



Pontificia Universidad
JAVERIANA
Cali

**IMPACTO MACROECONÓMICO EN TARIFAS ENERGÉTICAS COLOMBIANAS: ANÁLISIS CON
MODELO DE CIENCIA DE DATOS**

Ivan Alejandro Posada Mesa

Código 8992276

*Proyecto Aplicado para optar al título de
Magister en Ciencia de Datos*

Director(a)

Mgtr. Orlando Joaqui Barandica

FACULTAD DE INGENIERÍA Y CIENCIAS

MAESTRÍA EN CIENCIA DE DATOS

SANTIAGO DE CALI, JUNIO DE 2025

FICHA RESUMEN

TÍTULO DEL PROYECTO: Impacto macroeconómico en tarifas energéticas colombianas: análisis con modelo de ciencia de datos

1. **ÁREA DE TRABAJO:** Sector eléctrico colombiano.
2. **TIPO DE PROYECTO:** Investigación.
3. **ESTUDIANTE:** Ivan Alejandro Posada Mesa.
4. **CORREO ELECTRÓNICO:** alejandroposada@javerianacali.edu.co.
5. **DIRECCIÓN Y TELEFONO:** Ibagué – Tolima. 300 6509223.
6. **DIRECTOR:** Orlando Joaqui Barandica.
7. **VINCULACIÓN DEL DIRECTOR:** Profesor de planta Pontificia Universidad Javeriana de Cali
8. **CORREO ELECTRÓNICO DEL DIRECTOR:** orlando.joaqui@javerianacali.edu.co
9. **PALABRAS CLAVE:** Colombia - Sector energético – Variables Macroeconómicas – Tarifa de energía
10. **FECHA DE INICIO:** 02 de octubre de 2024
11. **DURACIÓN ESTIMADA:** 11 meses
12. **RESUMEN:** El presente proyecto aplicado se centró en el análisis de la relación dinámica entre factores macroeconómicos y las tarifas de energía eléctrica en Colombia, un tema de crucial importancia dado el impacto del costo energético en la competitividad económica, el bienestar social y la planificación sectorial. En el contexto colombiano, la volatilidad de las variables macroeconómicas y las particularidades del mercado energético hacen esencial comprender cómo los choques económicos se transmiten a los precios de la electricidad. La problemática abordada fue la necesidad de identificar y modelar estas interrelaciones para mejorar la capacidad de previsión de las tarifas, lo cual resulta fundamental para la toma de decisiones de los agentes del sector, los reguladores y los consumidores. Los objetivos propuestos incluyen la recolección y limpieza de datos tarifarios y macroeconómicos, la reducción de la dimensionalidad de las variables macro a través de PCA, el diseño y entrenamiento de un modelo VAR, el análisis de los impulsos de respuesta y el desarrollo y evaluación de modelos de pronósticos (VAR, SARIMAX, XGBoost y VAR+XGBoost). Los principales resultados obtenidos revelan relaciones dinámicas específicas entre ciertas variables macro y componentes tarifarias, y la evaluación de los modelos de pronóstico sugirió que la combinación VAR+XGBoost demostró, en general, la mayor capacidad para capturar la dinámica real de las series temporales. Las posibles aplicaciones de este trabajo radican en la mejora de la planificación estratégica de las empresas energéticas, la optimización de las políticas regulatorias, la gestión de riesgos financieros y la provisión de información más precisa a los consumidores. Las conclusiones más relevantes señalan la complejidad de la interacción entre la macroeconomía y las tarifas eléctricas, la utilidad del análisis de impulsos para identificar relaciones significativas y el potencial de los modelos híbridos como el VAR+XGBoost para mejorar la precisión de los pronósticos en este sector.

