generar un conjunto de datos serialmente dependientes.

Así, mediante la identificación de la estructura de dependencia latente que existe entre las observaciones, se puede modelar la serie. Una de las principales ventajas de este enfoque aplicado a las series de tiempo, es la gran flexibilidad que se logra para representar un buen número fenómenos reales mediante una sola clase general de modelos. Otra ventaja es la facilidad y precisión para realizar pronósticos.

**2.2.2. Modelos AR, MA y mixtos**

La estructura existente de dependencia entre los datos en un proceso estocástico, puede ajustarse a seis modelos que pueden describir una gran variedad de fenómenos reales. La principal característica de dichos modelos es que no involucran a las variables independientes en su construcción, es decir, a la variable del tiempo *t*, como se hace en el análisis clásico. En cambio, emplean la información que se encuentra en la serie misma para generar los pronósticos (Makridakis y Wheelwright 2000). Los modelos son los que se describen a continuación (Arnau 2001):

Proceso de ruido blanco – (at): Es un proceso formado por una secuencia de variables aleatorias mutuamente independientes e idénticamente distribuidas.

1. Ruido blanco: el proceso *t* se define como ruido blanco si cumple con las siguientes condiciones:

* La esperanza de o promedio es igual a 0 para todos los periodos *t*. Esto es E () = 0 para toda *t*.
* La varianza de es constante y por consiguiente independiente del tiempo. Esto es var () =

1. Independiente e idénticamente distribuido(iid): el proceso se dice ser iid si cumple con las siguientes condiciones:

* La esperanza de es constante pero no necesariamente igual a 0 para todos los periodos t. Esto es E () = µ.
* La varianza de es constante y por consiguiente independiente del tiempo. Esto es var () =
* es independiente de *k* y para todas las *t* y k con *t* *k*.

**1. Modelo no estacionario de corrido aleatorio - (I):** Correspondiente a aquellas situaciones en las que los impulsos aleatorios tienden a sumarse o a integrase en el tiempo. La integración refleja la presencia de un componente de tendencia.

**2. Modelo autorregresivo de orden p - AR(p)**: Se trata de un modelo en el que una determinada observación es predecible a partir de la observación inmediatamente anterior (modelo autorregresivo de primer orden) o a partir de las dos observaciones que le preceden (modelo autorregresivo de segundo orden). En este caso, la observación actual se define como la suma ponderada de una cantidad finita *p* de observaciones precedentes más un impulso aleatorio independiente.

Matemáticamente, un modelo autorregresivo tiene la forma siguiente:

donde es la variable dependiente y son las variables independientes.

En este caso, las variables independientes son valores de la misma variable (de aquí el nombre de auto), pero de periodos anteriores (*t* − *1*, *t* − *2*, · · ·, *t* – *p*). Por último, es el error, o término residual, que representa perturbaciones aleatorias que no pueden ser explicadas por el modelo.

El modelo se llama autorregresivo porque se asemeja a la ecuación de regresión:

Las variables independientes son simplemente valores rezagados de la variable dependiente con rezagos de tiempo 1, 2, 3, …, *p* periodos.

**3. Modelo de Medias Móviles de orden q - MA(q)**: En este modelo, una determinada observación está condicionada por los “impulsos aleatorios” de las observaciones anteriores. De esta forma la observación actual se define como la suma del impulso actual y de los impulsos aleatorios anteriores con un determinado peso.

El modelo de medias móviles tiene la siguiente forma

Donde es el residuo o error en el periodo *t* y son los valores anteriores del error. La ecuación es semejante a la anterior, con la excepción de que implica que la variable dependiente depende de valores previos del término residual más que de la variable misma.

**4. Modelo autorregresivo de medias móviles de orden p, q - ARMA (p, q)**: Este modelo es la combinación de las estructuras anteriores: modelo autorregresivo y modelo de medias móviles. Así, una observación está determinada tanto por observaciones anteriores, así como por “impulsos aleatorios” o también llamados “errores” de observaciones pasadas.

La forma general de un modelo autorregresivo de medias móviles es:

Se observa que la ecuación es sencillamente una combinación de las ecuaciones del modelo AR y MA.

En la ecuación anterior *p* simboliza los valores tomados por la variable durante los *p* periodos anteriores (*t* − *1*, *t* − *2*, *t – 3,* · · ·, *t* – *p*). El grado de influencia de cada valor anterior sobre el valor considerado de la variable viene dado por el parámetro correspondiente Por su parte *q* representa los errores o residuos de la variable durante los *q* momentos anteriores (*t* − *1*, *t* − *2*, *t – 3,* · · ·, *t* – *p*). El grado de influencia de cada uno de ellos viene ponderado por el parámetro correspondiente

Un modelo ARMA puede ajustarse a cualquier patrón de datos. Sin embargo, los valores de *p* y *q* se deben de especificar.

Por ejemplo, si *p* = 1 y *q* = 0, el modelo ARMA es:

El modelo derivado de esta ecuación es llamado modelo AR de primer orden y se escribe como AR (1) o ARMA (1,0).

Por último, *p* y *q* pueden cada una ser diferentes de 0, en cuyo caso existe un modelo ARMA mixto.

Por ejemplo, cuando *p* = 1 y *q* = 1, el modelo es un modelo ARMA (1,1) y es de la forma:

Es interesante darse cuenta que esta ecuación es no lineal y por lo tanto suficientemente general para describir una amplia variedad de patrones de datos.

Obviamente *p* y *q* pueden adoptar muchos otros valores, aun cuando raramente son mayores que 2.

**5. Modelo autorregresivo integrado de medias móviles de orden p,d,q - ARIMA (p,d,q)**: Al igual que un modelo ARMA, es la combinación de los modelos autorregresivo y el de medias móviles, con la particularidad de incluir un proceso de restablecimiento (el cual se denomina integración) de inestabilidad original presente en una serie de tiempo.

La forma general de un modelo ARIMA es semejante al de un modelo ARMA:

Donde:

: Es la serie inducida a la estabilidad

En ocasiones es apropiado incluir un término constante a los modelos AR, MA y ARIMA. En un modelo que contiene sólo parámetros de media móvil, el valor de la constante es la media de los valores de la serie = Mientras que en un modelo autorregresivo (Arnau 2001). La inclusión del término depende de la serie en estudio. Concretamente, depende del valor de la media. Si la media de todas las observaciones es cero o muy cercana, no se incluye término constante. Si la media es significativamente distinta de cero, se incluye el término.

Un rápido indicador, es conocer si la serie ha sido diferenciada. Si la serie ha sido diferenciada, la constante es eliminada, se puede probar que en una serie diferenciada . Por el otro lado si la serie no es diferenciada, se debe verificar el valor de *µ* (Arnau 2001).

**2.2.3. Metodología Box-Jenkins No estacional**

Existen distintas metodologías que ayudan a elegir el modelo que mejor se represente a la serie de tiempo de entre los diversos modelos existentes. La metodología más usada y difundida es la que propusieron los profesores G.E.P Box y G.M Jenkins en la década de los años 1970, en la cual lograron grandes avances en identificar, ajustar y verificar los modelos ARIMA adecuados (Hanke y Wichern 2006). Comúnmente se le conoce como Metodología Box-Jenkins. Se especifica en esta metodología únicamente el uso de los modelos ARIMA ya que los modelos autorregresivos (AR) y los de medias móviles (MA) se dicen que son las formas particulares de la clase general de modelos ARMA y este a su vez del modelo ARIMA.

Esta metodología se basa en tratar de determinar cuál es el modelo probabilístico que rige el comportamiento del proceso a lo largo del tiempo. Cabe apuntar en este momento la diferencia entre proceso y modelo. Se puede decir que un proceso es lo real, es decir, el fenómeno en sí, del cual se desconoce su mecanismo generador. Por el otro lado un modelo es solo la imitación o representación del proceso (Pankratz 1983).

Además, en cuanto al tipo de datos que son aptos para someterse al análisis mediante la modelación ARIMA, se puede decir debido a la experiencia, que la mayoría de las series de tiempo se pueden modelar mediante estos métodos. Por ese motivo esta metodología es ampliamente utilizada.

A continuación, se destacan algunas consideraciones importantes acerca de la naturaleza de los datos en la modelación ARIMA (Pankratz 1983):

• Los modelos ARIMA aplican tanto para datos discretos o continuos. Los datos discretos son aquellos que son medidos solamente en números enteros, nunca con cifras decimales. Mientras tanto los datos continuos son medidos en intervalos fraccionarios, es decir, con cifras decimales.

• Aunque la metodología ARIMA trata tanto con datos discretos y continuos, sólo se puede

aplicar a datos espaciados equidistantemente en el tiempo, en intervalos discretos de tiempo. Los datos medidos en intervalos discretos de tiempo pueden clasificarse en dos tipos: 1) datos que son producto de la acumulación durante un periodo de tiempo, por ejemplo, los ahorros de una persona en un mes. 2) datos que son producto de la medición instantánea periódicamente, por ejemplo, la medición de la presión en una tubería en intervalos de una hora.

• Para elaborar un modelo ARIMA se requiere una cierta cantidad de datos mínimo. Los profesores Box y Jenkins sugieren un mínimo de 50 observaciones. Un modelo ARIMA se puede aplicar a una serie con menor tamaño, realizando con especial cautela su interpretación. Para series con patrones estacionales se aconseja una serie con gran número de muestras observadas.

• Los modelos ARIMA son especialmente útiles en el tratamiento de series que presentan patrones estacionales.

• Los métodos Box-Jenkins aplican a series estacionarias y no estacionarias. Una serie estacionaria es aquella cuya media, varianza y función de autocorrelación permanecen constantes en el tiempo.

• Se asume que las perturbaciones aleatorias ( presentes en la serie, son independientes entre sí. No existe correlación entre ellas, por lo tanto, ningún patrón modelable.

Una vez hechas estas anotaciones, se puede entrar de lleno a la metodología. Se destacan en especial 5 etapas principales en la aplicación de esta metodología (Aguirre 1994):

1. Estacionariedad

2. Identificación

3. Estimación

4. Evaluación

5. Pronóstico

Cada una de estas etapas se desarrolla en las siguientes secciones.

**1. Estacionariedad**

Los procesos estocásticos se clasifican en estacionarios y no estacionarios. La idea de estacionariedad está relacionada con la estabilidad de la serie. Un proceso estacionario se describe como una secuencia de datos o valores que no presentan cambio sistemático en la media, ni cambio sistemático alguno en la varianza, se dice entonces que la serie es estable. De forma parecida, un proceso no estacionario es aquel que presenta cambios sistemáticos en la media y/o varianza; la serie de datos es inestable en el tiempo.

Aplicado dicho concepto a una secuencia cronológica y explicado en forma intuitiva, una serie de tiempo es estacionaria si sus propiedades estadísticas (media, varianza) son esencialmente constantes a través del tiempo. Una serie cuya media y/o varianza cambian a través del tiempo es una serie no estacionaria (Bowerman, O’Conell, Koehler 2007).

A partir de la definición previa, se puede advertir que una serie cuyos valores fluctúan respecto de una media constante, es decir, sin tendencia, es una serie estacionaria.

La Figura 6 muestra la gráfica de una serie de tiempo, en ella se observa que los valores de una serie de tiempo fluctúan respecto de una media constante, entonces es razonable pensar que la serie temporal es estacionaria. Si los valores no fluctúan respecto de una media constante o fluctúan con variación constante, entonces es razonable pensar que la serie de tiempo es no estacionaria, como se observa en la Figura 7

Este concepto es importante ya que el primer paso en la metodología Box-Jenkins es determinar si la serie de tiempo es estacionaria. En caso de que la serie de datos no sea estable a lo largo del tiempo, es necesario aplicar una transformación a la serie para inducirla a ello. Este paso es indispensable para poder aplicar las demás etapas de esta metodología.

Existen dos formas de conocer si una serie es estacionaria: por medio del gráfico de la serie y mediante la exploración de función de autocorrelación simple.

**a. Por medio de inspección gráfica**

Visualmente se puede advertir si la serie es estacionaria si se detectan elevaciones o inclinaciones en los datos. Cualquier patrón de este tipo expresa que la serie es inestable.

**b. Por medio de las funciones de autocorrelación**

Cuando visualmente no es factible determinar si la serie es estacionaria, usualmente se recurre a la función de autocorrelación simple (FAS), en especial en aquellas series con tendencia poco remarcada. El reconocimiento de la estabilidad se logra a partir de la variedad de comportamientos que esta función que puede mostrar (Bowerman, O’Conell, Koehler 2007).

En primer lugar, la función de autocorrelación simple se puede cortar o truncar. Para explicar este término, se dice que existe una espiga en el desfasamiento *k* en la FAS si es estadísticamente grande.

Se considera que existe una espiga en la función si el valor absoluto de:

es mayor que 2. La función de autocorrelación se trunca después del desfasamiento *k* si no hay espigas en los desfasamientos mayores que *k* en la función. Por ejemplo, si

Entonces la función de autocorrelación simple se trunca o se corta después del desfasamiento 2. Esto se ilustra en la Figura 8.

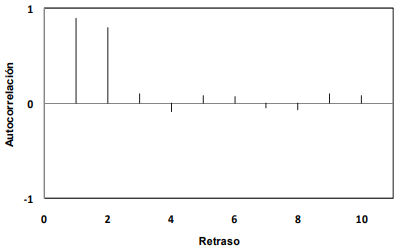


Figura 8 La función de autocorrelación se trunca

Fuente:

La salida muestra automáticamente el valor de *t* asociado a cada coeficiente y además muestra en la gráfica un par de bandas que representan el valor de dos desviaciones estándar 2. Si un coeficiente sobrepasa estas bandas ya sea en la parte de arriba como en la de abajo, significa que el valor de es mayor que 2.

Segundo, la función de autocorrelación simple decae ó se extingue si esta no se corta o muere, sino que decrece en una forma permanente. Puede extinguirse de tres maneras:

• En forma exponencial (Figura 10)

• En seno amortiguada (Figura 11)

• En combinación de 10 y 11 (Figura 12)

Además, puede cortarse con gran rapidez como se muestra en la Figura 13 o puede extinguirse con lentitud extrema como se muestra en la Figura 14.

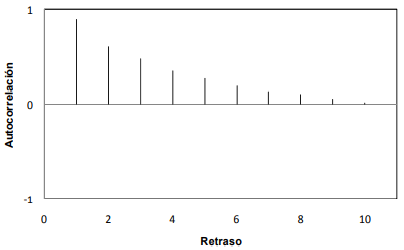


Figura 10 Se extingue en forma exponencial

Fuente:

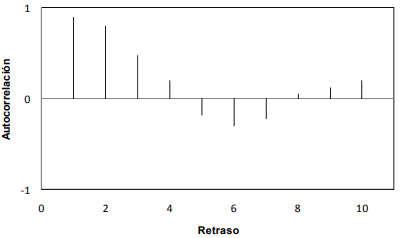


Figura 11 Se extingue en forma de seno amortiguada

Fuente:

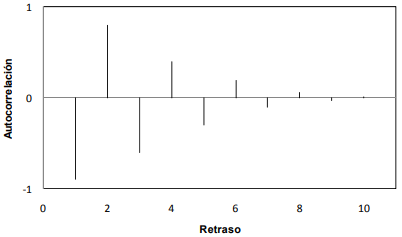


Figura 12 Se extingue en forma combinada

Fuente:

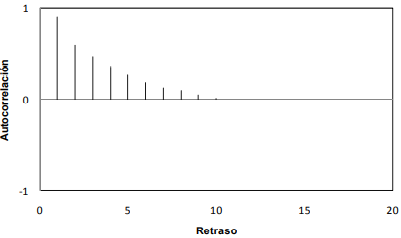


Figura 13 Se extingue con gran rapidez

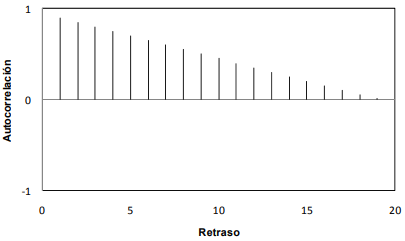


Figura 14 Se extingue con lentitud extrema

Fuente:

En general, se puede demostrar que:

Si la función de autocorrelación simple de los valores de la serie de tiempo se corta claramente con rapidez o si se corta rápidamente, entonces se deben considerar que los valores de la serie temporal son estacionarios.

Si la función de autocorrelación simple de los valores  de la serie de tiempo se corta con lentitud extrema, entonces se deben considerar que los valores de la serie de tiempo son no estacionarios

Cabe aclarar que los términos “claramente con rapidez” y con “lentitud extrema” es algo

arbitrario, y se puede determinar mejor a través de la experiencia. Además, la experiencia indica que en lo que se refiere a datos no estacionales, si la función de autocorrelación simple se trunca claramente con rapidez, casi siempre sucede así, después de un desfasamiento *k* que es menor o igual a 2 (Bowerman, O’Conell, Koehler 2007).

De esta simple forma se determina por medio de la FAS si la serie es estacionaria. Ahora bien, la realidad es que las series de tiempo normalmente no se comportan establemente, es decir, no son procesos estocásticos estacionarios. Cuando se presenta esa situación entonces es preciso aplicar una transformación a los valores de la serie para tener una serie temporal estacionaria.

Existen distintas maneras de inducir la estacionariedad en una serie de tiempo, la más usada es el método llamado de construcción de diferencias. Este método consiste, como su nombre lo dice, en obtener diferencias entre los mismos valores de la serie, para remover cualquier patrón de tendencia.

Las primeras diferencias de los valores de la serie de tiempo son:

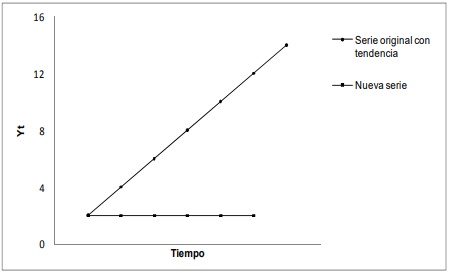
*t* = 2,3, …, *n*

La Tabla 2 muestra cómo se hace la diferenciación y cómo los datos con una tendencia lineal se convierten en estacionarios (horizontales) después de diferenciar. Gráficamente se puede observar esta transformación en la Figura 15

**Tabla 2 Obtención de las primeras diferencias en una serie**

|  |  |  |
| --- | --- | --- |
| Serie de datos | Primeras diferencias | Nueva serie |
| 2  4  6  8  10  12  14 | 4-2 = 2  6-4 = 2  8-6 = 2  10-8 = 2  12-10 = 2  14-12 = 2 | 2  2  2  2  2  2  2 |

**Fuente:** **Elaboración propia**



**Figura 15 Transformación de la serie de tiempo mediante las primeras diferencias**

**Fuente: Elaboración propia**

En ocasiones es necesario realizar más de una diferenciación para generar valores de series de tiempo estacionarias. Si los valores originales no son estacionarios y las primeras diferencias de los valores originales de la serie de tiempo tampoco son estacionarios, entonces podemos obtenerlas segundas diferencias de los valores originales de la serie de tiempo.

Las segundas diferencias de los valores de la serie de tiempo son:

*t* = 3,4, …, n

La Tabla 3 muestra cómo se obtienen las segundas diferencias a una serie de tiempo, se puede apreciar que las primeras diferencias siguen siendo no estacionarias, al aplicar una nueva diferenciación se obtiene una serie estacionaria. Cabe aclarar que la notación tan sólo se usa para indicar que la serie se transformó, ya sea mediante las primeras o las segundas diferencias.

En la gráfica de la Figura 16 se observa ambas transformaciones a la serie original.

**Tabla 3 Obtención de las segundas diferencias de una serie**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Series de datos | Primeras diferencias | Nueva serie | Segundas diferencias | Nueva serie |
| 5.6  15.5  29.0  46.0  66.7  91.0  118.9 | 15.5 - 5.6 = 9.9  29 - 15.5 = 13.5  46 - 29 = 17  66.7 - 46 = 20.7  91 - 66.7 = 24.3  118.9 - 91 = 27.9 | 9.9  13.5  17.1  20.7  24.3  27.9 | 13.5 – 9.9 = 3.6  17.1 - 13.5 = 3.6  20.7 - 17.1 = 3.6  24.3 – 20.7 = 3.6  27.9 - 24.3 = 3.6 | 3.6  3.6  3.6  3.6  3.6 |

**Fuente:**

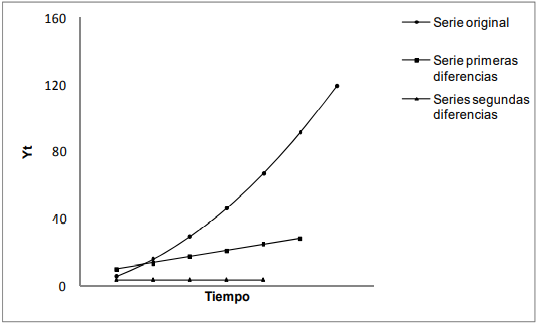


Figura 16 Transformación de la serie mediante las primeras y segundas diferencias

**Fuente:**

La experiencia indica que, si los valores de la serie de tiempo original son no estacionarios y no estacionales, entonces si se aplica la primera transformación por diferencias o la segunda transformación por diferencias por lo general se producen valores estacionarios de la serie de tiempo.

Una vez introducido el tema de estabilidad en una serie de tiempo, a fin esclarecer la diferencia entre un modelo ARMA y un modelo ARIMA. La diferencia se halla en que un modelo ARMA(p,q) es capaz de operar únicamente sobre series estacionarias, mientras que un modelo ARIMA(p,d,q) es capaz de operar tanto sobre series de tiempo no estacionarias, como en series estacionarias.

El modelo ARIMA(p,d,q) es capaz de operar sobre cualquier ya que se incluye el proceso de estabilización de la serie, por medio del parámetro *d*, que simboliza el grado de diferenciación aplicado a la serie de tiempo. Por ejemplo, si *d* =1, significa que se está trabajando con las primeras diferencias de la serie; si *d* = 2, se trabaja con las segundas diferencias y así sucesivamente.

**2. Identificación**

Una vez que se asegura que la serie es estable en el tiempo. El siguiente paso en la metodología Box-Jenkins es la identificación del modelo probable que rige el proceso de la serie de tiempo. Las ideas básicas de esta fase son las siguientes:

• La serie de tiempo que se encuentre en proceso de estudio cuenta con sus respectivas funciones de autocorrelación simple y parcial (FAS y FAP), que se denominan prácticas o calculadas.

• Por otra parte, cada una de las distintas configuraciones ARMA posee su FAS y FAP teóricas asociadas al modelo.

• Si la FAS y FAP calculadas de la serie a la que deseamos ajustar un modelo se asemeja a alguna o varias FAS y FAP teóricas, entonces podemos decir que el modelo ARMA teórico es un modelo tentativo para la serie.

Queda claro entonces que la identificación del modelo probable se realiza por medio de la comparación de las funciones de autocorrelación calculadas contra las teóricas (tanto simples como parciales).

Las funciones de autocorrelación calculadas sólo se usan como guías para escoger uno o varios modelos ARMA que parezcan apropiados, ya que tan sólo brindan una aproximación a la estructura más adecuada que debe considerarse.

Las funciones de autocorrelación teóricas por su parte se derivan de una familia de modelos ARMA propuestos por Box y Jenkins. Cabe aclarar que, aunque existen una infinidad de posibles procesos dentro de estos modelos ARMA, la mayoría de los procesos que ocurren normalmente en la práctica caben dentro de un número reducido de modelos. Es por esto que los modelos más comunes en la práctica, suelen ser precisamente los propuestos y desarrollados por estos profesores (Pankratz 1983)

De esta manera entonces los profesores Box y Jenkins sugirieron una completa familia de modelos teóricos de los cuales se puede escoger: AR(1), AR(2), MA(1), MA(2), ARMA(1,1).

Las principales características que se observan de las funciones de autocorrelación teóricas son:

**Tabla 4 Características FAS y FAP Teóricas (Pankrats)**

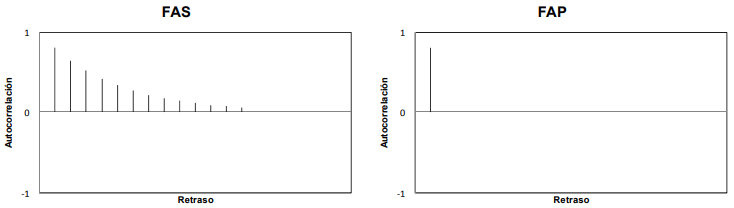
|  |  |  |
| --- | --- | --- |
| Modelo | FAS | FAP |
| AR | Decae a cero | Se trunca o se corta (después del retraso p) |
| MA | Se trunca o se corta (después del retraso q) | Decae a cero |
| ARMA | Decae a cero | Decae a cero |

**Fuente:**

A partir de la tabla anterior se describe en seguida a detalle las características de cada uno de los posibles modelos ARMA (Pankratz 1983).

**a. Modelo AR:** Las funciones de autocorrelación simples de los procesos autorregresivos decaen a cero. Es común que se use la expresion: “se extingue” . La forma en que se extingue la FAS puede ser de forma exponencial, en forma de seno amortiguado, etc. Ahora bien, las funciones de autocorrelación parcial contienen espigas significativas hasta el retraso *p*, después del cual la función “se corta”. El número de espigas indica, por tanto, el orden del proceso autorregresivo. Usualmente *p* no sobrepasa el valor de 2 o 3.

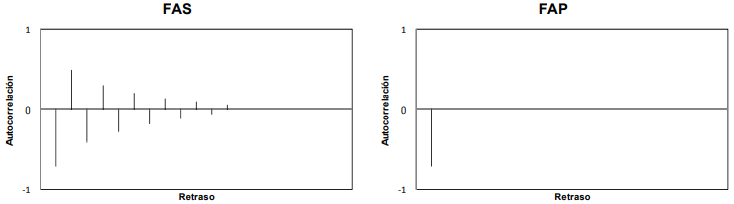
La Figura 17a y 17b (Aguirre 1994) muestra dos configuraciones posibles para un modelo AR(1) o bien ARMA(1,0). De esta figura se puede destacar que la FAS de un proceso autorregresivo de primer orden decae exponencialmente hacia cero, mientras que existe una espiga significativa en la FAP en el primer retraso. Además, si , La FAS decae positivamente hacia cero y la espiga en la FAP es positiva también. Si , la FAS decae hacia cero alternando signos, mientras que la espiga significativa en la FAP es negativa.

a)

**Figura 17a Funciones de autocorrelación simple y parcial para el modelo AR(1)**

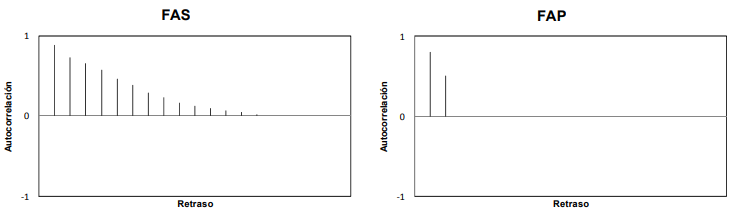
**Fuente:**

b)

 **Figura 17b Funciones de autocorrelación simple y parcial para el modelo AR(1)**

**Fuente:**

Para un modelo AR (2) o ARMA (2,0) existen una mayor variedad de patrones posibles que para un modelo AR (1). En general un proceso autorregresivo de segundo orden cuenta con una FAS que decae a cero tanto de forma exponencial como de seno amortiguado ó bien una mezcla de ambos. La FAP se caracteriza por tener espigas en los retrasos 1 y 2. El patrón exacto depende en las posibles combinaciones de signos para . En la Figura 18a y 18b se muestran dos de las cuatro posibles combinaciones.

 Figura 18a Funciones de autocorrelación simple y parcial teóricas para el modelo AR(2)

Fuente:

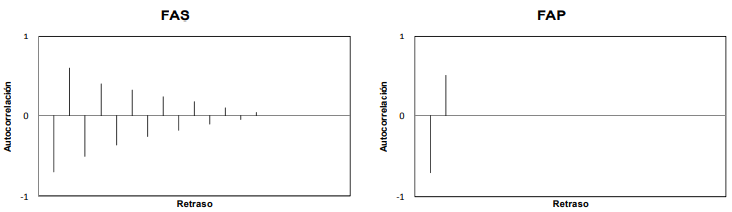


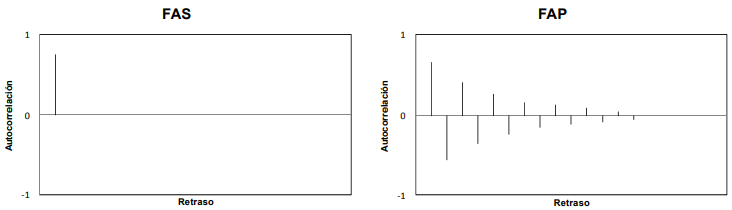
Figura 18b Funciones de autocorrelación simple y parcial teóricas para el modelo AR(2)

**Fuente:**

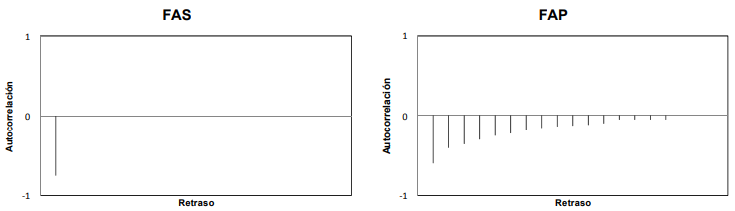
**b. Modelo MA**: En general la FAS de un proceso de medias móviles cuenta con espigas significativas hasta el retraso *q*, después del cual se corta la función. La FAP por su lado decae de forma exponencial o en forma de seno amortiguado. Normalmente el valor de *q* no es mayor que 2 o 3.

La Figura 19a 19b muestra las configuraciones de un modelo MA(1). Se ilustra como si < 0 en la FAS se presenta una espiga positiva mientras que la FAP decae exponencialmente a cero alternando signos. Si la espiga en la FAS es negativa, la FAP decae exponencialmente a cero con signo negativo.

a)

**Figura 19a Funciones de autocorrelación simple y parcial para el modelo MA(1)**

**Fuente:**

b)

**Figura 19b Funciones de autocorrelación simple y parcial para el modelo MA(1)**

**Fuente:**

Para un modelo MA(2) la Figura 20a y 20b muestra algunas formas que pueden tomar las funciones de autocorrelación. Se muestra que existen espigas en la FAS hasta el retraso *q*, en este caso hasta q = 2 y después de este retraso la función se corta o se extingue. La FAP decae a cero con distintas configuraciones.

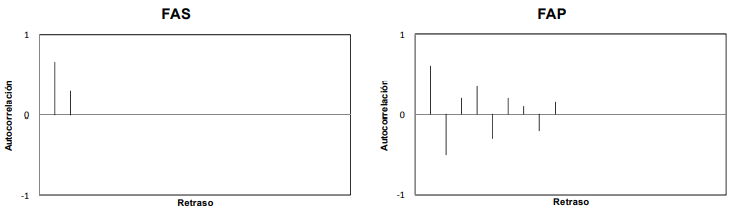
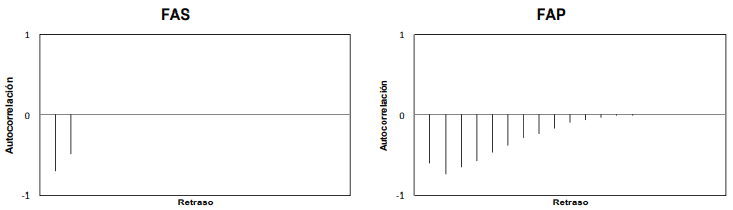


Figura 20a Funciones de autocorrelación simple y parcial teórica para el modelo MA(2)

**Fuente:**

Figura 20b Funciones de autocorrelación simple y parcial teórica para el modelo MA(2)

**Fuente:**

**c. Modelos ARMA**: Las funciones de autocorrelación combinan las características que se presentan en los modelos AR y los MA. En general la FAS decae a cero tanto exponencialmente como en forma de seno amortiguado después del retraso de tiempo p -q. La FAP teórica también decae a cero después del retraso p - q. Se muestran en la Figura 21a y 21b dos de las seis posibles combinaciones para el modelo teórico ARMA(1,1)

a)

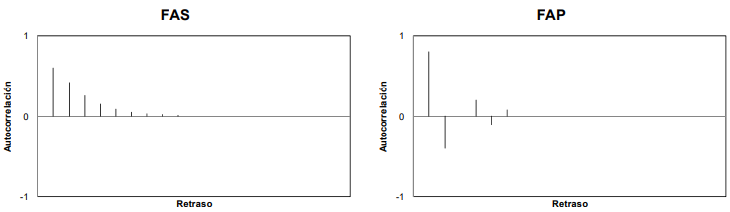


Figura 21a Funciones de autocorrelación simple y parcial teórica para el modelo ARMA(1,1)

**Fuente:**

b)

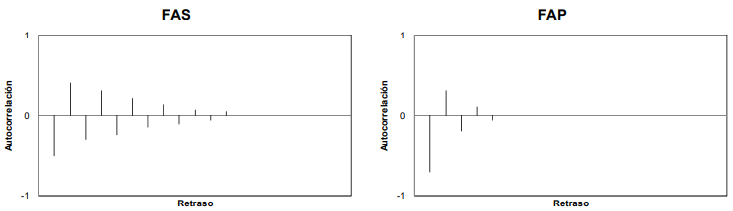


Figura 21b Funciones de autocorrelación simple y parcial teórica para el modelo ARMA(1,1)

**Fuente:**

Es probable que exista cierta ambigüedad al determinar un modelo ARIMA apropiado a partir de los patrones que provienen de las FAS y FAP parciales de la muestra. Cualquier modelo que se escoja, se considera simplemente tentativo: sólo es un candidato para ser considerado como el modelo ideal. Si no fuera el caso, se deberá intentar con un modelo alterno.

**3. Estimación**

En esta etapa se estiman los coeficientes del modelo escogido tentativamente en el paso anterior, existen distintos criterios para dicho propósito. Usualmente se realiza por medio de paquetes informáticos. Además, en esta etapa se puede pre-visualizar la adecuación del modelo a la serie de tiempo. Particularmente si los coeficientes no satisfacen ciertas condiciones matemáticas, el modelo es rechazado.

El principal criterio para el cálculo de los parámetros es el llamado ***Estimación de máxima***

***verosimilitud ó Máximum Likelihood*** (ML por sus siglas en ingles). Se ha demostrado que los parámetros estimados a través de este criterio reflejan con gran exactitud las características presentes en los datos de la serie de tiempo. Sin embargo, este método se ve limitado debido al gran tiempo que conlleva realizar la estimación, aún con el uso de la computadora (Pankratz 1983).

Por esta razón la mayoría de la literatura existente sugiere el uso del criterio de ***mínimos cuadrados o least squares*** (LS por sus siglas en ingles). Se ha demostrado que las estimaciones hechas mediante este criterio son muy cercanas a las del criterio de máxima verosimilitud. La estimación de los parámetros del modelo ARMA seleccionado, se realiza por medio de minimizar la suma de los cuadrados de los residuales SSR (sum of squared residuals). En general, estos estimados de los mínimos cuadrados deben obtenerse mediante un procedimiento no lineal de mínimos cuadrados. Un procedimiento no lineal de mínimos cuadrados es, sencillamente, un algoritmo que encuentra el mínimo de la suma de la función de errores cuadrados, mediante un proceso iterativo.

Los residuales de cualquier modelo ARMA se definen como:

Donde:

: residual de la serie

serie original

serie calculada con los parámetros estimados

Entonces, la suma de los cuadrados de los residuales se define como:

Esta última función es la que se debe minimizar, a través de la búsqueda de los parámetros del modelo ARMA que se desea ajustar.

Finalmente, obtenidos los parámetros del modelo ARMA propuesto, deben de cubrir en conjunto los siguientes requerimientos (Pankratz 1983):

**a.** El modelo debe tener parsimonia: Box y Jenkins enfatizaron la importancia del principio de parsimonia. Un modelo con parsimonia ajusta la información disponible sin usar coeficientes innecesarios. Es decir, se opta siempre por el modelo con el menor número de coeficientes. Este principio es importante ya que en la práctica un modelo con esta característica produce mejores predicciones. El objetivo en la modelación ARIMA no es encontrar el modelo exacto que represente al proceso generador de las observaciones, es más bien encontrar el modelo que se aproxime al verdadero proceso y que explique el comportamiento de la variable en estudio de forma adecuada y práctica.

**b.** El modelo debe ser estacionario e inversible: La metodología Box-Jenkins requiere que el modelo que se utiliza para la descripción y pronóstico de una serie de tiempo sea tanto estacionario como inversible.

La inversibilidad se refiere a que cualquier modelo ARIMA puede expresar a la serie en función de las Y observaciones pasadas, es decir etc. Esto es obvio para un modelo autorregresivo de orden *p*:

No es obvio para un modelo de medias móviles de orden *q*:

No obstante, se puede demostrar que un modelo de medias móviles puede expresar a la serie como una serie infinita de *Y* observaciones pasadas. Es decir, un modelo de medias móviles se puede transformar o invertir en un modelo autorregresivo de orden infinito.

La Tabla 5 (Bowerman, O’Conell, Koehler 2007) muestra las condiciones de estacionariedad e inversibilidad para los modelos ARMA más comunes. Las condiciones para modelos generales son complicadas por lo que no se tocara en esta tesis. Sin embargo, se puede decir que, si un modelo utiliza parámetros autorregresivos, una condición de estacionariedad necesaria (pero no suficiente) es que la suma de los valores de los parámetros autorregresivos es menor que 1. Si un modelo utiliza parámetros de media móvil, una condición de inversibilidad necesaria (pero no suficiente) es que la suma de los valores de los parámetros de la media móvil es menor que 1. Además, un modelo que utiliza sólo parámetros autorregresivos no tiene condiciones de inversibilidad y un modelo que utiliza sólo parámetros de media móvil no tiene condiciones de estacionariedad.

**Tabla 5 Condiciones de estacionariedad e invertibilidad**

|  |  |  |
| --- | --- | --- |
| Modelo | Condiciones de estacionariedad | Condiciones de inversibilidad |
| AR (1) |  | Ninguna |
| AR (1) |  | Ninguna |
| MA (1) | Ninguna |  |
| MA (2) | Ninguna |  |
| ARMA (1,1) |  |  |

c. Los coeficientes del modelo deben ser estadísticamente significativos: Se debe determinar si los coeficientes son importantes en el modelo o se deben excluir del mismo. Se hace uso de los valores del error estándar y el valor *t* asociado a cada uno de los coeficientes del modelo.

El valor de *t* se define como:

Ambos valores normalmente son calculados automáticamente por los programas estadísticos.

La exclusión de aquellos coeficientes que no son importantes en el modelo se hace mediante la hipótesis de que el coeficiente no es significativo, por tanto, si se rechaza esta hipótesis, el coeficiente es significativo y debe permanecer en el modelo. Como regla práctica se dice que es razonable incluir en el modelo un parámetro cuya estadística *t* absoluta es mayor que 2. Si el valor de *t* absoluto de un coeficiente es menor a 2, entonces debemos considerar seriamente, excluir el parámetro del modelo.

d. El modelo debe proporcionar un adecuado ajuste: Aplicar la modelación ARIMA a una serie de tiempo no es garantía de que se obtendrá un modelo que ajuste a la perfección, en especial en aquellas series donde existe gran cantidad de perturbaciones aleatorias que no puedan ser ajustadas al modelo. Suele suceder que, para una misma serie de tiempo, existen varios modelos que cumplen con las características de un modelo adecuado y proporcionan resultados similares, sin embargo, debe escogerse aquel que se ajuste mejor a la serie.

Una medida útil que ayuda a conocer el grado de ajuste del modelo a la serie es la ***RMSE*** (root-mean-squared error). Esta medida da a conocer la desviación estándar de los residuales , y se calcula como:

Donde:

: residuales del modelo

numero de residuales

numero de coeficientes del modelo

Generalmente se utiliza el valor de RMSE para compararlo con el de otros modelos estimados para la misma serie. Se preferirá aquel modelo cuyo RMSE tienda a tener menor valor.

**4. Evaluación del modelo**

Una vez estimados los coeficientes del modelo propuesto, la siguiente etapa en la metodología Box-Jenkins es la evaluación ó verificación del mismo. En este paso se comprueba la eficiencia del modelo y se decide si es estadísticamente adecuado.

Un modelo estadísticamente adecuado es aquel cuyos residuales son independientes entre sí. Es decir, si los residuales son completamente aleatorios.

Para comprender mejor el objetivo de esta etapa, es conveniente recordar algunas ideas importantes:

a. La idea básica de la metodología es modelar aquellos datos que estén correlacionados entre sí, mediante la combinación de términos AR y MA.

b. Los fenómenos reales siempre presentan perturbaciones aleatorias , también llamados choques aleatorios ó proceso de ruido blanco.

c. La metodología Box-Jenkins asume que las perturbaciones aleatorias pertenecientes a un proceso son independientes entre sí. No existe correlación alguna entre ellas. Por tanto, estas perturbaciones no se modelan mediante términos AR o MA.

En la práctica las perturbaciones aleatorias no se pueden identificar directamente en una serie de tiempo. Sin embargo, los residuales del modelo ARIMA, nos proporcionan un cálculo aproximado de las perturbaciones reales (Pankratz 1983).

Recordemos que los residuales de cualquier modelo ARIMA se definen como:

Donde:

: residual en el tiempo t

serie original

serie calculada con los parámetros estimados

Entonces, si los residuales muestran estar correlacionados entre sí, significa que existe un patrón que aún no ha sido tomado en cuenta por los términos autorregresivos y/o de medias móviles del modelo propuesto.

Por lo tanto, se concluye que, si los residuales están correlacionados de alguna manera, estos no son ruido blanco y se debe buscar otro modelo cuyos residuales sean completamente aleatorios.

**La función de autocorrelación simple de los residuales** es el instrumento que se utiliza para determinar si el modelo es estadísticamente adecuado. El cálculo de los coeficientes de autocorrelación se realiza por medio de la misma fórmula de autocorrelación, pero aplicada a los residuales del modelo ajustado (Pankratz 1983):

Donde:

: coeficiente de autocorrelación simple residual para un retraso de k periodos

: media de los residuales

: residual en el periodo

: residual en el periodo con k retrasos

n: número total de residuales

Mediante el análisis de esta función se busca que los coeficientes de autocorrelación sean cero o muy cercanos a él. Específicamente se pretende que los coeficientes no sean significativos. Esto se determina a través del estadístico *t* asociado a la función de autocorrelación simple de los residuales:

La mayoría de los programas de computadora realiza automáticamente el cálculo del estadístico *t*. Se despliega el gráfico de la función de autocorrelación de los residuales, tal como se observa en la Figura 22.

Por medio del gráfico y de las bandas se puede establecer si el valor de un coeficiente es significativo. Si uno o varios valores de autocorrelación sobrepasan las bandas de dos desviaciones estándar, significa que el ó los coeficientes son significativamente distintos de cero y, por lo tanto, los residuales no son independientes entre sí. El modelo no es estadísticamente adecuado. De modo contrario, si los valores de los coeficientes no sobrepasan la banda de puntos, indica que dichos valores no son significativos y por tanto se asume que los residuales son independientes entre sí y se acepta el modelo.

También una regla común es que, para los tres primeros rezagos de tiempo, si el valor del estadístico *t* calculado es menor que 1.25 y menor que 1.6 para rezagos posteriores, se concluya que las perturbaciones aleatorias son independientes entre sí (Pankratz 1983).