TABLA DE CONTENIDO

INTRODUCCIÓN	1
1. CONTEXTUALIZACIÓN DEL PROYECTO	2
1.1. Definición del problema.....	2
1.1.1. Planteamiento del problema.....	2
1.1.2. Formulación del problema	4
1.1.3. Sistematización.....	4
1.2. Objetivos	4
1.2.1. Objetivo General.....	4
1.2.2. Objetivos Específicos	4
1.3. Marco de Referencia	5
1.3.1. Marco Teórico	5
1.3.2. Antecedentes.....	6
2. RECOLECCIÓN Y LIMPIEZA DE DATOS	9
2.1. Base de datos de variables macroeconómicas colombianas	9
2.2. Base de datos de tarifas de energía	10
3. REDUCCIÓN DE LA DIMENSIONALIDAD DE LAS VARIABLES.	11
3.1. Base de datos de variables macroeconómicas colombianas	11
3.2 Base de datos de tarifas de energía	14
3.2.1. Análisis de varianza y correlación	15
3.3. Base de datos final	17
4. DISEÑAR Y ENTRENAR UN MODELO DE VECTORES AUTORREGRESIVOS (VAR) UTILIZANDO LAS TARIFAS DE ENERGÍA ELÉCTRICA Y LOS FACTORES MACROECONÓMICOS IDENTIFICADOS.....	19
4.1. Prueba de Augmented Dickey-Fuller (ADF).....	19
4.2. Prueba de Phillips-Perron (PP).....	19
4.3. Prueba Kwiatkowski-Phillips-Schmidt-Shin (KPSS)	20
4.4. Convertir las series en estacionarias.....	20
4.5. Modelo de Vector AutoRegresivo (VAR)	21
5. EVALUAR EL DESEMPEÑO DEL MODELO VAR EN TÉRMINOS DE PRECISIÓN Y CONFIABILIDAD.	23
5.1. Dimensiones.....	23
5.1.1. Dimensión 1.....	23
5.1.2. Dimensión 2.....	24

5.1.3.	Dimensión 3.....	24
5.1.4.	Dimensión 4.....	25
5.1.5.	Dimensión 5.....	26
5.1.6.	Impulso en la tarifa total calculada (Cu).....	26
5.2.	Variables Macroeconómicas Individuales	27
5.2.1.	Recibos Fiscales	28
5.2.2.	Precios al productor Nacional	29
5.2.3.	Precios al productor Interno	29
5.2.4.	Empleo en el sector de abastecimiento de electricidad, gas y agua.....	30
5.2.5.	Tipo de cambio (COP).....	30
5.2.6.	Crédito a negocios en circulación.....	31
5.2.7.	Deuda externa	32
5.2.8.	Población	32
5.2.9.	Demanda de energía	33
5.2.10.	Generación hidráulica	33
5.2.11.	Generación Fósil	34
5.3.	Análisis final	34
6.	IMPLEMENTAR EL MODELO VAR ENTRENADO PARA PREDECIR LAS TARIFAS DE ENERGÍA ELÉCTRICA EN FUNCIÓN DE LOS DISTINTOS PARÁMETROS MACROECONÓMICOS IDENTIFICADOS.	37
6.1.	Modelo de Vectores Autorregresivos (VAR).....	37
6.2.	Modelo Autorregresivo Integrado de Media Móvil con Explicativas (SARIMAX).....	38
6.3.	Algoritmo XGBoost (Extreme Gradient Boosting):.....	39
6.4.	Modelo Combinado VAR + XGBoost	40
6.5.	Variable G.....	40
6.6.	Variable T	41
6.7.	Variable D.....	42
6.8.	Variable C	43
6.9.	Variable PR	44
6.10.	Variable R	45
6.11.	Análisis final	46
7.	CONCLUSIONES Y TRABAJOS FUTUROS.....	48
7.1.	CONCLUSIONES	48
7.1.1.	Conclusión general.....	48