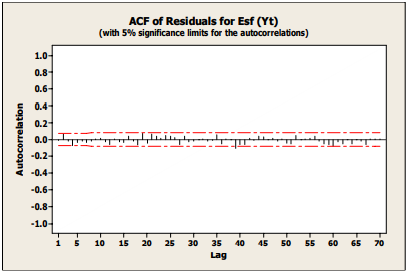


Figura 22 Función de autocorrelación simple de los residuales

Fuente:

Es necesario precisar que el cálculo del estadístico *t* tan sólo es una aproximación del valor real, lo que podría provocar que se consideren los residuos como aleatorios, cuando en realidad no lo son, o viceversa. Por esta razón se recomienda el uso del estadístico Ljung-Box, el cual también se enfoca al análisis de los coeficientes de autocorrelación de los residuales, pero lo hace de manera grupal. Es una prueba que considera un conjunto de *K* coeficientes de autocorrelación.

El estadístico Ljung-Box se define como:

Donde:

n: es el número de observaciones de la serie

d: es el grado de diferenciación

: es el cuadrado de

K: número de autocorrelaciones

El valor de *K* se fija arbitrariamente, la mayoría de los programas de computadora establecen *K* = 6,12,18 y 24.

La forma en que se prueba la eficiencia de un modelo a través de este estadístico sigue el mismo razonamiento, es decir, se trata de comprobar que los residuos no están correlacionados y, por tanto, las autocorrelaciones de los residuos deben ser pequeñas. Por consiguiente, debe ser pequeña. Ahora bien, juzgar que el valor de es un valor pequeño, es sumamente subjetivo. Es por eso que se utilizan ciertas condiciones para establecer un punto de rechazo de la suficiencia del modelo. Se utiliza una escala de distribución chi-cuadrado, a la cual se asocia un valor *p* que representa la probabilidad de juzgar de manera correcta el valor de .Se puede demostrar que, si el valor de *p* es pequeño, el valor de es grande y el modelo no es adecuado y viceversa (Bowerman, O’Conell, Koehler 2007).

De forma práctica, si el valor de *p* es mayor que 0.01 pero menor que 0.05, se considera que el modelo es inaceptable sin lugar a dudas. Si el valor de *p* es mayor que 0.05 entonces se dice que es razonable concluir que el modelo es suficiente. No es del alcance de esta tesis demostrar los valores anteriores, pero se consideran aceptables y válidos debido a la amplia difusión que tienen en la literatura existente.

También cabe destacar que la razón de mencionar el valor de *p* para analizar la suficiencia de un modelo es que en los paquetes informáticos se incluye con frecuencia este valor. Además, es una manera práctica y rápida de evaluar la suficiencia del modelo. Una manera de comprobar que el modelo es estadísticamente adecuado.

**5. Pronóstico**

La fase final de la metodología Box-Jenkins es pronosticar valores futuros de la serie de tiempo. En primer lugar, se debe comprender cómo realizar predicciones puntuales a partir de un modelo ARIMA ya estimado. En segundo lugar, se debe estudiar cómo estimar límites de probabilidad alrededor de una estimación puntual, conocidos particularmente como intervalos de confianza. Y por último se debe conocer y probar la suficiencia de los pronósticos.

A partir de este punto y con el fin de explicar la generación de pronósticos, se asumirá que se conoce al modelo ARIMA en su totalidad, es decir, se conoce sus coeficientes y se ha probado que el modelo es adecuado. Se hace notar que, en la práctica, estas consideraciones son verdaderas si se ha identificado y estimado los parámetros del modelo conforme a la aplicación de las etapas anteriores.

El pronóstico de una serie de tiempo se hace por medio de predicciones ó estimaciones puntuales. Así, una predicción puntual se define como el valor de la variable *Y* en el tiempo *t*, calculado por medio del modelo ARIMA ajustado a la serie.

La forma más conveniente para empezar a realizar predicciones puntuales es escribir el modelo incluyendo el restablecimiento de la no estacionariedad. Hay que recordar que, en el primer paso de la metodología, se induce la estacionariedad en la serie de tiempo (sólo en caso que se requiera), por medio de diferencias. Para restablecer las características originales de la serie, se realiza la operación inversa, es decir, se integra a la serie. A este proceso inverso se le conoce como integración. Se advierte la importancia del restablecimiento de condiciones iniciales, ya que es lógico que se requieran para obtener pronósticos adecuados y apegados a la realidad.

Para ejemplificar cómo se escribe un modelo considerando la integración de la serie, tomemos un modelo ARIMA (1,1,0) cuya ecuación es:

Donde:

: primeras diferencias de la serie

Simplemente se sustituye en la ecuación del modelo:

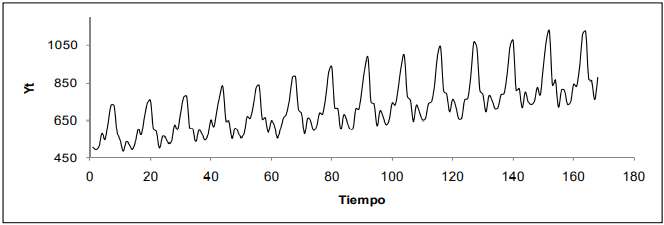
Despejando :

**2.2.4. Metodología Box-Jenkins Estacional**

La metodología vista hasta ahora, está enfocada a series de tiempo no estacionales. Sin embargo, en la práctica se presentan frecuentemente series con estacionalidad. Este hecho no se había remarcado anteriormente debido a que las etapas de la metodología para series estacionales no varían en su concepto y siguen las mismas pautas.

Las variaciones que deben tenerse en cuenta al aplicar modelos ARIMA a series estacionales consisten principalmente en tres aspectos: el reconocimiento del periodo estacional, la estabilización de la serie y la adición de términos autorregresivos y/o de medias móviles referidos a *L* periodos de tiempo anteriores donde L = periodo estacional. A continuación, se presentan aquellas consideraciones extras que deben hacerse en cada una de las etapas.

**1. Estacionariedad**: Uno de los requisitos para poder iniciar la identificación y estimación del modelo es que la serie debe ser estacionaria. Recordemos que, si la serie presenta algún tipo de patrón ascendente o descendente, la serie es inestable o no estacionaria, por lo que la serie debe ser transformada para lograr su estabilización. La Figura 23 muestra una serie estacional donde se muestra un aumento en los valores de la serie.

Figura 23 Serie de tiempo no estacionaria

**Fuente:**

Para series estacionales, la estabilidad se logra por medio de las transformaciones siguientes:

• Diferenciación regular: Es la el mismo tipo de diferenciación presentado en la sección. Al igual que en series no estacionales, se puede requerir las primeras o las segundas diferencias de la serie.

Primeras diferencias

• Diferencias estacionales: Se obtienen a partir de restar el valor de la variable L periodos

anteriores.

• Combinación de ambas:

El periodo estacional *L* se obtiene de acuerdo a lo visto en la modelación clásica, con la ayuda de las funciones de autocorrelación o en su caso con el periodograma.

Para conocer si una serie estacional es estacionaria desde un principio o bien una vez aplicada alguna de las transformaciones anteriores, se sigue el mismo razonar que el visto en la sección anterior.

Si la función de autocorrelación simple de los valores etc. de la serie de tiempo se corta claramente con rapidez o si se corta rápidamente para los primeros retrasos (1,2, 3) y para los retrasos de *L* periodos de tiempo (*L, 2L, 3L, etc*.), entonces se deben considerar que los valores de la serie estacional son estacionarios. (Bowerman, O’Conell, Koehler 2007)

Si la función de autocorrelación simple de los valores etc. para los primeros retrasos (1,2, 3) y para los retrasos de *L* periodos de tiempo (L, 2L, 3L, etc.) de la serie de tiempo, se cortan con lentitud extrema, entonces se deben considerar que los valores de la serie de tiempo son no estacionarios

Además de estas transformaciones, normalmente en series estacionales, se suelen aplicar transformaciones del tipo:

Estas transformaciones se deben a la variabilidad que puede existir entre distintos periodos de tiempo. La gráfica de la figura anterior (Figura 4.23) muestra el ejemplo de una serie estacional con variabilidad estacional, es decir, se tiene un incremento en el rango de valores en cada periodo de tiempo.

Ahora bien, las Figuras 24 y 25 muestran la transformación de la misma serie, utilizando y respectivamente. Se puede observar que la última transformación quita totalmente la variación estacional presente en la serie, mientras que la otra tan sólo la atenúa ligeramente.

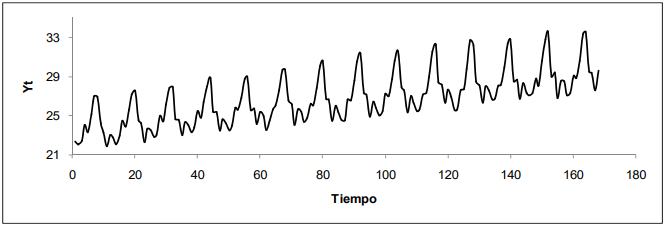


Figura 24 Raíz cuadrada de

Fuente:

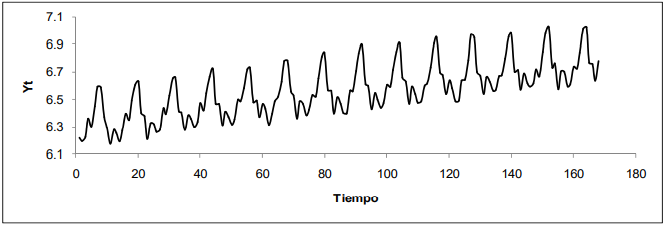


Figura 25 Logaritmo natural de

Fuente:

**2. Identificación:** Para series estacionales, se realiza de la misma forma que para una serie no estacional, es decir, observando el comportamiento de la FAS y la FAP, tanto para los primeros periodos de tiempo (t = 1, 2, 3, 4, 5) como para aquellos alejados *L* periodos de tiempo (*t = L, 2L, 3L, 4L*).

El modelo propuesto para series estacionales estará conformado tanto por coeficientes autorregresivos y de media móvil ordinarios, como de coeficientes autorregresivos y medias móviles estacionales de orden P y Q respectivamente.

Un modelo autorregresivo estacional de orden P sería:

Se puede demostrar que la FAS se extingue en los retrasos *L, 2L, 3L*,…,etc., mientras que la FAP tiene autocorrelaciones significativas (espigas) en los retrasos *L, 2L,…, PL*.

De forma similar, un modelo de media móvil estacional de orden Q sería:

En este modelo la FAS tiene autocorrelaciones significativas (espigas) en los retrasos *L, 2L,…,QL* y autocorrelaciones no significativas en los demás. La FAP se extingue en los retrasos de tiempo *L, 2L, 3L*,…, *etc.* (Bowerman, O’Conell, Koehler 2007).

**3. Estimación, Evaluación y Pronóstico**: Las tres etapas finales se efectúan de la misma manera que el visto para series no estacionales. Únicamente en el estadístico Ljung-Box, se calcula mediante la fórmula:

Donde:

d: grado de diferenciación regular

D: grado de diferenciación estacional

L: periodo estacional

**2.2.5. Ventajas y desventajas de la modelación ARIMA**

La modelación ARIMA presenta importantes ventajas frente a la modelación clásica. La principal de ellas es el gran grado de ajuste que proporciona a la mayoría de las series de tiempo. A diferencia de la modelación clásica en donde se ajusta una serie a un modelo matemático ya establecido, los modelos ARIMA se ajustan a una serie en particular.

Utilizando la metáfora, como lenguaje, la modelación ARIMA viene a ser como la confección de un traje que debe ser ajustado a una serie de tiempo particular. De acuerdo a esta analogía, el analista de series de tiempo utiliza instrumentos (los modelos estadísticos), materiales (los datos) y planes (estrategias de construcción del modelo) para la definitiva especificación del modelo.

Otras ventajas de los modelos ARIMA sobre los métodos clásicos de tratamiento de series de tiempo son:

• Los conceptos que se utilizan para la modelación ARIMA, se derivan de sólidas teorías de la probabilidad clásica y de la estadística matemática

• Los modelos ARIMA son una familia de modelos, no simplemente un único modelo

• Box y Jenkins desarrollaron una completa estrategia que sirve de guía para escoger un

apropiado modelo dentro de esta gama de modelos

• Un apropiado modelo ARIMA produce óptimas predicciones

Por otro lado, dentro de las dificultades que se pueden presentar en la modelación ARIMA, se encuentra el problema que supone para el analista, recién iniciado en la teoría y la ejecución de la metodología Box-Jenkins, la correcta elección de un modelo adecuado en los primeros intentos. Es por ello que la modelación de un proceso ARIMA ha sido referida por algunos autores como un arte. Quizá sería preferible utilizar el término habilidad, debido a que las técnicas básicas en la modelación de los procesos ARIMA son fácilmente accesibles y ajustables tras un aprendizaje relativamente mediano.

Una desventaja más es que, aunque se utilizan programas de cómputo en el cálculo de los coeficientes de un modelo ARIMA probable, la aplicación de la metodología Box-Jenkins sigue siendo una labor manual, es decir, requiere ser ejecutada por el analista. Hay que recordar que no cualquier modelo propuesto proporciona necesariamente el ajuste adecuado.

**2.2.6. Modelos ARIMA con variables de Intervención**

Aplicación de modelos ARIMA con variables de intervención, debido a que la serie temporal a predecir puede verse afectada por fenómenos externos, tales como inundaciones o huelgas, por ejemplo.

Estos fenómenos intervienen en el comportamiento de la serie, por lo que se debe evaluar

se efecto mediante la introducción en el modelo ARIMA de variables artificiales binarias. Por lo tanto, se incorporan como variables Dummy en forma de impulsos temporales y escalones. Si no se incorporan dichas variables se pueden producir sesgos en las estimaciones, así como un elevado error en las predicciones.

En el análisis realizado en este proyecto, se puede reconocer dos grandes grupos.

Variables laboralidad: Son variables de tipo impulso o de tipo escalón que recogen periodos que afectan directamente a la demanda de Gas.

Variables calendario: Marcan la estacionalidad en el comportamiento de la demanda, ya sea mensual o diaria.

**2.2.7. Modelos ARMA con variables externas (ARMAX)**

Además de incorporar valores autorregresivos y de media móvil, se pueden incorporar variables externas en forma de regresores. Son variables que provienen de una fuente externa distinta a la de la serie de predicción.

Dichos modelos presentan los siguientes parámetros: ARIMA(p,q,n), siendo *p* la parte autorregresiva, *q* la parte media móvil y *n* las variables externas. La nueva ecuación quedaría de la siguiente forma:

**CAPITULO III: METODOLOGIA, FORMULACION DE LOS MODELOS PROYECTIVOS Y ANALISIS DE LAS VARIABLES OBJETIVOS**

**3.1. DATOS**

El cálculo se realiza con la utilización del Software EViews y con los datos de demanda histórica para el período de enero 2005 a diciembre 2021, en resolución anual, es decir la serie se compone de 17 observaciones, siendo el horizonte de predicción entre enero 2022 a diciembre 2026. Las variables socioeconómicas (explicativas) se proyectaron con el modelo de Hold y Winters sin periodicidad del 2022 al 2026.

**3.2. MODELO DE REGRESIÓN UNIECUACIONAL**

Las variables que se utilizaron en el modelo son:

* DEM: Demanda de gas natural seco de Camisea para generadores eléctricos (MMPC)
* PBI: Producto bruto interno del Perú (MMS)
* PGT: Producción total de gas natural seco de Camisea (MMPC)
* PET: Producción eléctrica total del SEIN (GWh)
* PEG: Producción eléctrica del gas natural seco de Camisea (GWh)
* TAR: Tarifa de gas natural seco en boca de pozo para generadores eléctricos (US$/MMBTU)

En teoría, la relación entre la demanda (DEM) el PBI y las otras variables, se investiga bajo el supuesto de que la economía doméstica produce bienes que se usan para el consumo (Dornbush, 1980); (Rose, 1990). Sin embargo, en esta investigación la demanda se mide como la relación con las otras variables, la relación es preferible ya que no es sensible a unidades de medida y se puede interpretar como demanda nominal o real (Bahmani-Oskooee, 1991).

Entonces, la demanda se basa en una función que depende directamente del PBI, PGT, PET, PEG, TAR, se puede escribir de un modo general:

[1]

Una función concreta muy usual es la lineal, o sea

[2]

y otro tipo es la potencial generalizada, o función de Cobb-Douglas,

[3]

En que como es sabido, dan las elasticidades (constantes siempre) de la variable DEM con respecto a las variables explicativas

; ; ; ; [4]

Tomando logaritmos en [3] se tiene que es una función lineal cuyos parámetros se cuantifican del mismo modo que [2].

[5]

Se hace preciso, pues, reconocer el carácter estocástico de las relaciones, introduciendo

un término de perturbación de tipo aleatorio que designaremos por ɛ. Así se tiene que [2] se convierte en [6] y [5] en [7]

[6]

[7]

Donde:

: Coeficientes de las variables

: Termino de perturbación aleatoria

c: Intercepto

Se dice de una serie temporal es estacionaria si

**3.2.1. Prueba de estacionariedad**

**a) Prueba visual**

Observar en gráficas, estas pruebas no formales no son definitivas, si la serie muestra tendencia entonces sugiere una variación en la media, lo cual puede indicar la no estacionariedad.

Visualizar el correlograma, el correlograma es una representación de la función de autocorrelación, un correlograma que desciende lentamente es típico de variables no estacionarias. Un correlograma que desciende rápidamente o cuasi aleatorio es típico de variables estacionarias.

**b) Prueba de Dickey-Fuller aumentada**

Es un test formal exigente. Tiene la ventaja de que la hipótesis nula no es si la serie es o no ruido blanco, sino si tiene una raíz unitaria.

Donde , y pasamos el test

, si no puede rechazarse la nula (p-valor > 0.05) la serie es no estacionaria y tiene raíz 1

, si se rechaza la nula (p-valor < 0.05) la serie es estacionaria y tiene raíz 0

**3.2.2. Colinealidad entre variables explicativas**

Otra hipótesis de interés es la relativa a la inexistencia de colinealidad entre las variables independientes o explicativas. O sea, las variables deben ser independientes entre sí. Cuando alguno de los coeficientes de correlación es elevado (próximo a ±1) tendremos un indicio de la existencia de colinealidad imperfecta en el modelo.

El Método basado en la correlación muestral entre variables explicativas: Es calcular la correlación lineal simple existente entre pares de variables explicativas. Si hacemos esto para los k regresores del modelo, obtenemos una matriz R con la forma:

**3.2.3. Procedimiento de cointegracion: Engle-Granger**

Se formulan dos ecuaciones una primera ecuación de largo plazo tal que genere un determinado error y una segunda ecuación de corrección de error o de corto plazo.

En términos de aplicación como se hace un análisis de cointegracion:

1. Se formula una primera ecuación de largo plazo, sacando logaritmos a todas las variables, a esa primera ecuación de regresión sus residuos van a estar autocorrelados, entonces, los residuos que tenga esta primera ecuación es precisamente los errores de cointegracion que es la que nosotros vamos a corregir.

Formulación de la ecuación de largo plazo:

2. Se formula una segunda ecuación de corrección de error, sacándole una diferenciación a todas sus variables directas, pero considerando un rezago adicional de 1 y el error también rezagado en 1, en la medida que le sacamos diferenciación a los regresores la *c* tiende a cero o sea el intercepto tiende a cero.

3. Una vez logrado que las variables estén cointegradas, con los regresores apropiados por un lado consistentes por otro lado representativos con un D.W. cercano a 2 y R2 **≈** 1 o en diferencias R2 **≈** 0.65

Formulación de la ecuación de corrección de error o de corto plazo:

4. La proyección, la ecuación de largo plazo lo que ve es el comportamiento evolutivo tendencial de la serie, entonces lo que ajusta esta ecuación es a esa expresión, si la tendencia tiene una determinada pendiente o característica, eso se va reflejar en el futuro sin detalle, y ese cuanto en detalle lo explica la ecuación de corto plazo ( 1 o 2 años), marca la pauta de como ira, esta ecuación de proyección mira lo mas cercano y sobre esa base nosotros hacemos la proyección.

**3.2.4. Causalidad de Granger**

Una vez confirmado que las series en estudio son integradas de orden 1 será necesario estudiar si efectivamente existe una causalidad entre las variables y .

Para el objetivo de este trabajo solo será relevante si existe causalidad en una sola dirección o si existe en ambas direcciones, también llamada causalidad bilateral . Para ello haremos uso de las herramientas definidas anteriormente en la sección anterior.

## 3.3. ANALISIS DE LAS VARIABLES OBJETIVOS

Cuando se formula un modelo econométrico que explique a la demanda de un determinado país en función de todas sus variables económicas, sociales y también técnicas, supuestamente se está garantizando el mejor modelo, porque se está involucrando a todas las variables.

Pero cuando se le aplican los condicionante económicos y estadísticos no cumplen, de tal forma que un gran trabajo de investigación que arrancó con n variables explicativas se queda solamente con el PBI y otra variable.

Entonces el arte de potenciar los alcances del aprendizaje y la investigación es justamente de ir descartando una por una todas las variables explicativas en función de los condicionantes estadístico y económico, ese es el procedimiento, no cuanto más variables explicativas tenga mi modelo será mejor, hay países que su demanda es explicada solo por su PBI.

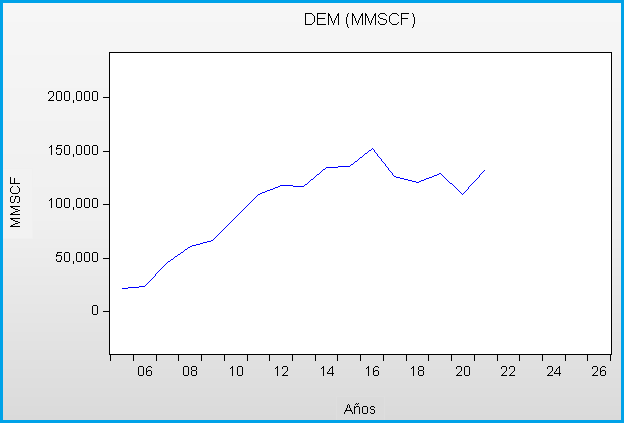
Esto no significa que no tenga dependencia de las demás variables, sino que el efecto de uno lo adsorbe al del otro y lo que buscamos es un elemento de contraste, es decir queremos explicar indirectamente con cualquiera de las variables a la demanda y resulta que una es más predominante y por lo tanto esa es la que me va explicar de manera más aproximada a la demanda.