7.1.2. Conclusiones específicas.....	48
7.2. TRABAJOS FUTUROS.....	49
8. REFERENCIAS BIBLIOGRÁFICAS	51

LISTA DE ILUSTRACIONES

Ilustración 1 Lo que se paga en la factura de energía	2
Ilustración 2 Relación entre antecedentes	8
Ilustración 3. Variables que más afectan la dimensión 1	12
Ilustración 4.. Variables que más afectan la dimensión 2	12
Ilustración 5.. Variables que más afectan la dimensión 3	13
Ilustración 6. Variables que más afectan la dimensión 4	13
Ilustración 7. Variables que más afectan la dimensión 5	14
Ilustración 8 Frecuencia de comercializadores.....	15
Ilustración 9 Histogramas por componentes.....	15
Ilustración 10. Varianza y desviación estándar tarifaria	16
Ilustración 11. Flujo para la creación de la base de datos final	18
Ilustración 12 Series de tiempo no estacionarias.....	20
Ilustración 13 Series de componentes estacionarios	21
Ilustración 14 Series variables macroeconómicas estacionarias.....	21
Ilustración 15. Expresión del modelo VAR.....	22
Ilustración 16. Impulsos de respuesta: Dimensión 1.....	23
Ilustración 17. Impulsos de respuesta: Dimensión 2.....	24
Ilustración 18. Impulsos de respuesta: Dimensión 3.....	24
Ilustración 19. Impulsos de respuesta: Dimensión 4.....	25
Ilustración 20. Impulsos de respuesta: Dimensión 5.....	26
Ilustración 21. Impulsos de respuesta: Todas las dimensiones en el CU.....	26
Ilustración 22. Variables macroeconómicas individuales	28
Ilustración 23. Impulsos de respuesta: Recibos fiscales.....	28
Ilustración 24. Impulsos de respuesta: Precios al productor nacional	29
Ilustración 25. Impulsos de respuesta: Precios al productor interno.....	29
Ilustración 26. Impulsos de respuesta: Empleo en el sector eléctrico, gas y agua.....	30
Ilustración 27. Impulsos de respuesta: Tipo de cambio (COP)	30
Ilustración 28. Impulsos de respuesta: Crédito a negocios en circulación	31
Ilustración 29. Impulsos de respuesta: Deuda externa	32
Ilustración 30. Impulsos de respuesta: Población	32

Ilustración 31. Impulsos de respuesta: Demanda de energía	33
Ilustración 32. Impulsos de respuesta: Generación hidráulica	33
Ilustración 33. Impulsos de respuesta: Generación fósil	34
Ilustración 34. Expresión del modelo SARIMAX	38
Ilustración 35. Predicción Agregada del Modelo XGBoost	39
Ilustración 36. Función Objetivo Minimizada por XGBoost	39
Ilustración 37. Pronóstico de la variable G	40
Ilustración 38. Pronóstico de la variable T	41
Ilustración 39. Pronóstico de la variable D	42
Ilustración 40. Pronóstico de la variable C	43
Ilustración 41. Pronóstico de la variable PR	44
Ilustración 42. Pronóstico de la variable R	45

INTRODUCCIÓN

El sector energético colombiano se caracteriza por la alta variabilidad de las tarifas de energía eléctrica, lo que ha generado un impacto significativo en los usuarios, especialmente en los regulados, quienes representan alrededor del 70% de la demanda total de energía en el país. Esta variabilidad ha sido atribuida a diversos factores, entre los que se destacan los componentes de la tarifa, como la generación, transmisión, distribución, comercialización, restricciones y pérdidas. Sin embargo, se desconoce en gran medida la influencia de las variables macroeconómicas en estos componentes y su impacto directo en el precio final de la tarifa.

En este contexto, surgió la necesidad de comprender mediante el uso de ciencia de datos el papel de las variables macroeconómicas en la dinámica de las tarifas de energía eléctrica en Colombia. Este trabajo de investigación tuvo como objetivo principal evaluar el impacto de factores macroeconómicos sobre el costo de la tarifa de energía eléctrica en Colombia mediante el uso de modelos de Vectores Autorregresivos (VAR) y técnicas de aprendizaje supervisado.