### 3.3.1. Variable Dependiente: demanda de gas natural seco para generadores eléctricos

La demanda de gas es una serie anual estable en el tiempo por eso ello que se tomó para realizar el modelo econométrico. La estructura de la demanda de gas en el Perú a nivel de generación, la estadística lo manejan los generadores, a nivel de consumo las ventas son reportadas al COES, al Osinergim en resolución mensual.

Esta será la variable dependiente u objetivo del modelo econométrico a formular, esta demanda está conformada principalmente por el consumo histórico de las centrales termoeléctricas por empresas, que en su conjunto representan el mayor volumen de la demanda total del SEIN y es publicado por el COES. Los datos se encuentran disponibles en su página web: <https://www.coes.org.pe/Portal/Publicaciones/Estadisticas/>.

Se quiere que esta demanda de gas sea explicada por otras variables energéticas y no energéticas, que pueden ser sociales y económicas y que obedezcan a la teoría económica, pero que también cumplan con todos los estadísticos de medición y contraste. En la Figura 26 se ilustra el comportamiento histórico 2005 al 2021 de la demanda de gas para generadores eléctricos en (MMPC).



**Figura 26: Demanda de GN seco para GE (2005 - 2021) en forma anual**

**Fuente:** **COES**

### 3.3.2. Variable Independiente: Producto Bruto Interno

Un factor importante de las variables independientes, es que se parte de la premisa que son más accesibles a conocer el futuro de cada una de ellas, son más fáciles de proyectar que la propia demanda (variable dependiente). Por otro lado, diríamos que estas proyecciones, se encuentran publicados o elaborados por otras fuentes de confianza en cuanto a valores proyectados en el medio.

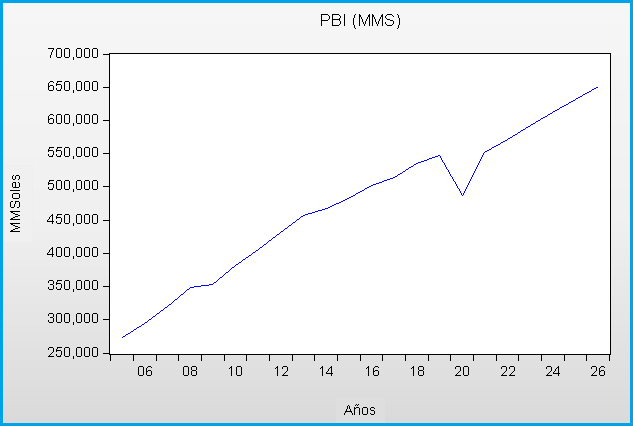
Teniendo en cuenta que el PBI si guarda relación con la demanda y teniendo en cuenta que la demanda de gas es un bien necesario que tiene que ser consumido pero ese bien para ser consumido, tiene que tener una persona o la organización su poder adquisitivo para adquirir ese bien, entonces del pbi podemos decir del punto de vista de la ley económica si guarda relación con la demanda.

Ahora para el PBI nadie se arriesga, ninguna identidad privada ni estatal, ni de corte internacional en hacer proyecciones de largo plazo regional o nacional, el banco mundial tiene una proyección de los próximos dos años, la banca privada (BCP, Scoctiabank) publican solamente de año y medio, y cuando nosotros nos preguntamos se puede o no proyectar esto frente a una circunstancia de la demanda de gas y si alguien podría hacerlo.

Algún estudio podría hacer una aproximación, ante esta situación, si existen quien lo pueda hacer, por ejemplo, apoyo consulting es una empresa confiable en el Perú de las estadísticas que se manejan y también del proceso de encuestado y las proyecciones correspondientes, ellos pueden calcular el PBI con un determinado margen de error, entonces hay quien lo pueda hacer, entonces diríamos que ya tenemos esta proyección.

El PBI es la variable independiente del modelo econométrico a formular. Se utiliza la variable PBI porque se considera que la actividad económica del SEIN está relacionada de manera directa con la actividad económica total del país y de la demanda interconectada. En ese sentido, se tomará como datos de entrada la serie histórica del PBI (a precios constantes de 2007) de los últimos 17 años publicados por el Banco Central de Reserva del Perú (BCRP). Dichos datos se encuentran disponibles en su página web: https://www.bcrp.gob.pe/estadisticas.html

En cuanto a la proyección de esta variable, se ha considerado los valores estimados en el reporte de inflación a diciembre de 2018 publicado por el BCRP. En la Figura 27 se muestra el comportamiento histórico del PBI nacional en millones de soles desde el año 2005 al 2026 contabilizada en resolución anual.



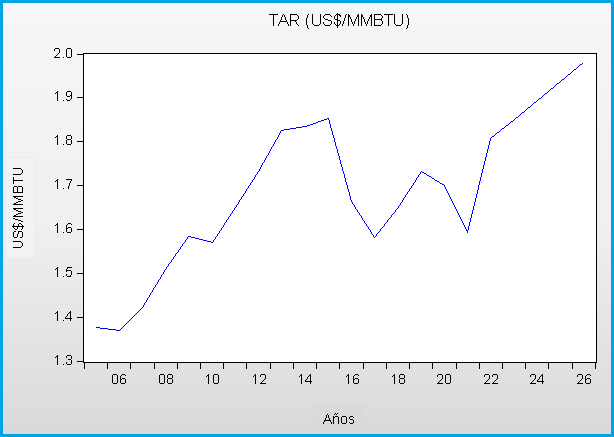
**Figura 27: Evolución del PBI (2005 – 2026)**

**Fuente:** **BCRP**

### 3.3.3. Variable Independiente: Precio del gas en Boca de Pozo

Las empresas eléctricas y el propio estado en su plan de fiscalización, control y planificación de la energía eléctrica, así como los vendedores, trasportadores y distribuidores manejan valores de la tarifa de gas en boca de pozo, conscientemente también esta proyección es alcanzable, entonces ya se tendría.

Es la variable independiente del modelo econométrico a formular. Se ha considerado para el período de evaluación la tarifa de energía de los clientes regulados publicado por el OSINERGMIN y que se reajusta a partir del primer día de cada año calendario. La fuente es la página web: <https://bit.ly/3K1k00S>, en la Figura 28 se observa la proyección del precio del gas en boca de pozo del 2005 a 2026contabilizada en resolución anual.



**Figura 28: Proyección del Precio del gas en boca de pozo**

**Fuente: OSINERGMIN**

### 3.3.4. Variable Independiente: Precio del Petróleo

Es la variable independiente del modelo econométrico a formular. Tanto el precio del petróleo Brent como el WTI son los precios más representativos del petróleo a nivel mundial, si bien en Europa el gas natural de Siberia abastece la mayoría de los países europeos, ese precio del gas natural si guarda una relación con el precio del petróleo, pero en América del sur, no guarda una relación por las características del mercado peruano.

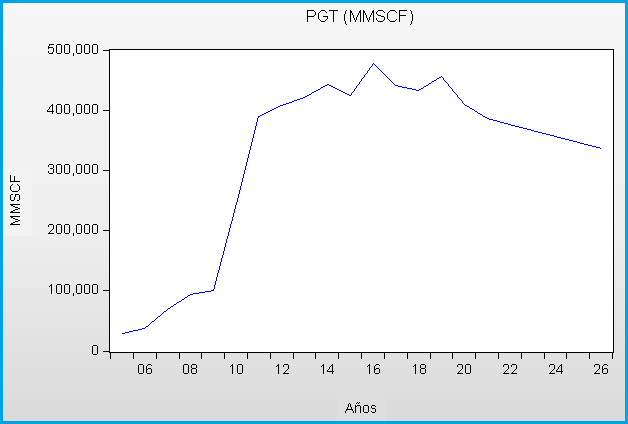
Por este motivo y por la curva de proyección errática que muestra no se consideró como variable explicativa.

### 3.4.5. Variable Independiente: Producción Total de gas natural seco de Camisea

Es la variable independiente del modelo econométrico a formular. Su estadística histórica es del año 2005 al 2021 publicado por Perupetro. La fuente, su página web <https://cutt.ly/AO9pJqP>

En la Figura 29 se ilustra el comportamiento histórico desde el año 2005 al 2026 de la producción fiscalizada de gas natural total seco de Camisea en resolución anual.

.



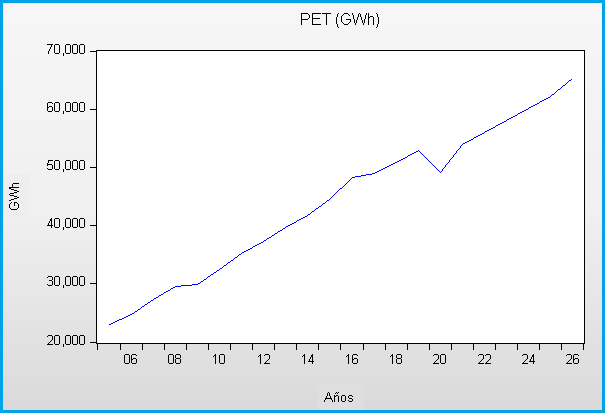
**Figura 29: Proyección de la Producción Total de G.N.**

**Fuente:** **Perupetro Estadísticas anual de hidrocarburos 2020**

### 3.3.6. Variable Independiente: Producción Eléctrica Total del SEIN

Es la variable independiente. La estadística histórica se tiene desde el año 2005 al 2021. La fuente: https://www.coes.org.pe/Portal/areas/Publicaciones/documentos/capitulo04.htm

En la Figura 30 se ilustra el comportamiento histórico desde el año 2005 a 2026 de la producción eléctrica total del SEIN en GWh, en resolución anual.



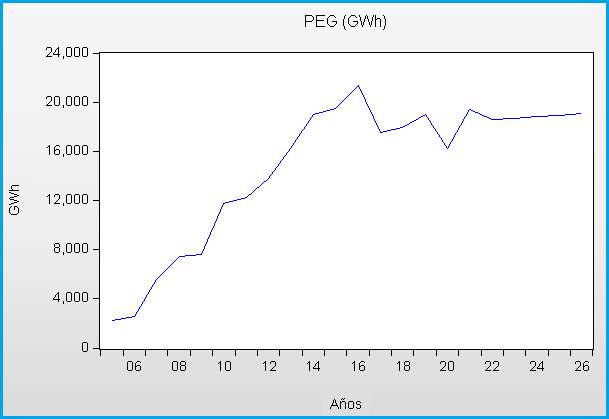
**Figura 30: Proyección de la Producción Eléctrica Total del SEIN**

**Fuente:** **COES**

### 3.3.7. Variable Independiente: Producción Eléctrica con gas natural seco de Camisea

#### Es la variable independiente del modelo econométrico a formular. Su estadística histórica se tiene desde el año 2005 al 2021 en resolución anual, publicado por el Ministerio de energía y minas.

La fuente: <http://www.minem.gob.pe/_estadisticaSector.php?idSector=5> . En la Figura 31 se ilustra el comportamiento histórico de la producción eléctrica con gas natural seco de Camisea en GWh, desde el año 2005 a 2026 en resolución anual.



**Figura 31: Proyección de la Producción Eléctrica con G.N. seco**

**Fuente: MINEM**

**3.4. MODELO UNIVARIANTE ARIMA**

### 3.4.1. Origen y estructura de los datos

Los datos utilizados para el estudio son los proporcionados por el Comité de Operaciones del Sistema Interconectado Nacional (COES). dichos datos se encuentran disponibles en su página web <https://www.coes.org.pe/Portal/Publicaciones/Estadisticas/>. Luego de obtener la base de datos total del consumo de energía registrado por el COES y habiendo realizado un análisis previo de la calidad de información, se seleccionó la demanda de gas para centrales térmicas exclusivamente.

La base de datos final se trabajó en Microsoft Excel, y luego mediante tablas se obtuvo la demanda de energía en resolución mensual para la proyección. A continuación, se detallarán las sucesivas etapas que nos permitirán determinar el modelo que mejor se ajusta al comportamiento de la serie de datos.

### 3.4.2. Procedimiento

Con objeto de exponer de forma concisa el procedimiento aplicativo de la teoría estocástica mediante análisis de series temporales ARIMA Univariante que seguiremos para lograr nuestra estimación con un nivel de ajuste adecuado, se detallan brevemente los pasos:

**1. Visualización de la serie**

Se visualizar la serie en niveles (serie vs tiempo) y las autocorrelaciones (fac y facp), para ver si es estacionaria, con lo cual se concluye que la serie de demanda de energía es un proceso estocástico no estacionario ni lineal (la serie no es estacionaria).

**2. Transformaciones para estabilizar la serie**

Mediante transformaciones se puede lograr una serie estacionaria. Para lo cual se debe lograr que la serie tenga:

* Estabilidad en Varianza (λ).
* Estabilidad en Media Regular (d)
* Estabilidad en Media estacional (D)

Así, el modelo ARIMA se puede expresar como:

Donde µ es la media de con potencia de transformación λ, siendo

**3. Identificación de las órdenes autorregresivas y medias móviles del modelo ARIMA**

* Las ordenes autorregresivas regulares (p).
* Las ordenes autorregresivas estacionales (P).
* Las órdenes a medias móviles regulares (q).
* Las órdenes a medias móviles estacionales (Q).

Determinación del modelo ARIMA, mediante los resultados obtenidos en procedimiento anterior de la forma y la función analítica del modelo ARIMA calculado.

**4. Estimación de parámetros y pronóstico**

Se estimará los parámetros y c de la función analítica*.* Finalmente,la predicción de la demanda de energía para los próximos cinco años.

**5. validación de resultados**.

Se procederá mediante dos formas:

* **Fijación de horizonte de validación**

Fijar un horizonte histórico menor que el horizonte real. Se procederá a aplicar el modelo ARIMA para el horizonte seleccionado. Se determina el pronóstico para el horizonte de validación. Se determina el error absoluto medio porcentual MAPE.

* **Validación de modelo ARIMA óptimo mediante la prueba de residuos**

Consistirá en verificar a través de la gráfica autocorrelaciones ACF del error del modelo ARIMA optimo, si los residuos están incorrelados es decir caen dentro de la banda de confianza, entonces se dirá que el error se comporta como ruido blanco.

# CAPITULO IV: ANALISIS Y RESULTADOS DE LA INVESTIGACION

## 4.1. PROYECCION DEL MODELO DE REGRESION UNIECUACIONAL

### 4.1.1. Importación del archivo Excel al programa Eviews y definición de variables

como se muestra en la Tabla 6.

**Tabla 6: Series de las variables del modelo econométrico**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Años** | **PBI(MMS)** | **PGT(MMSCF)** | **TAR(US$/MMBTU)** | **DEM(MMSCF)** | **PEG(GWh)** | **PET(GWh)** |
| 2005 | 273971.15 | 28441 | 1.377 | 22099.13 | 2264.88 | 23001.48 |
| 2006 | 294597.83 | 37584 | 1.369 | 23833.09 | 2568.51 | 24762.78 |
| 2007 | 319693 | 69006 | 1.423 | 45859.97 | 5573.51 | 27254.93 |
| 2008 | 348870 | 93090 | 1.508 | 61317.61 | 7409.85 | 29558.71 |
| 2009 | 352693 | 99723 | 1.583 | 66236.43 | 7643.44 | 29807.25 |
| 2010 | 382081 | 244300 | 1.571 | 88223.71 | 11724.79 | 32426.83 |
| 2011 | 406256 | 388900 | 1.649 | 109446.34 | 12267 | 35217.43 |
| 2012 | 431199 | 407590 | 1.732 | 118129.8 | 13822.9 | 37321.18 |
| 2013 | 456435 | 421160 | 1.827 | 117205.91 | 16251.83 | 39669.43 |
| 2014 | 467308 | 442960 | 1.836 | 134721 | 19012.94 | 41795.89 |
| 2015 | 482506 | 424900 | 1.854 | 136147.25 | 19523.06 | 44540.04 |
| 2016 | 501581 | 477490 | 1.665 | 152233.48 | 21321.43 | 48326.42 |
| 2017 | 514215 | 441120 | 1.582 | 126228.46 | 17533.65 | 48993.25 |
| 2018 | 534665 | 432310 | 1.648 | 120719.51 | 17920.22 | 50816.79 |
| 2019 | 546161 | 455560 | 1.731 | 129550.04 | 18962.9 | 52889.14 |
| 2020 | 485992.07 | 409220 | 1.699 | 109562.62 | 16221.8 | 49186.64 |
| 2021 | 551829 | 385420 | 1.593 | 132955.15 | 19366.2 | 53990.35 |
| 2022 | 571694.84 | 375535 | 1.808 |  | 18573.74 | 55925.16 |
| 2023 | 591704.16 | 365651.2 | 1.852 |  | 18695.43 | 58008.65 |
| 2024 | 611230.4 | 355766.8 | 1.895 |  | 18817.12 | 60092.14 |
| 2025 | 63078.77 | 345882.4 | 1.938 |  | 18938.8 | 62175.64 |
| 2026 | 650344.26 | 335997.95 | 1.981 |  | 19060.49 | 65197.61 |

**Fuente: Elaboración propia**

### 4.1.2. Especificación Funcional

Se propone una formulación uniecuacional de regresión, que relacione la variable dependiente con las variables explicativas y la perturbación aleatoria, dada por la siguiente expresión:

Donde:

DEM: Variable dependiente representativa de la demanda energética

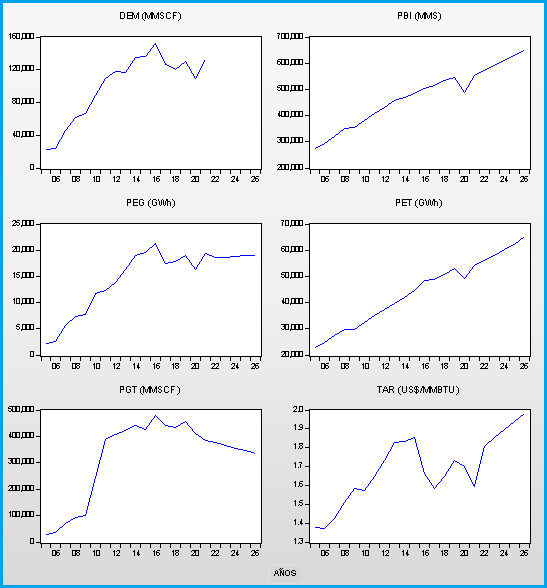
PBI, PEG, PET, PGT, TAR: Variables explicativas

: Términos a ser estimados

: Termino de perturbación estacionaria

**4.1.3. Visualización de la función de cada una de las variables**

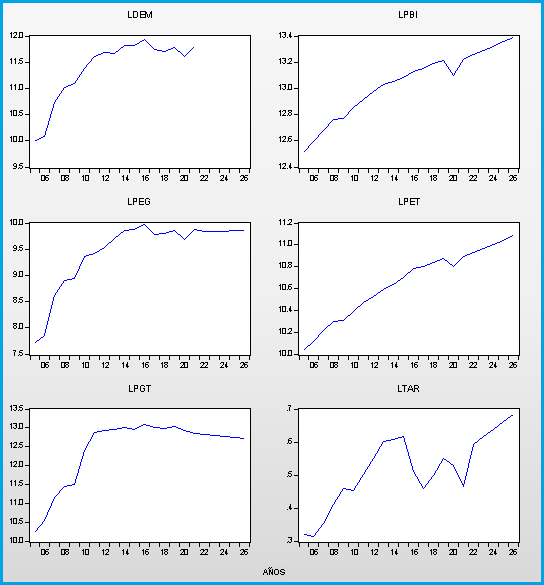
Toda serie de tiempo en resolución anual tiene tendencia, pero no tiene estacionacionalidad, en la Figura 32 se muestra las variables en resolución anual.

**Figura 32: función de cada una de las variables**

**Fuente:** **Elaboración propia**

### 4.1.4. Transformación de las variables

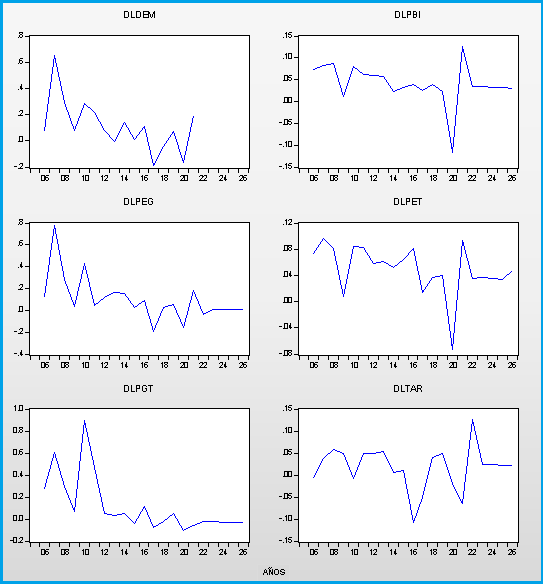
Función logaritmo de cada una de las variables,una forma de linealizar y estacionalizar las variables es sacando el logaritmo a las series, es decir estabilizar una posible variabilidad que se pueda dar, las series no son estacionarias como se muestra en la Figura 33.



**Figura 33 Variables con la función logaritmo**

**Fuente: Elaboración propia**

**Función diferencia de logaritmo de cada una de las variables**

Para seguir estacionalizando las series de las variables calculamos la diferencia logarítmica de las series, en la Figura 34 se muestra que las series se hicieron estacionarias.

**Figura 34: Variables con la función diferencia de logaritmo**

**Fuente: elaboración propia**

**4.1.5. Prueba de estacionariedad**

Se observa que las seis variables se hicieron estacionarias con la diferenciación logarítmica

de orden uno, con estas variables se tiene las condiciones para conformar una ecuación de

representativa, y hacer la prueba de raíz unitaria como se muestra en las siguientes Tablas.

**Tabla 7: prueba de raíz unitaria de DLDEM**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Null Hypothesis: DLDEM has a unit root | | | |  |
| Exogenous: None | | |  |  |
| Lag Length: 2 (Automatic - based on SIC, maxlag=3) | | | | |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  | t-Statistic | Prob.\* |
|  |  |  |  |  |
|  |  |  |  |  |
| Augmented Dickey-Fuller test statistic | | | -2.917873 | 0.0071 |
| Test critical values: | 1% level |  | -2.754993 |  |
|  | 5% level |  | -1.970978 |  |
|  | 10% level |  | -1.603693 |  |
|  |  |  |  |  |
|  |  |  |  |  |
| \*MacKinnon (1996) one-sided p-values. | | | |  |

**Fuente:** **Elaboración propia**

**Tabla 8: prueba de raíz unitaria de DLPBI**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Null Hypothesis: DLPBI has a unit root | | | |  |  |
| Exogenous: Constant | | |  |  |  |
| Lag Length: 0 (Automatic - based on SIC, maxlag=4) | | | | |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  | t-Statistic | Prob.\* |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| Augmented Dickey-Fuller test statistic | | | -5.288405 | 0.0004 |  |
| Test critical values: | 1% level |  | -3.808546 |  |  |
|  | 5% level |  | -3.020686 |  |  |
|  | 10% level |  | -2.650413 |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| \*MacKinnon (1996) one-sided p-values. | | | |  |  |

**Fuente:** **Elaboración propia**

**Tabla 9: prueba de raíz unitaria de DLPEG**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Null Hypothesis: DLPEG has a unit root | | | |  |
| Exogenous: None | | |  |  |
| Lag Length: 2 (Automatic - based on SIC, maxlag=4) | | | | |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  | t-Statistic | Prob.\* |
|  |  |  |  |  |
|  |  |  |  |  |
| Augmented Dickey-Fuller test statistic | | | -4.712078 | 0.0001 |
| Test critical values: | 1% level |  | -2.699769 |  |
|  | 5% level |  | -1.961409 |  |
|  | 10% level |  | -1.606610 |  |
|  |  |  |  |  |
|  |  |  |  |  |
| \*MacKinnon (1996) one-sided p-values. | | | |  |

**Fuente:** **Elaboración propia**

**Tabla 10: prueba de raíz unitaria de DLPET**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Null Hypothesis: DLPET has a unit root | | | |  |
| Exogenous: Constant | | |  |  |
| Lag Length: 0 (Automatic - based on SIC, maxlag=4) | | | | |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  | t-Statistic | Prob.\* |
|  |  |  |  |  |
|  |  |  |  |  |
| Augmented Dickey-Fuller test statistic | | | -4.739354 | 0.0014 |
| Test critical values: | 1% level |  | -3.808546 |  |
|  | 5% level |  | -3.020686 |  |
|  | 10% level |  | -2.650413 |  |
|  |  |  |  |  |
|  |  |  |  |  |
| \*MacKinnon (1996) one-sided p-values. | | | |  |

**Fuente:** **Elaboración propia**

**Tabla 11: prueba de raíz unitaria de DLPGT**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Null Hypothesis: DLPGT has a unit root | | | |  |
| Exogenous: None | | |  |  |
| Lag Length: 0 (Automatic - based on SIC, maxlag=4) | | | | |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  | t-Statistic | Prob.\* |
|  |  |  |  |  |
|  |  |  |  |  |
| Augmented Dickey-Fuller test statistic | | | -2.475099 | 0.0163 |
| Test critical values: | 1% level |  | -2.685718 |  |
|  | 5% level |  | -1.959071 |  |
|  | 10% level |  | -1.607456 |  |
|  |  |  |  |  |
|  |  |  |  |  |
| \*MacKinnon (1996) one-sided p-values. | | | |  |

**Fuente:** **Elaboración propia**

**Tabla 12: prueba de raíz unitaria de DLTAR**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Null Hypothesis: DLTAR has a unit root | | | |  |
| Exogenous: None | | |  |  |
| Lag Length: 0 (Automatic - based on SIC, maxlag=4) | | | | |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  | t-Statistic | Prob.\* |
|  |  |  |  |  |
|  |  |  |  |  |
| Augmented Dickey-Fuller test statistic | | | -3.593189 | 0.0011 |
| Test critical values: | 1% level |  | -2.685718 |  |
|  | 5% level |  | -1.959071 |  |
|  | 10% level |  | -1.607456 |  |
|  |  |  |  |  |
|  |  |  |  |  |
| \*MacKinnon (1996) one-sided p-values. | | | |  |

**Fuente:** **Elaboración propia**

**4.1.6. Determinación de colinealidad entre variables explicativas**

En la prueba de colinealidad al logaritmo de las variables explicativas, se observa valores de correlación cercanos a 1, lo que significa que hay una fuerte relación de correlación entre las variables lo que no es apropiado para hacer un análisis de regresión, como se observa en la Tabla 13.

**Tabla 13: Matriz de correlación de variables con logaritmos**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **LPBI** | **LPEG** | **LPET** | **LPGT** | **LTAR** |
| **LPBI** | 1 | 0.910377 | 0.995098 | 0.860458 | 0.838438 |
| **LPEG** | 0.910377 | 1 | 0.888339 | 0.973021 | 0.837077 |
| **LPET** | 0.995098 | 0.888339 | 1 | 0.83558 | 0.804905 |
| **LPGT** | 0.860458 | 0.973021 | 0.83558 | 1 | 0.806361 |
| **LTAR** | 0.838438 | 0.837077 | 0.804905 | 0.806361 | 1 |

**Fuente:** **Elaboración propia**

En la prueba de colinealidad a la diferencia de logaritmo de las variables explicativas, se observa que ya se puede trabajar con cada una de ellas, como se observa en la Tabla 14.

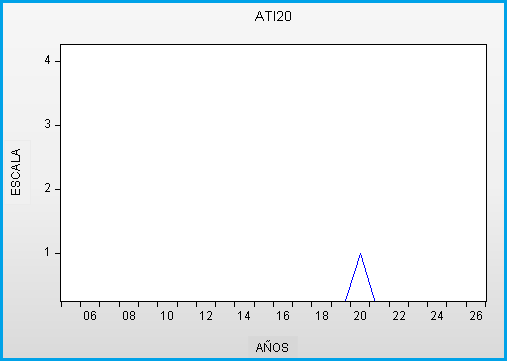
**Tabla 14: Matriz de correlación de variables con diferencia de logaritmos**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **DLPBI** | **DLPEG** | **DLPET** | **DLPGT** | **DLTAR** |
| **DLPBI** | 1 | 0.57758 | 0.922337 | 0.437566 | 0.017769 |
| **DLPEG** | 0.57758 | 1 | 0.667593 | 0.756288 | 0.077921 |
| **DLPET** | 0.922337 | 0.667593 | 1 | 0.554169 | -0.055078 |
| **DLPGT** | 0.437566 | 0.756288 | 0.554169 | 1 | 0.051912 |
| **DLTAR** | 0.017769 | 0.077921 | -0.055078 | 0.051912 | 1 |

**Fuente:** **Elaboración propia**

**4.1.7. Modelo Econométrico Uniecuacional**

Entonces se formula una ecuación de prueba, pero antes se dio un breve vistazo al gráfico de la variable demanda con la finalidad de filtrar los efectos externos o atípicos, se revisó a lo largo del tamaño muestral del 2005 al 2021, y se obtiene el atípico debido al efecto del covid19 en el año 2020, al ser un caso puntual se opta por su filtrado, se creó la función tipo pulso (ATI20), como se muestra en la Figura 35.



**Figura 35: Atípico del año 2020**

**Fuente:** **Elaboración propia**

Entonces una ecuación de regresión tendrá la siguiente forma:

La bondad o grado de ajuste del modelo, medido por el coeficiente de determinación R2 es 0.88, pero el estadístico p-valor es mayor que 0.05, lo cual significa que los residuos de esta regresión tienen un componente de autocorrelación de primer orden, es decir no son un ruido blanco, entonces esta ecuación de prueba no es consistente por lo que iríamos a la teoría de cointegracion formulada por Granger.

La ecuación estimada se muestra en la Tabla 15.

**Tabla 15: Estimación de la ecuación de prueba**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dependent Variable: DLDEM | | |  |  |
| Method: Least Squares | | |  |  |
| Date: 08/27/22 Time: 04:09 | | |  |  |
| Sample (adjusted): 2006 2021 | | |  |  |
| Included observations: 16 after adjustments | | | |  |
|  |  |  |  |  |
|  |  |  |  |  |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|  |  |  |  |  |
|  |  |  |  |  |
| DLPEG | 0.745927 | 0.116676 | 6.393134 | 0.0000 |
| DLPGT | 0.083336 | 0.094542 | 0.881476 | 0.3929 |
|  |  |  |  |  |
|  |  |  |  |  |
| R-squared | 0.882489 | Mean dependent var | | 0.112155 |
| Adjusted R-squared | 0.874095 | S.D. dependent var | | 0.199544 |
| S.E. of regression | 0.070804 | Akaike info criterion | | -2.341328 |
| Sum squared resid | 0.070185 | Schwarz criterion | | -2.244755 |
| Log likelihood | 20.73063 | Hannan-Quinn criter. | | -2.336383 |
| Durbin-Watson stat | 2.623827 |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**Fuente:** **Elaboración propia**

**4.1.8. Análisis de cointegración**

Formulamos un modelo econométrico de naturaleza biecuacional con dos componentes una de largo plazo y otra de corto plazo con variables que sean no necesariamente de origen individual estacionaria.

**Formulación de la ecuación de largo plazo:**

(2)

.

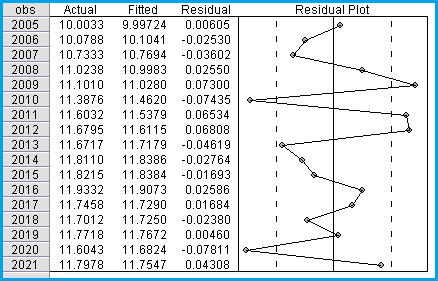
La ecuación de largo plazo estimada se muestra en la tabla 16, se observa que los Prob. son menores que 0.05 y el R2 de 0.99%, pero el Durbin Watson tiene un valor de 2.41 fuera del rango establecido, esta ecuación de cointegracion ha generado un residuo que son los errores de la ecuación de cointegracion y están correlados, bajo ese sentido tengo que extraer esa función error y corregirlo.

**Tabla 16: Ecuación de largo plazo estimada**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dependent Variable: LOG(DEM) | | |  |  |
| Method: Least Squares | | |  |  |
| Date: 07/11/22 Time: 04:42 | | |  |  |
| Sample (adjusted): 2005 2021 | | |  |  |
| Included observations: 17 after adjustments | | | |  |
|  |  |  |  |  |
|  |  |  |  |  |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|  |  |  |  |  |
|  |  |  |  |  |
| LOG(PBI) | -0.468569 | 0.182697 | -2.564731 | 0.0235 |
| LOG(PEG) | 0.792065 | 0.097932 | 8.087922 | 0.0000 |
| LOG(PGT) | 0.148004 | 0.063681 | 2.324161 | 0.0370 |
| C | 8.227304 | 1.863566 | 4.414817 | 0.0007 |
|  |  |  |  |  |
|  |  |  |  |  |
| R-squared | 0.993889 | Mean dependent var | | 11.38052 |
| Adjusted R-squared | 0.992478 | S.D. dependent var | | 0.598849 |
| S.E. of regression | 0.051936 | Akaike info criterion | | -2.875279 |
| Sum squared resid | 0.035066 | Schwarz criterion | | -2.679228 |
| Log likelihood | 28.43987 | Hannan-Quinn criter. | | -2.855791 |
| F-statistic | 704.7439 | Durbin-Watson stat | | 2.503430 |
| Prob(F-statistic) | 0.000000 |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**Fuente:** **Elaboración propia**

En la figura 36 se muestra el grafico de la función error [16].



**Figura 36: Grafico del residuo de la ecuación de cointegración**

**Fuente:** **Elaboración Propia**

**Estimación de la ecuación de corrección de error o ecuación de corto plazo**

La ecuación de corrección de error se conseguirá colocando a todas las variables directas en diferenciación logarítmica acompañada de su rezago número uno, y con diferencias al error y al ATI20 y luego seleccionar a todas aquellas que sean representativas Se formula una ecuación base:

(3)

La ecuación de base estimada se muestra en la tabla 17.

**Tabla 17: Ecuación base**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dependent Variable: DLDEM | | |  |  |
| Method: Least Squares | | |  |  |
| Date: 01/28/23 Time: 20:24 | | |  |  |
| Sample (adjusted): 2007 2021 | | |  |  |
| Included observations: 15 after adjustments | | | |  |
|  |  |  |  |  |
|  |  |  |  |  |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|  |  |  |  |  |
|  |  |  |  |  |
| DLDEM(-1) | 0.482359 | 0.182457 | 2.643681 | 0.2302 |
| DLPBI | 0.340411 | 1.544587 | 0.220390 | 0.8619 |
| DLPBI(-1) | 1.719511 | 0.929931 | 1.849075 | 0.3156 |
| DLPEG | 0.690591 | 0.083589 | 8.261764 | 0.0767 |
| DLPEG(-1) | -0.489804 | 0.166050 | -2.949733 | 0.2081 |
| DLPET | -0.761171 | 1.063701 | -0.715588 | 0.6046 |
| DLPET(-1) | -0.759377 | 1.179263 | -0.643943 | 0.6358 |
| DLPGT | 0.129265 | 0.067736 | 1.908356 | 0.3073 |
| DLPGT(-1) | 0.056205 | 0.073258 | 0.767223 | 0.5834 |
| DLTAR | -0.479928 | 0.301154 | -1.593627 | 0.3568 |
| DLTAR(-1) | 0.317382 | 0.332007 | 0.955951 | 0.5143 |
| ERROR(-1) | -1.897446 | 0.520595 | -3.644766 | 0.1705 |
| DATI20 | -0.091806 | 0.100421 | -0.914210 | 0.5285 |
| C | -0.003495 | 0.025762 | -0.135655 | 0.9142 |
|  |  |  |  |  |
|  |  |  |  |  |
| R-squared | 0.999281 | Mean dependent var | | 0.114596 |
| Adjusted R-squared | 0.989940 | S.D. dependent var | | 0.206300 |
| S.E. of regression | 0.020692 | Akaike info criterion | | -5.759569 |
| Sum squared resid | 0.000428 | Schwarz criterion | | -5.098722 |
| Log likelihood | 57.19677 | Hannan-Quinn criter. | | -5.766608 |
| F-statistic | 106.9758 | Durbin-Watson stat | | 2.558505 |
| Prob(F-statistic) | 0.075548 |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**Fuente:** **Elaboración propia**

Se observa que todas las variables tienen un Prob mayores a 0,05 son no significativas, entonces todas no son regresores de la diferenciación logarítmica entonces se procede a eliminar aquellas que son menos representativas, las que están más lejanas de 0.05 y se va eliminando hasta que todas sean representativas y el Durbin Watson cerca de 2, se logró seleccionar todas aquellas que son representativas, cuya expresión es la siguiente.

Formulación de la ecuación de corrección de error encontrada:

(4)

La ecuación de corrección de error estimada se muestra en la tabla18, se observa los Prob. son menores a 0,05 son significativas, el valor de R2 es 96% y el Durbin-Watson es 1.93, está en el rango establecido, entonces esta ecuación de corrección de error encontrada nos servirá para el cálculo de la proyección.

**Tabla 18: Ecuación de corrección de error**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dependent Variable: DLDEM | | |  |  |
| Method: Least Squares | | |  |  |
| Date: 07/11/22 Time: 05:10 | | |  |  |
| Sample (adjusted): 2006 2021 | | |  |  |
| Included observations: 16 after adjustments | | | |  |
|  |  |  |  |  |
|  |  |  |  |  |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|  |  |  |  |  |
|  |  |  |  |  |
| DLPEG | 0.683898 | 0.069416 | 9.852117 | 0.0000 |
| DLPGT | 0.132095 | 0.056200 | 2.350456 | 0.0352 |
| ERROR(-1) | -1.218471 | 0.231409 | -5.265447 | 0.0002 |
|  |  |  |  |  |
|  |  |  |  |  |
| R-squared | 0.962489 | Mean dependent var | | 0.112155 |
| Adjusted R-squared | 0.956718 | S.D. dependent var | | 0.199544 |
| S.E. of regression | 0.041514 | Akaike info criterion | | -3.358219 |
| Sum squared resid | 0.022404 | Schwarz criterion | | -3.213359 |
| Log likelihood | 29.86576 | Hannan-Quinn criter. | | -3.350801 |
| Durbin-Watson stat | 1.996267 |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**Fuente:** **Elaboración propia**

**Ecuación de corto plazo de proyección**

Se construye la ecuación de proyección reemplazando en la variable error(*-1*) de la ecuación de corrección de error los valores de los coeficientes de la ecuación de cointegracion, como se muestra en la ecuación de proyección obtenida.

Se formula la ecuación de proyección:

(5)

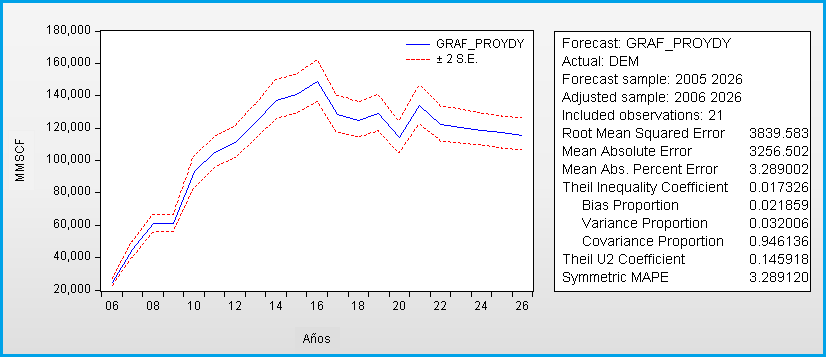
La ecuación de proyección estimada se muestra en la tabla 19, Tal como puede observarse, la ecuación de regresión tiene los estadísticos de prueba el Durbin-Watson, p-valores y el R2 valores significantes. Entonces el modelo estimado exhibe buenos resultados estadísticos tanto en bondad de ajuste como en bondad de predicción por lo que se realizara la proyección.

**Tabla 19: Ecuación de proyección estimada**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dependent Variable: D(LOG(DEM)) | | | |  |
| Method: Least Squares | | |  |  |
| Date: 07/11/22 Time: 05:39 | | |  |  |
| Sample (adjusted): 2006 2021 | | |  |  |
| Included observations: 16 after adjustments | | | |  |
|  |  |  |  |  |
|  |  |  |  |  |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|  |  |  |  |  |
|  |  |  |  |  |
| D(LOG(PEG)) | 0.683893 | 0.069418 | 9.851861 | 0.0000 |
| D(LOG(PGT)) | 0.132090 | 0.056200 | 2.350335 | 0.0352 |
| LOG(DEM(-1))-(-0.468569\*LOG(PBI(-1))+0.792065\*LOG(PEG(-1))+0.148004\*LOG(PGT(-1))+8.227304) | -1.218456 | 0.231411 | -5.265325 | 0.0002 |
|  |  |  |  |  |
|  |  |  |  |  |
| R-squared | 0.962488 | Mean dependent var | | 0.112155 |
| Adjusted R-squared | 0.956716 | S.D. dependent var | | 0.199544 |
| S.E. of regression | 0.041514 | Akaike info criterion | | -3.358188 |
| Sum squared resid | 0.022405 | Schwarz criterion | | -3.213328 |
| Log likelihood | 29.86550 | Hannan-Quinn criter. | | -3.350770 |
| Durbin-Watson stat | 1.996162 |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**Fuente:** **Elaboración propia**

### 4.1.9. Proyección dinámica de largo plazo de la demanda de gas para G.E.

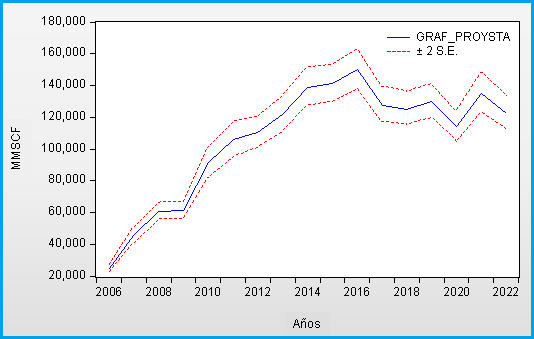
La proyección dinámica muestra la proyección histórica desde el 2005 al 2021 y la parte proyectada del 2022 al 2026, nuestro MAPE salió 3.28% lo que significa que el ajuste de calidad de proyección de nuestro modelo es bueno, como se muestra en la figura 37.

**Figura 37: Proyección dinámica de la demanda de gas natural seco para G.E.**

**Fuente:** **Elaboración propia**

**Proyección estática de la demanda periodo 2005-2021**

La proyección estática muestra la proyección histórica desde el 2005 al 2021, el MAPE salió 3.18% lo que significa que el ajuste de proyección de nuestro modelo es bueno, para hacer proyecciones con fines de planificación, como se muestra en la figura 38



**Figura 38: Proyección estática de la demanda**

**Fuente:** **Elaboración propia**

Seguidamente se muestran los valores numéricos obtenidos en la predicción desde enero 2005 a diciembre 2026, en la tabla 20, los cuales corresponden la figura 32 anteriormente mostrado.