Para alcanzar este objetivo, se realizó una metodología que incluye la recopilación de datos históricos de tarifas y variables macroeconómicas relevantes, la identificación de las variables con mayor impacto mediante técnicas de aprendizaje no supervisado, como análisis de componentes principales (PCA), la evaluación de su relación con la variabilidad de las tarifas a través de la aplicación de modelos vectoriales autorregresivos por sus siglas en inglés VAR y la implementación del modelo VAR entrenado para predecir las tarifas de energía eléctrica en función de los distintos parámetros macroeconómicos identificados. Los resultados de este estudio contribuyen a una mejor comprensión del funcionamiento del mercado energético colombiano y sirven de base para que en un futuro se puedan diseñar políticas públicas más efectivas que garanticen un acceso a la energía asequible y sostenible para todos los colombianos.

Este estudio permitió identificar las variables macroeconómicas que más han incidido en la variabilidad del precio de la energía eléctrica en Colombia, y cómo estas han impactado a los usuarios regulados. Específicamente, se identificaron un conjunto de variables macroeconómicas relevantes que han tenido y que tendrían a futuro un impacto significativo en el costo de la tarifa de energía eléctrica, se cuantificó el impacto individual de cada variable macroeconómica identificada en el costo de la tarifa de energía eléctrica, y se analizó la dinámica del impacto de las variables macroeconómicas a lo largo del tiempo y finalmente se examinó la relación entre la variabilidad de las tarifas y el comportamiento de las variables macroeconómicas identificadas.

1. CONTEXTUALIZACIÓN DEL PROYECTO

1.1. Definición del problema

1.1.1. Planteamiento del problema

Las tarifas de energía eléctrica en Colombia han experimentado una variabilidad significativa en los últimos años, afectando a los usuarios regulados (Se considera usuario regulado a toda persona natural o jurídica cuyas compras de electricidad están sujetas a tarifas establecidas por la comisión de regulación de energía y gas – CREG [1]) de manera considerable generando de esta forma un impacto significativo en los hogares y las empresas, afectando su presupuesto, competitividad y bienestar general. Si bien el cálculo de las tarifas se basa en seis componentes (Generación, Transmisión, Distribución, Comercialización, Restricciones y Pérdidas [2]), la comprensión de las variables macroeconómicas que inciden en estos componentes y su impacto directo en el precio final de la tarifa aún no es del todo clara.

Diversos factores contribuyen a esta variabilidad. Entre los más relevantes se encuentran la tarifa de energía y la demanda del usuario. La tarifa de energía se establece por estrato y mercado, y varía según la región y el tipo de consumidor, por otro lado, la demanda se ve afectada por el consumo de energía eléctrica del usuario durante un periodo específico.

Como se mencionaba anteriormente la tarifa está compuesta por 6 componentes, en marzo de 2024 las componentes que experimentaron un mayor aumento fue la de generación con 11.3% y la de distribución con 9.6% las cuales representan el 73% del total de la tarifa [3].