**Tabla 20: Demanda histórico real y la predicción mediante el modelo Econométrico**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Años** | **PBI(MMS)** | **PGT(MMSCF)** | **TAR(US$/MMBTU)** | **DEM(MMSCF)** | **PEG(GWh)** | **PET(GWh)** |
| 2005 | 273971.15 | 28441 | 1.377 | 22099.13 | 2264.88 | 23001.48 |
| 2006 | 294597.83 | 37584 | 1.369 | 23833.09 | 2568.51 | 24762.78 |
| 2007 | 319693 | 69006 | 1.423 | 45859.97 | 5573.51 | 27254.93 |
| 2008 | 348870 | 93090 | 1.508 | 61317.61 | 7409.85 | 29558.71 |
| 2009 | 352693 | 99723 | 1.583 | 66236.43 | 7643.44 | 29807.25 |
| 2010 | 382081 | 244300 | 1.571 | 88223.71 | 11724.79 | 32426.83 |
| 2011 | 406256 | 388900 | 1.649 | 109446.34 | 12267 | 35217.43 |
| 2012 | 431199 | 407590 | 1.732 | 118129.8 | 13822.9 | 37321.18 |
| 2013 | 456435 | 421160 | 1.827 | 117205.91 | 16251.83 | 39669.43 |
| 2014 | 467308 | 442960 | 1.836 | 134721 | 19012.94 | 41795.89 |
| 2015 | 482506 | 424900 | 1.854 | 136147.25 | 19523.06 | 44540.04 |
| 2016 | 501581 | 477490 | 1.665 | 152233.48 | 21321.43 | 48326.42 |
| 2017 | 514215 | 441120 | 1.582 | 126228.46 | 17533.65 | 48993.25 |
| 2018 | 534665 | 432310 | 1.648 | 120719.51 | 17920.22 | 50816.79 |
| 2019 | 546161 | 455560 | 1.731 | 129550.04 | 18962.9 | 52889.14 |
| 2020 | 485992.07 | 409220 | 1.699 | 109562.62 | 16221.8 | 49186.64 |
| 2021 | 551829 | 385420 | 1.593 | 132955.15 | 19366.2 | 53990.35 |
| 2022 | 571694.84 | 375535 | 1.808 | 122029 | 18573.74 | 55925.16 |
| 2023 | 591704.16 | 365651.2 | 1.852 | 120543.2 | 18695.43 | 58008.65 |
| 2024 | 611230.4 | 355766.8 | 1.895 | 118677.5 | 18817.12 | 60092.14 |
| 2025 | 63078.77 | 345882.4 | 1.938 | 117034.6 | 18938.8 | 62175.64 |
| 2026 | 650344.26 | 335997.95 | 1.981 | 115415.1 | 19060.49 | 65197.61 |

,

**Fuente:** **Elaboracion propia**

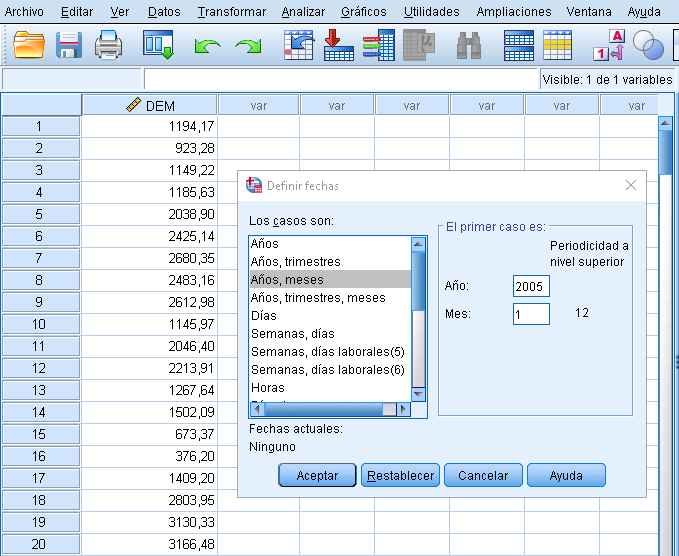
## 4.2. PROYECCION DEL MODELO UNIVARIANTE ARIMA EN RESOLUCION MENSUAL

La metodología de Box-Jeking nos dice que debemos seguir 5 pasos [11].

### 4.2.1. Paso 1: Prueba de estacionariedad

#### **Preparación del modelo**

Para el cálculo del modelo ARIMA univariante, se utilizó el Software SPSS-25, con los datos de la demanda histórica en resolución mensual en MMPC para el período de enero 2005 a diciembre 2021, es decir la serie se compone de 204 observaciones, siendo el horizonte de predicción entre enero 2021 a diciembre 2026.

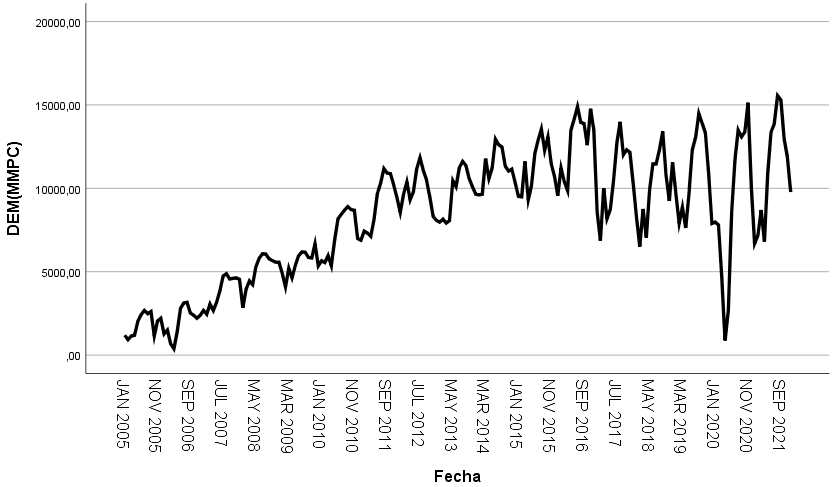
******Cargamos la data en el SPSS y configuramos el periodo de trabajo, En la Figura 39 podemos observar la periodicidad que será utilizada para nuestro estudio.

**Figura 39: Registro mensual de la variable demanda**

**Fuente:** **Elaboración propia**

#### **1. Visualización de la serie demanda de gas natural para generadores eléctricos**

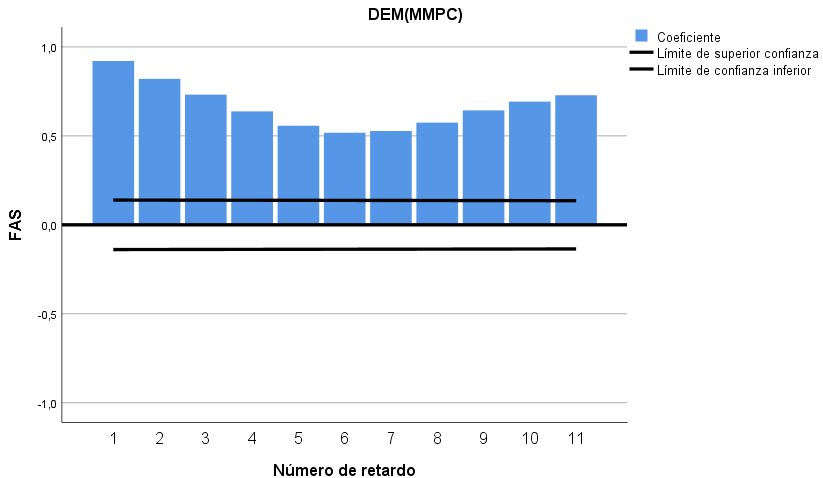
**Patrones de la serie demanda**, se observa en la Figura 40, que la serie presenta tendencia, variaciones estacionales y aleatoriedades, por lo que se concluye que la serie no es estacionaria.



**Figura 40: Muestra la evolución de la demanda de gas natural para G.E.**

**Fuente:** **Elaboración propia**

**Función de autocorrelación,** se observa, que la función de autocorrelación no corresponde a ningún tipo de patrón de los modelos estacionarios AR, MA o ARMA, entonces la serie es no estacionaria de origen, como se muestra en la Figura 41, entonces se tiene que aplicar sus transformaciones.

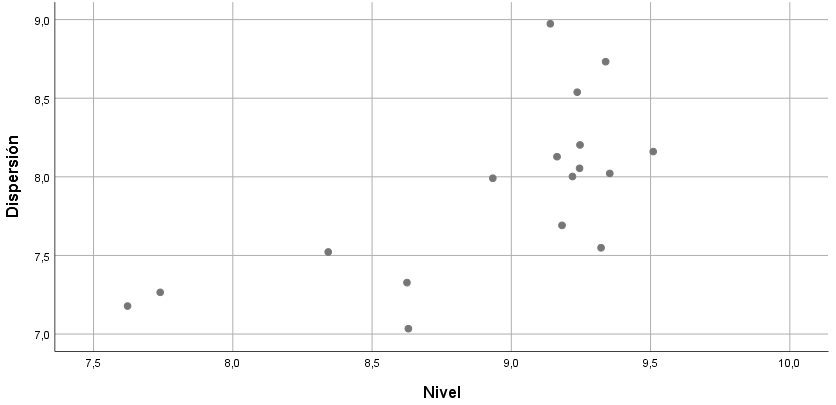


**Figura 41: funciones de autocorrelación (FAC)**

**Fuente:** **Elaboración propia**

**2. Análisis de estabilidad en varianza**

Para estabilizar en varianza bajo el modelo Box-Cox se realizó la prueba de Levene, como se muestra en la Figura 42, en el cual se agrupo los datos mensuales de la serie demanda en forma anual, obteniendo que la potencia de transformación de Levene es aproximadamente λ = 0.333.

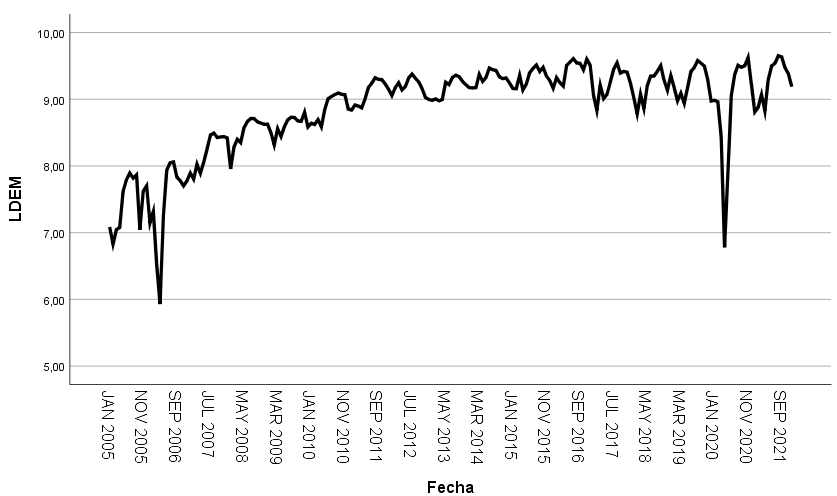


**Figura 42: Grafico de dispersión versus nivel de DEM por año (2005-2021).**

**Fuente:** **Elaboración propia**

De lo anterior, considerando que la potencia de transformación estimada mediante, la prueba de Levene fue 0.333 se seleccionó *λ*= 0, es decir la serie original requiere ser transformada para que tenga estabilidad en Varianza. Por lo tanto, se demuestra que no es estable en varianza. Para que la serie tenga estabilidad en Varianza debe ser transformada a , con la siguiente expresión.

En la figura 43 se muestra la serie transformada estabilizada en varianza.



**Figura 43: Logaritmo de la demanda versus el tiempo**

**Fuente:** **Elaboración propia**

**3. Análisis de estabilidad en media y estacional**

Con la aplicación de diferencias regulares y estacionales, es posible estabilizar la media de la serie.

Si la serie en estudio presenta tendencia creciente o decreciente, su FAC tendrá una estructura positiva con decrecimiento lento hacia cero, entonces esta tendencia puede estabilizarse aplicando sucesivas diferencias regulares:

Para estabilizar la media, se diferenció la parte regular (d=1) de la serie en estudio, en la Tabla 21 se muestra la serie estabilizada en media.

**Tabla 21. Serie estabilizada en media**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Números de caso de valores no perdidos** | |  | |
| **Nombre de la serie** | **Primero** | **Ultimo** | **N de casos validos** | **Creación de función** |
|  |  |
| LDEM\_dmin1 | 2 | 204 | 203 | DIFF (LDEM, 1) |

**Fuente:** **Elaboración propia**

**La estacionalidad**, se manifiesta como una pauta regular de comportamiento periódico en la serie. Entonces la estacionalidad se puede eliminar aplicando diferencias sucesivas estacionales D, de periodos:

Para estabilizar la estacionalidad, se diferenció la parte estacional (*D* = 1) de la serie en estudio, en la Tabla 22 se muestra la serie estabilizada en estacionalidad.

**Tabla 22. Serie estabilizada en estacionalidad**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Números de caso de valores no perdidos** | |  | |
| **Nombre de la serie** | **Primero** | **Ultimo** | **N de casos validos** | **Creación de función** |
|  |  |
| LDEM\_dmay1 | 2 | 204 | 203 | DIFF (LDEM, 1) |

**Fuente:** **Elaboración propia**

En la Tabla 23 se muestra la serie estabilizada en media y estacionalidad.

**Tabla 23. Serie estabilizada en media y estacionalidad**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Números de caso de valores no perdidos** | |  | |
| **Nombre de serie** | **Primero** | **Ultimo** | **N de casos validos** | **Creación de función** |
| LDEM\_dmin1dmay1 | 14 | 204 | 191 | SDIFF(LDEM\_dmin1,1, 12) |

**Fuente:** **Elaboración propia**

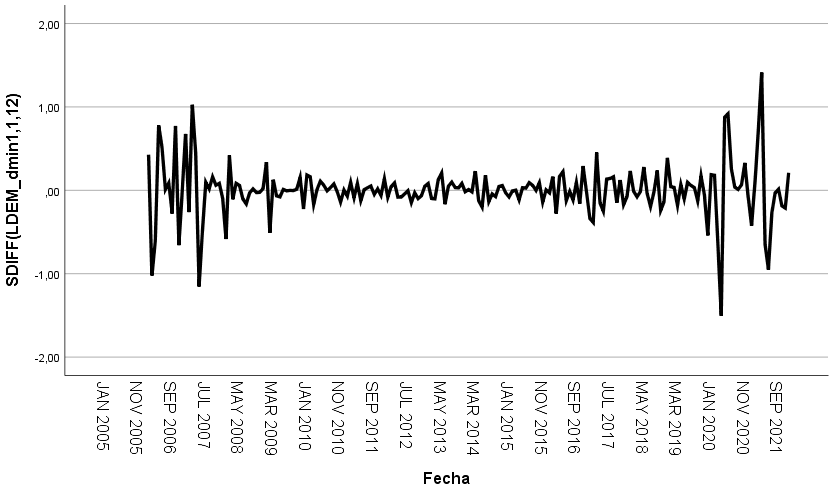
De las tres series diferenciadas (con nombres dmin1, dmay1 y dmin1dmay1), se obtiene su media y su desviación típica, La serie rdem-dmin1dmay1, tiene mayor estabilidad en media por tener el menor valor, por lo tanto, la serie es estacionaria, como se muestra en la Tabla 24.

**Tabla 24. Estadísticos descriptivos de la serie demanda**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **N** | **Media** | **Desviación** |
| **rdem-dmin1** | 203 | 0.0104 | 0.27067 |
| **rdem-dmay1** | 192 | 0.1138 | 0.39647 |
| **rdem-dmin1dmay1** | 191 | -0.005 | 0.31908 |

**Fuente:** **Elaboración propia**

En la Figura 44 se muestra el gráfico de frecuencia del estadístico descriptivo seleccionado.



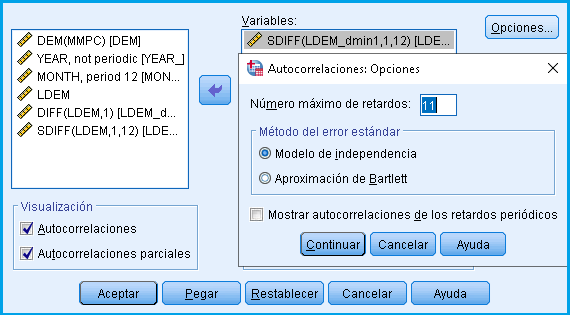
**Figura 44: Grafico de frecuencia del estadístico descriptivo rdem-dmin1dmay1 (d1D1)**

**Fuente:** **Elaboración propia**

**4.2.2. Paso 2: Identificación del modelo ARIMA**

**Determinación de las ordenes *p,q* de la parte regular y *P,Q* de la parte estacional.**

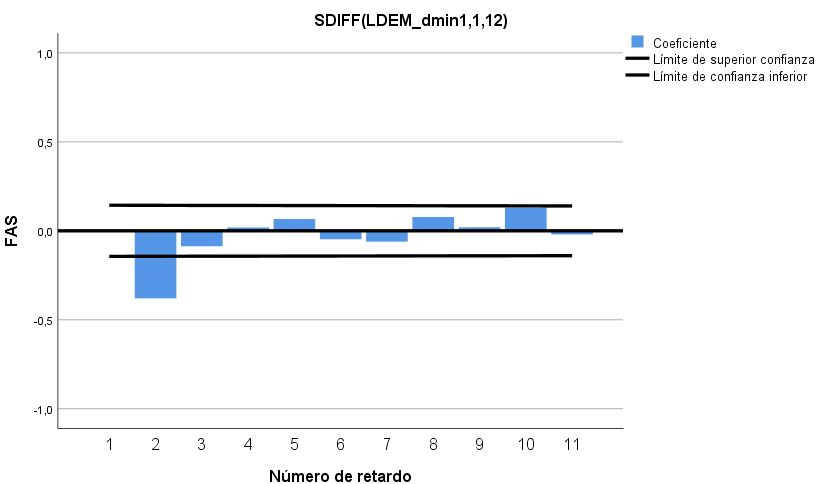
Para la parte regular el número máximo de retardos de la periodicidad es 12 menos 1, es decir 11 es el número máximo de retardos, en la figura 45 se muestra la configuración que se hace en el programa SPSS para la obtención de la autocorrelación de la orden regular.



**Figura 45: Configuración en el SPSS para obtener la orden regular**

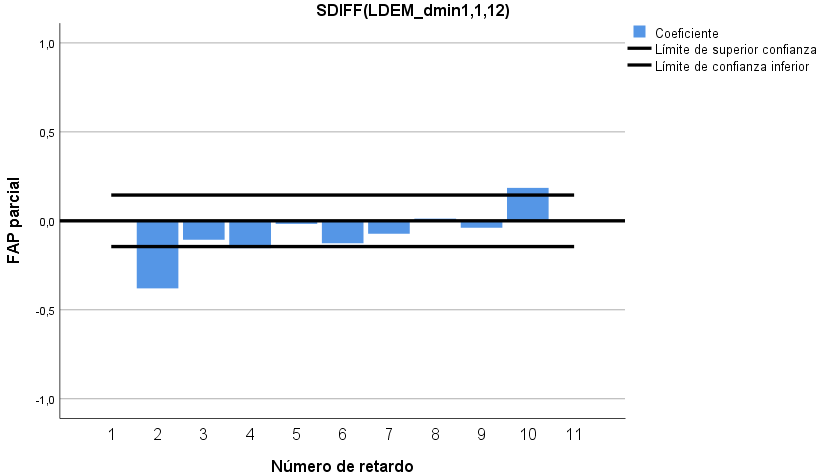
**Fuente: Elaboración propia**

En los gráficos ACF y ACFP se identificó una orden regular, existen escalones 1 y 9 que salen fuera del intervalo de confianza, por lo tanto, para la parte regular se considerara un modelo mixto ARIMA(1,1,1), cuya característica de la ACF y ACFP se muestra en las figuras 46 y 47.



**Figura 46: Función de autocorrelación regular de laserie en el periodo 2005-2021.**

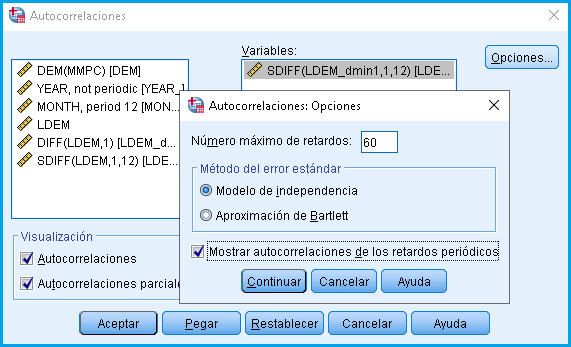
**Fuente:** **Elaboración propia**



**Figura 47: Función de autocorrelación parcial regular de la serie**

**Fuente:** **Elaboración propia**

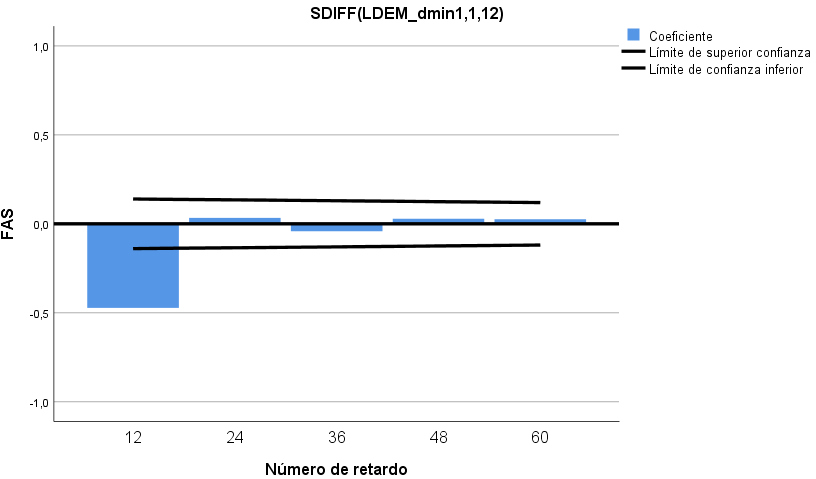
Para la parte estacional el número máximo de retardos es cinco veces el periodo es decir 60. En la figura 48 se muestra la configuración en el programa SPSS para la obtención de la orden estacional.



**Figura 48: Configuración en el SPSS para la orden estacional**

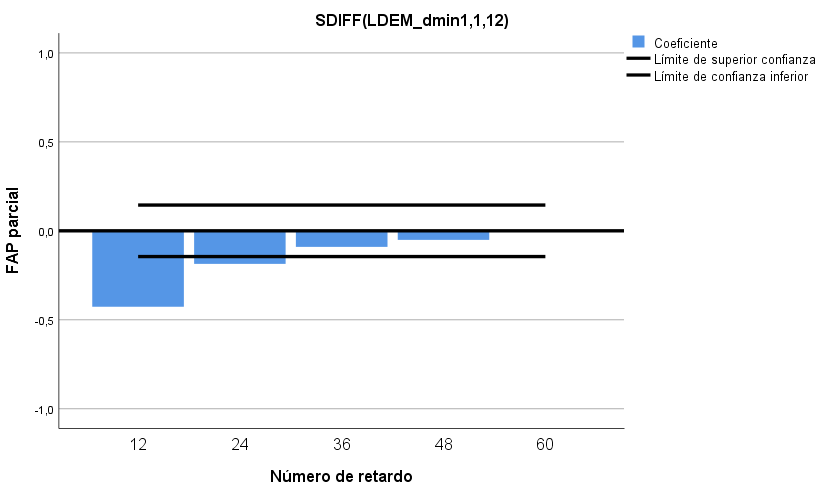
**Fuente: Elaboración propia**

En los gráficos ACF y ACFP se identificó una orden estacional de media móvil MA (1), *Q* = 1 entonces, *P*=0. Por tanto, la parte estacional es: ARIMA(*P,D,Q*) = ARIMA(0, 1 ,1), cuya característica de la ACF y ACFP se muestra en la figura 49 y 50.



**Figura 49: Función de autocorrelación estacional periodo 2005-2021.**

**Fuente: Elaboración propia**



**Figura 50: Función de autocorrelación parcial estacional de la serie**

**Fuente: Elaboración propia**

Finalmente, el modelo calculado presenta la siguiente estructura:

ARIMA (*p,d,q*)x(*P,D,Q*) = ARIMA (*1,1,1*)x(*0,1,1*).

En tal sentido, se analizarán dos modelos identificados que se muestran en la tabla 25, los cuales se analizaran desde el punto de vista de validación y se escogerá el modelo que de la mejor aproximación.

**Tabla 25: Modelos Identificados**

|  |  |  |
| --- | --- | --- |
| **N° Modelo** | **Variable** | **Modelo ARIMA** |
| **1** | Demanda de GN para GE(MMPC) | (1,1,1)x(0,1,1) |

**Fuente:** **Elaboración propia**

**4.2.3. Paso 3 Estimación del modelo identificado**

**Modelo 1 ARIMA**

La fórmula analítica del modelo ARIMA encontrado, se encuentra partiendo de la Ecuación General del Modelo:

Donde:

, Es el polinomio de retardo autorregresivo estacional.

, Es el polinomio de retardo autorregresivo regular.

, Es el polinomio de retardo media móvil estacional.

, Es el polinomio de retardo media móvil regular.

= función error

C = Constante

L = Operador retardo,

D = 1, d = 1 Diferenciación

S = 12, Periodo de ciclo

Dado que:(*p,d,q*)x(*P,D,Q*) = **(1,1,1)(0,1,1)**

P=0, reemplazando se encuentra que

p=1, reemplazando se encuentra que

Q=1, reemplazando se encuentra que

q=1, reemplazando se encuentra que

Reemplazando en la función general obtenemos la función analítica del modelo ARIMA calculado.

Considerando

Para la serie el cuadro de salida en SPSS del modelo ARIMA se muestra en la tabla 26:

**Tabla 26: Estimación de Parámetros incluyendo la constante**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Estimación | SE | t | Sig. |
| Constante |  | -,003 | ,001 | -4,320 | ,000 |
| AR | Retardo 1 | ,676 | ,076 | 8,890 | ,000 |
| Diferencia |  | 1 |  |  |  |
| MA | Retardo 1 | ,998 | ,518 | 1,926 | ,056 |
| Diferencia estacional |  | 1 |  |  |  |
| MA, estacional | Retardo 1 | ,811 | ,074 | 11,009 | ,000 |

**Fuente:** **Elaboración propia**

Donde, (p,d,q)x(P,D,Q) = **(1,1,1)(0,1,1)**

0.676, tiene representatividad

0.998, no tiene representatividad dado que la significancia obtenida es mayor del 0.05

0.811, tiene representatividad

C = -0.03, tiene representatividad

Recalculando con el SPSS sin considerar la constante y considerando los atípicos, obtenemos la tabla 27:

**Tabla 27: Resumen de los parámetros del modelo con valores atípicos sin constate**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Estimación | SE | t | Sig. |
| AR | Retardo 1 | ,709 | ,089 | 7,948 | ,000 |
| Diferencia |  | 1 |  |  |  |
| MA | Retardo 1 | ,926 | ,053 | 17,398 | ,000 |
| Diferencia estacional |  | 1 |  |  |  |
| MA, estacional | Retardo 1 | ,629 | ,072 | 8,726 | ,000 |

**Fuente:** **Elaboración propia**

Se obtienen los valores finales de los parámetros (p,d,q)x(P,D,Q) = **(1,1,1)(0,1,1)**

0.709, tiene representatividad

0.926, tiene representatividad

0.629, tiene representatividad

Reemplazando los valores de los parámetros, la ecuación final de la serie es la siguiente:

**Modelo ARIMA optimo**

El modelo ARIMA óptimo obtenido es: (1,1,1)(0,1,1) cuya ecuación final de la serie es la siguiente: Considerando

En la tabla 28 y 29 se muestra el mejor ajuste del modelo ARIMA y los valores atípicos obtenidos de a la serie.

**Tabla 28: Ajuste del modelo**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Estadístico de ajuste | Media | SE | Mínimo | Máximo | Percentil |  |  |  |  |  |  |
|  |  |  |  |  | 5 | 10 | 25 | 50 | 75 | 90 | 95 |
| R cuadrado estacionaria | ,845 | . | ,845 | ,845 | ,845 | ,845 | ,845 | ,845 | ,845 | ,845 | ,845 |
| R cuadrado | ,915 | . | ,915 | ,915 | ,915 | ,915 | ,915 | ,915 | ,915 | ,915 | ,915 |
| RMSE | 1,069,262 | . | 1,069,262 | 1,069,262 | 1,069,262 | 1,069,262 | 1,069,262 | 1,069,262 | 1,069,262 | 1,069,262 | 1,069,262 |
| MAPE | 9,464 | . | 9,464 | 9,464 | 9,464 | 9,464 | 9,464 | 9,464 | 9,464 | 9,464 | 9,464 |
| MaxAPE | 56,086 | . | 56,086 | 56,086 | 56,086 | 56,086 | 56,086 | 56,086 | 56,086 | 56,086 | 56,086 |
| MAE | 741,600 | . | 741,600 | 741,600 | 741,600 | 741,600 | 741,600 | 741,600 | 741,600 | 741,600 | 741,600 |
| MaxAE | 3,689,321 | . | 3,689,321 | 3,689,321 | 3,689,321 | 3,689,321 | 3,689,321 | 3,689,321 | 3,689,321 | 3,689,321 | 3,689,321 |
| BIC normalizado | 14,362 | . | 14,362 | 14,362 | 14,362 | 14,362 | 14,362 | 14,362 | 14,362 | 14,362 | 14,362 |

**Fuente:** **Elaboración propia**

**Tabla 29: Valores atípicos**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  | Estimación | SE | t | Sig. |
| Feb-06 | Aditivo |  | ,613 | ,094 | 6,549 | ,000 |
| Abr-06 | Aditivo |  | -,803 | ,095 | -8,453 | ,000 |
| Jun-06 | Transitorio | Magnitud | ,444 | ,117 | 3,784 | ,000 |
|  |  | Factor de decrecimiento | ,756 | ,174 | 4,332 | ,000 |
| Oct-06 | Aditivo estacional |  | ,691 | ,100 | 6,940 | ,000 |
| Ene-07 | Cambio de nivel |  | ,486 | ,114 | 4,258 | ,000 |
| Ene-08 | Aditivo |  | -,348 | ,092 | -3,800 | ,000 |
| Mar-20 | Aditivo |  | 2,790 | ,333 | 8,390 | ,000 |
|  | Transitorio | Magnitud | -3,491 | ,351 | -9,935 | ,000 |
|  |  | Factor de decrecimiento | ,619 | ,053 | 11,710 | ,000 |
| Jun-20 | Transitorio | Magnitud | ,464 | ,129 | 3,590 | ,000 |
|  |  | Factor de decrecimiento | ,775 | ,160 | 4,851 | ,000 |

**Fuente:** **Elaboración propia**

La figura 51, muestra la serie real en el periodo 2005-2021, de color azul y la serie que fue modelada en el mismo periodo, de color rojo, se observa que la aproximación de la serie predictiva es buena con respecto a la serie real, por lo tanto, el procedimiento seguido y valores que se obtienen son adecuados, dado el bajo porcentaje de error que arroja el modelo.

#### 

**Figura 51: Serie demanda real VS pronosticado periodo (2005-2021)**

**Fuente:** **Elaboración propia**

### 4.2.4. Paso 4: Evaluación

#### **1. Validación con horizonte predictivo**

Para la validación, fijamos un horizonte histórico parcial (HHP) menor que el horizonte histórico real (HHR), de tal forma que la diferencia entre el tamaño histórico parcial y el real debe ser del tamaño similar al horizonte predictivo, en este caso fijamos el horizonte parcial a enero 2019 y el predictivo a diciembre 2021, por tanto, el Horizonte de Validación (HV) es de 36 meses. Aplicamos el modelo ARIMA para el horizonte seleccionado y determinamos el pronóstico para el horizonte de validación. Con los resultados determinamos el error absoluto medio porcentual MAPE que está conformado por:

Donde;

TH: Tamaño del horizonte histórico y predictivo

En la tabla 14, se muestra los resultados y la comparación con el modelo determinista Winter Aditivo, resulta un MAPE de 9.464% para el modelo (1,1,1)(0,1,1), por lo tanto, este modelo encontrado es eficiente.

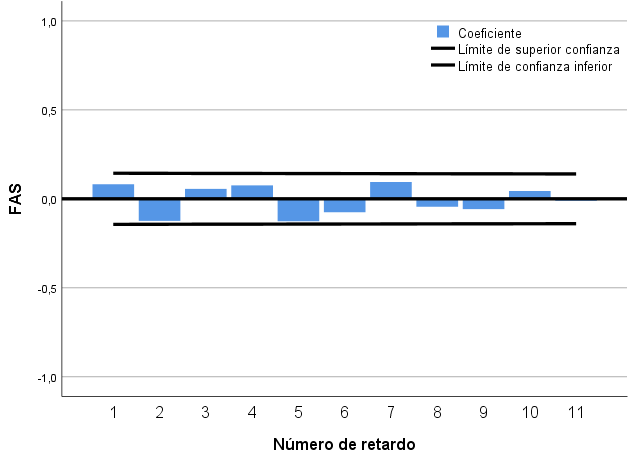
**Tabla 30: Estimación de parámetros**

|  |  |  |
| --- | --- | --- |
| **Modelo** | **Horizonte** | **MAPE (%)** |
| ARIMA | Ene.05 – Dic. 21 | 9.464 |
| Winter Aditivo | Ene.05 – Dic. 21 | 15.445 |

**Fuente: Elaboración propia**

**2. Validación del modelo ARIMA óptimo mediante la prueba de residuos.**

Mediante el software SPSS se aplica autocorrelaciones ACF al error del modelo ARIMA optimo (1,1,1)(0,1,1), se muestra que los residuos están incorrelados es decir caen dentro de la banda de confianza, como se muestran en la figura 52.



**Figura 52: Residuo de ruido de la demanda para modelo ARIMA optimo**

**Fuente: Elaboración propia**

Así mismo en la tabla 31, el estadístico de Box – Ljung muestra que todos los valores son significativos porque son mayores de 0.05, en tal sentido se comprueba que este modelo es el óptimo.

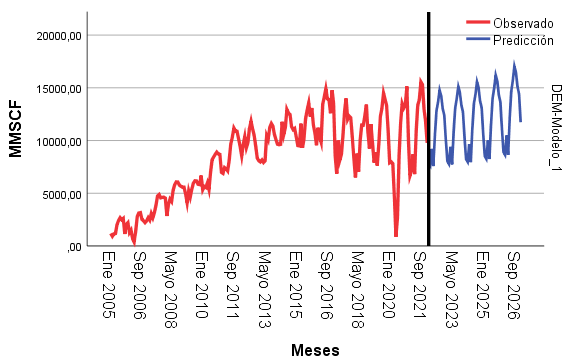
**Tabla 31: Autocorrelaciones del Residuo de ruido de la demanda del modelo optimo**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Retardo | Autocorrelación | Desv. Error |  | Estadístico de Box-Ljung |  |
|  |  |  | Valor | gl | Sig.b |
| 1 | ,082 | ,072 | 1,296 | 1 | ,255 |
| 2 | -,124 | ,072 | 4,279 | 2 | ,118 |
| 3 | ,056 | ,071 | 4,891 | 3 | ,180 |
| 4 | ,075 | ,071 | 6,008 | 4 | ,199 |
| 5 | -,126 | ,071 | 9,145 | 5 | ,103 |
| 6 | -,076 | ,071 | 10,284 | 6 | ,113 |
| 7 | ,095 | ,071 | 12,079 | 7 | ,098 |
| 8 | -,045 | ,070 | 12,479 | 8 | ,131 |
| 9 | -,058 | ,070 | 13,158 | 9 | ,156 |
| 10 | ,044 | ,070 | 13,553 | 10 | ,194 |
| 11 | -,011 | ,070 | 13,577 | 11 | ,257 |

**Fuente:** **Elaboración propia**

**4.2.5.Paso 5:** **Proyección de la demanda de gas natural para generadores eléctricos**

Finalmente se procedió a proyectar la serie que fue estudiada, la Figura 53, muestra la serie modelada en el periodo 2005-2021(color rojo) y la serie proyectada en el periodo 2022-2026(color azul), que fue el objetivo.

**Figura 53: Serie proyectada de la demanda energética para el periodo 2022-2026**

**Fuente:** **Elaboración propia**

Seguidamente se muestra en la Tabla 32, los valores numéricos obtenidos por el modelo en la predicción desde enero 2005 a diciembre 2026, los cuales corresponden la Figura 53 anteriormente mostrado.

**Tabla 32: Consumo histórico real y la predicción mediante el modelo ARIMA**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Consumo Histórico (MMCSF) | Año | Mes | Fecha | Proyección (MMSCF) |
| 1194.17 | 2005 | 1 | Ene-05 |  |
| 923.28 | 2005 | 2 | Feb-05 |  |
| 1149.22 | 2005 | 3 | Mar-05 |  |
| 1185.63 | 2005 | 4 | Abr-05 |  |
| 2038.9 | 2005 | 5 | May-05 |  |
| 2425.14 | 2005 | 6 | Jun-05 |  |
| 2680.35 | 2005 | 7 | Jul-05 |  |
| 2483.16 | 2005 | 8 | Ago-05 |  |
| 2612.98 | 2005 | 9 | Set-05 |  |
| 1145.97 | 2005 | 10 | Oct-05 |  |
| 2046.4 | 2005 | 11 | Nov-05 |  |
| 2213.91 | 2005 | 12 | Dic-05 |  |
| 1267.64 | 2006 | 1 | Ene-06 |  |
| 1502.09 | 2006 | 2 | Feb-06 | 1830.96 |
| 673.37 | 2006 | 3 | Mar-06 | 1051.03 |
| 376.2 | 2006 | 4 | Abr-06 | 341.8 |
| 1409.2 | 2006 | 5 | May-06 | 1520.06 |
| 2803.95 | 2006 | 6 | Jun-06 | 2744.7 |
| 3130.33 | 2006 | 7 | Jul-06 | 2866.6 |
| 3166.48 | 2006 | 8 | Ago-06 | 2684.26 |
| 2519.56 | 2006 | 9 | Set-06 | 3044.4 |
| 2390.66 | 2006 | 10 | Oct-06 | 2127.79 |
| 2211.45 | 2006 | 11 | Nov-06 | 2026.54 |
| 2382.13 | 2006 | 12 | Dic-06 | 2246.4 |
| 2682.37 | 2007 | 1 | Ene-07 | 2073.63 |
| 2448.66 | 2007 | 2 | Feb-07 | 1753.09 |
| 3062.22 | 2007 | 3 | Mar-07 | 2186.6 |
| 2675.71 | 2007 | 4 | Abr-07 | 3049.85 |
| 3164.28 | 2007 | 5 | May-07 | 4220.14 |
| 3856.45 | 2007 | 6 | Jun-07 | 3931.31 |
| 4745.87 | 2007 | 7 | Jul-07 | 4583.47 |
| 4881.42 | 2007 | 8 | Ago-07 | 4820.62 |
| 4565.09 | 2007 | 9 | Set-07 | 4543.8 |
| 4604.02 | 2007 | 10 | Oct-07 | 4273.02 |
| 4632.3 | 2007 | 11 | Nov-07 | 4205.23 |
| 4541.54 | 2007 | 12 | Dic-07 | 4915.96 |
| 2851.22 | 2008 | 1 | Ene-08 | 2047.29 |
| 3958.69 | 2008 | 2 | Feb-08 | 2987.26 |
| 4447.72 | 2008 | 3 | Mar-08 | 3945.8 |
| 4225.74 | 2008 | 4 | Abr-08 | 4128.3 |
| 5298.93 | 2008 | 5 | May-08 | 5718.68 |
| 5812.57 | 2008 | 6 | Jun-08 | 6302.69 |
| 6076.11 | 2008 | 7 | Jul-08 | 6946.66 |
| 6065.59 | 2008 | 8 | Ago-08 | 6367.25 |
| 5771.37 | 2008 | 9 | Set-08 | 5832.52 |
| 5667.56 | 2008 | 10 | Oct-08 | 5721.53 |
| 5567.74 | 2008 | 11 | Nov-08 | 5557.6 |
| 5574.33 | 2008 | 12 | Dic-08 | 5894.58 |
| 4905.46 | 2009 | 1 | Ene-09 | 4219.68 |
| 4089.02 | 2009 | 2 | Feb-09 | 4156.89 |
| 5215.84 | 2009 | 3 | Mar-09 | 4516.92 |
| 4636.98 | 2009 | 4 | Abr-09 | 5004.39 |
| 5375.06 | 2009 | 5 | May-09 | 6313.66 |
| 5960.1 | 2009 | 6 | Jun-09 | 6512.89 |
| 6187.08 | 2009 | 7 | Jul-09 | 7064.66 |
| 6170.46 | 2009 | 8 | Ago-09 | 6641.24 |
| 5852.65 | 2009 | 9 | Set-09 | 6151.78 |
| 5823.57 | 2009 | 10 | Oct-09 | 6004.18 |
| 6675.1 | 2009 | 11 | Nov-09 | 5902.14 |
| 5345.06 | 2009 | 12 | Dic-09 | 6890.4 |
| 5655.53 | 2010 | 1 | Ene-10 | 4490.81 |
| 5548.53 | 2010 | 2 | Feb-10 | 4773.74 |
| 5971.35 | 2010 | 3 | Mar-10 | 6267.01 |
| 5359.71 | 2010 | 4 | Abr-10 | 5601.12 |
| 6924.62 | 2010 | 5 | May-10 | 6893.06 |
| 8176.06 | 2010 | 6 | Jun-10 | 7959.26 |
| 8434.43 | 2010 | 7 | Jul-10 | 8935.29 |
| 8690.09 | 2010 | 8 | Ago-10 | 8571.91 |
| 8903.9 | 2010 | 9 | Set-10 | 8238.43 |
| 8726.05 | 2010 | 10 | Oct-10 | 8628.68 |
| 8684.08 | 2010 | 11 | Nov-10 | 8919.05 |
| 6986.62 | 2010 | 12 | Dic-10 | 8099.43 |
| 6887.63 | 2011 | 1 | Ene-11 | 6376.71 |
| 7434.28 | 2011 | 2 | Feb-11 | 6216.06 |
| 7320.34 | 2011 | 3 | Mar-11 | 8157.9 |
| 7120.29 | 2011 | 4 | Abr-11 | 6818.47 |
| 8127.01 | 2011 | 5 | May-11 | 9096.91 |
| 9675.95 | 2011 | 6 | Jun-11 | 9619.74 |
| 10299.45 | 2011 | 7 | Jul-11 | 10517.99 |
| 11181.25 | 2011 | 8 | Ago-11 | 10611.14 |
| 10909.19 | 2011 | 9 | Set-11 | 10878.1 |
| 10882.22 | 2011 | 10 | Oct-11 | 10699.65 |
| 10189.46 | 2011 | 11 | Nov-11 | 11051.98 |
| 9419.17 | 2011 | 12 | Dic-11 | 9143.56 |
| 8554.32 | 2012 | 1 | Ene-12 | 8723.73 |
| 9645.99 | 2012 | 2 | Feb-12 | 8291.44 |
| 10381.23 | 2012 | 3 | Mar-12 | 10116.11 |
| 9304.78 | 2012 | 4 | Abr-12 | 9597.8 |
| 9796.25 | 2012 | 5 | May-12 | 11354.65 |
| 11167.36 | 2012 | 6 | Jun-12 | 11786.31 |
| 11831.96 | 2012 | 7 | Jul-12 | 12321.46 |
| 11076.76 | 2012 | 8 | Ago-12 | 12709.71 |
| 10494 | 2012 | 9 | Set-12 | 11373.99 |
| 9480.18 | 2012 | 10 | Oct-12 | 10949.12 |
| 8315.66 | 2012 | 11 | Nov-12 | 10049.48 |
| 8081.21 | 2012 | 12 | Dic-12 | 8190.04 |
| 7967.97 | 2013 | 1 | Ene-13 | 7931.15 |
| 8152.44 | 2013 | 2 | Feb-13 | 8527.03 |
| 7914.04 | 2013 | 3 | Mar-13 | 9174.19 |
| 8068.38 | 2013 | 4 | Abr-13 | 7826.94 |
| 10459.73 | 2013 | 5 | May-13 | 9752.65 |
| 10096.29 | 2013 | 6 | Jun-13 | 12298.75 |
| 11242.09 | 2013 | 7 | Jul-13 | 11347.98 |
| 11617.29 | 2013 | 8 | Ago-13 | 11712.57 |
| 11375.73 | 2013 | 9 | Set-13 | 11552.01 |
| 10580.12 | 2013 | 10 | Oct-13 | 11149.98 |
| 10088.55 | 2013 | 11 | Nov-13 | 10304.05 |
| 9643.19 | 2013 | 12 | Dic-13 | 9583.14 |
| 9603.59 | 2014 | 1 | Ene-14 | 9245.59 |
| 9644.41 | 2014 | 2 | Feb-14 | 9856.38 |
| 11776.85 | 2014 | 3 | Mar-14 | 10080.42 |
| 10586.41 | 2014 | 4 | Abr-14 | 11014.4 |
| 11215.43 | 2014 | 5 | May-14 | 12686.19 |
| 12948.29 | 2014 | 6 | Jun-14 | 12403.34 |
| 12603.66 | 2014 | 7 | Jul-14 | 14108.96 |
| 12478.91 | 2014 | 8 | Ago-14 | 13142.47 |
| 11329.97 | 2014 | 9 | Set-14 | 12500.92 |
| 11037.69 | 2014 | 10 | Oct-14 | 11208.59 |
| 11151.34 | 2014 | 11 | Nov-14 | 10846.04 |
| 10344.34 | 2014 | 12 | Dic-14 | 10651.27 |
| 9519.43 | 2015 | 1 | Ene-15 | 10156.51 |
| 9489.75 | 2015 | 2 | Feb-15 | 9949.92 |
| 11605.15 | 2015 | 3 | Mar-15 | 10786.73 |
| 9312.85 | 2015 | 4 | Abr-15 | 10956.51 |
| 10180.27 | 2015 | 5 | May-15 | 11176.23 |
| 12071.18 | 2015 | 6 | Jun-15 | 11871.84 |
| 12880.83 | 2015 | 7 | Jul-15 | 12981.97 |
| 13546.85 | 2015 | 8 | Ago-15 | 13267 |
| 12279.45 | 2015 | 9 | Set-15 | 12999.8 |
| 13083.95 | 2015 | 10 | Oct-15 | 12013.42 |
| 11471.85 | 2015 | 11 | Nov-15 | 12678.68 |
| 10705.69 | 2015 | 12 | Dic-15 | 10931.96 |
| 9555.63 | 2016 | 1 | Ene-16 | 10331.75 |
| 11215.73 | 2016 | 2 | Feb-16 | 9943.72 |
| 10388.42 | 2016 | 3 | Mar-16 | 12800.82 |
| 9860.78 | 2016 | 4 | Abr-16 | 9602.79 |
| 13482.78 | 2016 | 5 | May-16 | 11431.59 |
| 14142.97 | 2016 | 6 | Jun-16 | 15074.66 |
| 14896.81 | 2016 | 7 | Jul-16 | 14897.09 |
| 13951.28 | 2016 | 8 | Ago-16 | 15255 |
| 13888.42 | 2016 | 9 | Set-16 | 13277.5 |
| 12592.71 | 2016 | 10 | Oct-16 | 13910.15 |
| 14767.72 | 2016 | 11 | Nov-16 | 12087.73 |
| 13490.23 | 2016 | 12 | Dic-16 | 13452.8 |
| 8588.26 | 2017 | 1 | Ene-17 | 12277.58 |
| 6861.35 | 2017 | 2 | Feb-17 | 9785.92 |
| 10000.67 | 2017 | 3 | Mar-17 | 8223.35 |
| 8175.71 | 2017 | 4 | Abr-17 | 9512.91 |
| 8708.42 | 2017 | 5 | May-17 | 10593.46 |
| 10460.03 | 2017 | 6 | Jun-17 | 10516.49 |
| 12701.12 | 2017 | 7 | Jul-17 | 11746.57 |
| 13986.34 | 2017 | 8 | Ago-17 | 12948.38 |
| 12003.92 | 2017 | 9 | Set-17 | 13421.71 |
| 12308.29 | 2017 | 10 | Oct-17 | 11875.99 |
| 12155.13 | 2017 | 11 | Nov-17 | 12679.32 |
| 10279.22 | 2017 | 12 | Dic-17 | 11468.95 |
| 8250.87 | 2018 | 1 | Ene-18 | 8601.16 |
| 6504.25 | 2018 | 2 | Feb-18 | 8199.1 |
| 8760.15 | 2018 | 3 | Mar-18 | 8368.58 |
| 7050.43 | 2018 | 4 | Abr-18 | 8003.38 |
| 9927.53 | 2018 | 5 | May-18 | 8651.65 |
| 11465.99 | 2018 | 6 | Jun-18 | 11456.42 |
| 11459.17 | 2018 | 7 | Jul-18 | 12844.3 |
| 12320.26 | 2018 | 8 | Ago-18 | 12182.73 |
| 13405.57 | 2018 | 9 | Set-18 | 11546.56 |
| 10779.94 | 2018 | 10 | Oct-18 | 12979.83 |
| 9250.01 | 2018 | 11 | Nov-18 | 11161.18 |
| 11545.33 | 2018 | 12 | Dic-18 | 8809.08 |
| 9676.13 | 2019 | 1 | Ene-19 | 9157.41 |
| 7898.87 | 2019 | 2 | Feb-19 | 8440.78 |
| 8871.8 | 2019 | 3 | Mar-19 | 9847.94 |
| 7628.36 | 2019 | 4 | Abr-19 | 7692.72 |
| 9684.43 | 2019 | 5 | May-19 | 9665.87 |
| 12303.3 | 2019 | 6 | Jun-19 | 11236.94 |
| 13034.16 | 2019 | 7 | Jul-19 | 13005.27 |
| 14487.09 | 2019 | 8 | Ago-19 | 13544.81 |
| 13908.37 | 2019 | 9 | Set-19 | 13899.15 |
| 13321.07 | 2019 | 10 | Oct-19 | 12542.63 |
| 10846.48 | 2019 | 11 | Nov-19 | 12362.66 |
| 7889.98 | 2019 | 12 | Dic-19 | 11135.6 |
| 7975.07 | 2020 | 1 | Ene-20 | 7017.21 |
| 7802.27 | 2020 | 2 | Feb-20 | 7111.73 |
| 4610.81 | 2020 | 3 | Mar-20 | 4663.46 |
| 880.73 | 2020 | 4 | Abr-20 | 916.95 |
| 2684.12 | 2020 | 5 | May-20 | 2547.75 |
| 8533.54 | 2020 | 6 | Jun-20 | 8445.42 |
| 11679.27 | 2020 | 7 | Jul-20 | 11191.88 |
| 13490.03 | 2020 | 8 | Ago-20 | 13867.39 |
| 13086.89 | 2020 | 9 | Set-20 | 13874.49 |
| 13372.44 | 2020 | 10 | Oct-20 | 12650.18 |
| 15131.33 | 2020 | 11 | Nov-20 | 12058.85 |
| 10106.86 | 2020 | 12 | Dic-20 | 12880.8 |
| 6691.27 | 2021 | 1 | Ene-21 | 9048.18 |
| 7165.27 | 2021 | 2 | Feb-21 | 6499.65 |
| 8693.02 | 2021 | 3 | Mar-21 | 8817.13 |
| 6817.07 | 2021 | 4 | Abr-21 | 7475.73 |
| 10871.68 | 2021 | 5 | May-21 | 9066.88 |
| 13371.77 | 2021 | 6 | Jun-21 | 12791.76 |
| 13872.49 | 2021 | 7 | Jul-21 | 14159.41 |
| 15557.88 | 2021 | 8 | Ago-21 | 14495.43 |
| 15300.4 | 2021 | 9 | Set-21 | 14459.48 |
| 12969.65 | 2021 | 10 | Oct-21 | 14194.48 |
| 11864.66 | 2021 | 11 | Nov-21 | 12639.8 |
| 9779.56 | 2021 | 12 | Dic-21 | 9673.46 |
|  | 2022 | 1 | Ene-22 | 7846.61 |
|  | 2022 | 2 | Feb-22 | 7610.26 |
|  | 2022 | 3 | Mar-22 | 9205.11 |
|  | 2022 | 4 | Abr-22 | 7578.53 |
|  | 2022 | 5 | May-22 | 10584.01 |
|  | 2022 | 6 | Jun-22 | 12814.37 |
|  | 2022 | 7 | Jul-22 | 13669.36 |
|  | 2022 | 8 | Ago-22 | 14780.21 |
|  | 2022 | 9 | Set-22 | 14245.12 |
|  | 2022 | 10 | Oct-22 | 13020.45 |
|  | 2022 | 11 | Nov-22 | 12404.97 |
|  | 2022 | 12 | Dic-22 | 10072.51 |
|  | 2023 | 1 | Ene-23 | 8043.77 |
|  | 2023 | 2 | Feb-23 | 7783.19 |
|  | 2023 | 3 | Mar-23 | 9404.04 |
|  | 2023 | 4 | Abr-23 | 7739.21 |
|  | 2023 | 5 | May-23 | 10808.2 |
|  | 2023 | 6 | Jun-23 | 13088.33 |
|  | 2023 | 7 | Jul-23 | 13965.97 |
|  | 2023 | 8 | Ago-23 | 15106.7 |
|  | 2023 | 9 | Set-23 | 14565.92 |
|  | 2023 | 10 | Oct-23 | 13319.57 |
|  | 2023 | 11 | Nov-23 | 12695.74 |
|  | 2023 | 12 | Dic-23 | 10313.35 |
|  | 2024 | 1 | Ene-24 | 8244.17 |
|  | 2024 | 2 | Feb-24 | 7984.1 |
|  | 2024 | 3 | Mar-24 | 9654.53 |
|  | 2024 | 4 | Abr-24 | 7951.29 |
|  | 2024 | 5 | May-24 | 11112.2 |
|  | 2024 | 6 | Jun-24 | 13465.55 |
|  | 2024 | 7 | Jul-24 | 14377.88 |
|  | 2024 | 8 | Ago-24 | 15562.18 |
|  | 2024 | 9 | Set-24 | 15014.5 |
|  | 2024 | 10 | Oct-24 | 13738.26 |
|  | 2024 | 11 | Nov-24 | 13102.84 |
|  | 2024 | 12 | Dic-24 | 10650.54 |
|  | 2025 | 1 | Ene-25 | 8523.61 |
|  | 2025 | 2 | Feb-25 | 8263.31 |
|  | 2025 | 3 | Mar-25 | 10001.68 |
|  | 2025 | 4 | Abr-25 | 8244.52 |
|  | 2025 | 5 | May-25 | 11531.77 |
|  | 2025 | 6 | Jun-25 | 13985.37 |
|  | 2025 | 7 | Jul-25 | 14944.78 |
|  | 2025 | 8 | Ago-25 | 16188.36 |
|  | 2025 | 9 | Set-25 | 15630.61 |
|  | 2025 | 10 | Oct-25 | 14312.86 |
|  | 2025 | 11 | Nov-25 | 13661.14 |
|  | 2025 | 12 | Dic-25 | 11112.68 |
|  | 2026 | 1 | Ene-26 | 8905.06 |
|  | 2026 | 2 | Feb-26 | 8643.31 |
|  | 2026 | 3 | Mar-26 | 10473.08 |
|  | 2026 | 4 | Abr-26 | 8642.02 |
|  | 2026 | 5 | May-26 | 12099.71 |
|  | 2026 | 6 | Jun-26 | 14688.2 |
|  | 2026 | 7 | Jul-26 | 15710.51 |
|  | 2026 | 8 | Ago-26 | 17033.47 |
|  | 2026 | 9 | Set-26 | 16461.56 |
|  | 2026 | 10 | Oct-26 | 15087.33 |
|  | 2026 | 11 | Nov-26 | 14413.25 |
|  | 2026 | 12 | Dic-26 | 11734.93 |

Fuente: Elaboración propia

**4.2.6. Comparación de los modelos Ecométrico y ARIMA**

En la tabla 32 se muestra los resultados obtenidos con el modelo Econométrico y el modelo ARIMA para el periodo enero 2005 a diciembre 2025. El parámetro utilizado para la evaluación es el indicador error absoluto porcentual medio (MAPE), para lo cual se han empleado los valores históricos de enero 2005 a junio 2021, Con el método ARIMA se ha obtenido el modelo (1,1,1)(0,1,1) cuyos resultados obtenidos tienen un error absoluto porcentual medio (MAPE) de 9.46%, frente al 3.42% del modelo Econométrico.

**Tabla 32: Comparación del MAPE**

|  |  |  |
| --- | --- | --- |
| **Modelo** | **Horizonte** | **MAPE** |
| **ARIMA en resolución mensual**  **Econométrico en resolución anual** | Enero 2005-Diciembre 2026  Enero 2005-Diciembre 2026 | 9.46%  3.42% |

**Fuente:** **Elaboración propia**

# CAPITULO V: CONCLUSIONES Y RECOMENDACIONES

Luego de haber realizado todo el proceso de copilación, validación, procesamiento, análisis de las corridas con los datos. Se ha llegado a las siguientes:

## 5.1 CONCLUSIONES

### 5.1.1. Modelo econométrico

1. Bajo el plano estadístico, el modelo de regresión lineal por mínimos cuadrados ordinarios que describe y proyecta a la variable dependiente, cumplió con todos los condicionantes estadísticos.

2. En la estimación de la ecuación de prueba en diferencias logarítmicas, el Durbin Watson sale fuera del rango de 2, lo cual significa que los residuos de esta regresión tienen un componente de autocorrelación de primer orden, es decir no son un ruido blanco, lo que significa que esta ecuación de prueba no es consistente.

3. De acuerdo al análisis de cointegracion, se formuló un modelo econométrico de naturaleza biecuacional con dos componentes una de largo plazo y otra de corto plazo con variables que sean no necesariamente de origen individual estacionaria.

4. En la estimación de la ecuación de corrección de error o ecuación de corto plazo de proyección, se observa los pvalores menores a 0,05, el valor de R2 es 96% y el Durbin-Watson es 1.99, está en el rango establecido, Entonces el modelo estimado exhibe buenos resultados estadísticos tanto en bondad de ajuste como en predicción.

5. En la estimación de la proyección dinámica en el periodo 2020-2026, el indicador estadístico más utilizado que mide la calidad de la proyección es el error absoluto porcentual medio (MAPE), y se observa que el MAPE es 3.28, es un buen indicador para proyecciones de mediano largo plazo y que la metodología es eficiente para hacer proyecciones con fines de planificación.

### 5.1.2. Modelo ARIMA

1. En la visualización de la serie demanda y, según la función de autocorrelación, los patrones de la serie presentan tendencia, aleatoriedades y variaciones estacionales mensuales, y la función de autocorrelación no corresponde a ningún tipo de patrón de los modelos estacionarios AR, MA o ARMA, entonces se concluye que la serie no es estacionaria.

2. En las transformaciones necesarias de la serie, para lograr estabilidad en Media Regular y Media Estacional, el estadístico descriptivo con menor media obtenida es de -0.005 correspondiente al aplicar una diferencia estacional combinada a la serie demanda, por lo tanto, para el modelamiento ARIMA se tomaron valores de orden d=1 y D=1, entonces esta serie es estable en varianza y media, es decir la serie original transformada ya es estacionaria.

3. En la Identificación Orden regular AR(p) MA(q) o ARMA(p,q), se obtuvieron para el cálculo el modelo mixto ARIMA(1,1,1). En la Identificación Orden estacional AR(P) MA(Q)o ARMA(P,Q), se obtuvo, el modelo ARIMA(P,D,Q) = ARIMA(0, 1 ,1).

4. En la validación de modelo ARIMA óptimo mediante la prueba de residuos, los residuos no están correlados es decir caen dentro de la banda de confianza, así también el estadístico de Box – Ljung muestra que todos los valores pasan la prueba de significancia, en tal sentido se comprueba que este modelo es el óptimo.

5. En la comparación de modelos, cuando hablamos del MAPE tanto en la OLADE (Organización Latinoamericana de Energía) o en la AIE (agencia internacional de energía), cuando se ve planificación de energía de oferta y demanda en energía eléctrica allí el MAPE es menor del 5% cuando la serie está en resolución anual y menor del 3% cuando la serie está en resolución mensual, en nuestro caso estamos hablando de la variable demanda de gas natural, por lo tanto no podemos también aplicarle la misma condición del MAPE, si al final lo que se quiere es buscar un caudal medio para garantizar un determinado despacho, por lo tanto el modelo ARIMA con este MAPE que da y además comparado con el modelo Winter que es bueno bajo el enfoque determinístico, se ve que es mejor por lo tanto la resolución de este modelo aplicado a esta variable demanda de gas es bueno.

Con el método ARIMA se ha obtenido el modelo Xtln(1,1,1)(0,1,1) cuyos resultados obtenidos tienen un MAPE de 9.46%, frente al 3.28% del modelo Econométrico. Con esto se demuestra que el método Econométrico es la que más se ajusta en la predicción de series temporales con estructura energética como lo es la demanda de gas para generadores eléctricos

## 5.2. RECOMENDACIONES

1. Se recomienda cumplir el proceso de datos, los supuestos estadísticos y econométricos y la validación como elemento de proyección para obtener las características matemáticas del modelo econométrico, la expresión final que se obtenga será el nuevo modelo econométrico de proyección de la demanda energética.

2. Para objetivos académicos, realizar las proyecciones de la demanda energética bajo dos metodologías diferentes para poder hacer comparaciones en el nivel de ajuste que tenga la metodología, respecto a los resultados del análisis proyectivo econométrico que se realice, se obtendría potenciar los alcances del aprendizaje y la investigación aplicada.

3. Se recomienda proyectar la variable objetivo de acuerdo al horizonte histórico que se dispone, si se tiene un horizonte histórico mensual o anual de corto plazo, el horizonte de proyección debe ser de corto plazo (1 a 2 años), si se dispone de un horizonte histórico mensual o anual de largo plazo, se puede realizar proyecciones de largo plazo (6 a 25) años.

BIBLIOGRAFIA

[1] Box, Jenkins, Reinsel, Time Series Analysis: Forecasting & Control, Wiley series in probability and Statistics. 1976

[2] Butrón F. Cesar (2019) “El sector eléctrico peruano: Situación y perspectivas”, COES, Febrero, 2019.

[3] Caridad, J. Ocerin (2012) Econometría: Modelos econométricos y series temporales con los paquetes Utsp y TSP, Tomo 1: Modelos econométricos uniecuacionales, Editorial Reverte, S.A.

[4] Chong Fuentes Manuel Aurelio, tesis “Proyección de series de tiempo para el consumo de la energía eléctrica a clientes residenciales en Ecuador”, Universidad de Concepción Facultad de Ciencias Económicas y Administrativas, Concepción-Chile 2016.

[5] Granger Clive W. J. (2004) “Análisis de series temporales, cointegración y aplicaciones”, Universidad de California-San Diego, Revista asturiana de economía - RAE Nº 30 2004.

Greene, W. H. (2008). Econometric Analysis. Pearson/Prentice Hall.

[6] Gonzales Chávez, Salome. (2019) COES, Proyección de la Demanda Vegetativa del SEIN Mediante Modelos Econométricos Periodo 2021-2030, Lima-Perú.

[7] Gonzales Chávez, Salome (2019) “Planeamiento y gestión de sistemas eléctricos”, Universidad Nacional de Ingeniería, Lima-Perú, 2019.

[8] Gil, S. (2007). Gas natural en la Argentina: presente y futuro. Escuela de Ciencia y Tecnología, UNSAM, y Departamento de Física, FCE y N-UBA, 17(101), pp. 26-36

[9] González, G. Rodríguez, D. y Ríos, P. (2018). Estimation of Residential Natural Gas Consumption in Medellin-Antioquia. IEEE LATIN AMERICA TRANSACTIONS, 16(3).

[10] Greene, W. H. (2008). Econometric Analysis. Pearson/Prentice Hall.

[11] Gujarati, D. Porter, D. (2009) “Econometría”, Quinta edición, McGraw-Hill, Impreso en México.

[12] Laub y Quijandria (2019) “Estudio técnico y económico para la determinación de las tarifas en barra Mayo 2019 – Abril 2020”, Sub comité de generadores, Enero 2019.

[13] Loria E. (2007) “Econometría con aplicaciones”, Universidad autónoma de México, Pearson educación, México, 2007.

[14] Mauricio, J. (2007) Introducción al análisis de series temporales, Universidad complutense de Madrid.

[15] MEM, (2021). Principales indicadores del sector eléctrico a nivel nacional. Recuperado: https://www.minem.gob.pe/minem/archivos/1%20Cifras%20preliminares%20del%20Sector%20Electrico%20-%20Enero%202021-Rev2.pdf

[16] Montero, R. (2013) “Variables no estacionarias y cointegracion” Documento de trabajo en economía aplicada. Universidad de Granada. España. Marzo, 2013.

[17] Osinergim (2014) Supervisión de contratos de centrales de generación y líneas de transmisión de energía eléctrica en operación, Octubre 2014.

[18] Reyes Francisco M. (2016), “Predicción de la demanda eléctrica: Comparativa ARIMA – Redes Neuronales mediante software SPSS”, Universidad de Sevilla, España.

[19] Statistical Review of World Energy (2020) | 69th edition. Recuperado de https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy/natural-gas.html

[20] Velásquez, J. Olaya, Y. y Franco, C. (2011). Análisis y predicción de series de tiempo en mercados de energía usando el leguaje R.  Dyna, N° 165, pp. 287­296 Medellín-Colombia.

[21] Zarate, I. M. Guzmán, E. S. Hernández, J. M. y Rebollar S. R. (2015). Análisis económico del consumo de gas natural para generación de energía eléctrica en México: 2005-2014. Pistas Educativas N° 115.

[22] Zúñiga, S. (2004) Econometría practica con Excel, Universidad católica del Norte, Julio, 2004.

**https://www.linkedin.com/in/roberto-alvarez-208a1825b/**