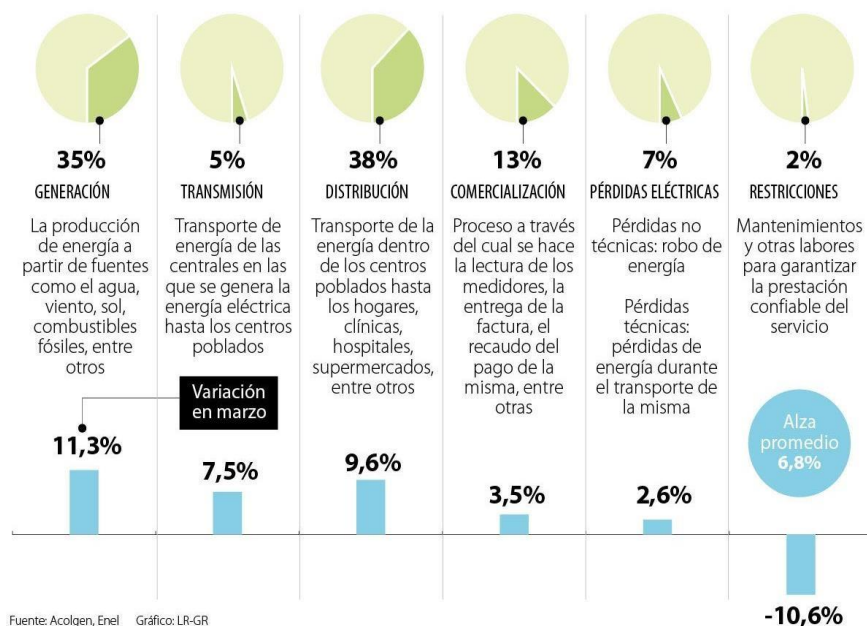


Ilustración 1 Lo que se paga en la factura de energía

En marzo de 2024, la tarifa de energía eléctrica en Colombia experimentó un incremento del 19,51% en comparación con marzo de 2023, siendo las ciudades más afectadas Santa Marta, Riohacha, Barranquilla, Valledupar y Montería, con aumentos que superaron el 30% [4]. A este incremento en la tarifa se suma un aumento del 12,06% en el consumo de energía del mercado regulado durante marzo de 2024, en comparación con el mes anterior. Las regiones con mayor aumento en el consumo fueron Guaviare, Tolima, Huila y Caquetá, con un promedio del 15,54% [5]. Cabe destacar que el mercado regulado representa alrededor del 70% de la demanda total de energía en el país, mientras que el mercado no regulado representa el 30% restante. En conjunto, estos factores han generado un aumento significativo en el costo final de la energía para los usuarios en Colombia. Esta situación ha motivado al gobierno colombiano a prestar especial atención al sector eléctrico tal como se evidencio en mayo de 2024 donde el actual presidente de Colombia Gustavo Petro le “ordenó al Ministerio de Minas y Energía bajar las tarifas de energía del país en menos de un mes” [6] y es que no se debe olvidar que el servicio de la energía no tiene un sustituto, por lo que se considera un servicio esencial en el país.

Algunas Variables macroeconómicas que afectan directamente algunos componentes de la tarifa pueden ser el IPP (Índice de precios al productor) tasa con la cual se indexan los contratos de compra de energía entre otros costos asociados al negocio, los precios internacionales de los combustibles, ya que algunas tecnologías de generación requieren de combustible para su funcionamiento, como lo pueden ser las plantas térmicas, o el Producto Interno Bruto (PIB) se sabe por estudios anteriores que el crecimiento del PIB está estrechamente relacionado con la demanda de energía eléctrica. Un mayor crecimiento económico implica un aumento en la demanda de energía, lo que puede generar presiones sobre los precios, especialmente en el componente de generación [7].

Las variables macroeconómicas mencionadas anteriormente interactúan de manera compleja y dinámica, generando un impacto significativo en los precios de la energía eléctrica en Colombia. Es importante comprender estas interacciones para formular políticas públicas efectivas que permitan mitigar el impacto negativo de estas variables en los usuarios, especialmente en los más vulnerables.

La compleja interacción entre estos factores dificulta comprender en su totalidad las causas de la variabilidad tarifaria y, por ende, formular soluciones efectivas. Si la situación actual persiste, se prevén consecuencias negativas como una mayor presión sobre el presupuesto familiar, especialmente las de bajos ingresos, para cubrir sus necesidades básicas, en relación con esto dichos hogares podrían verse desproporcionadamente afectados por los aumentos de precios, ampliando la brecha entre diferentes sectores de la población, adicional la incertidumbre sobre los precios de la energía podría desincentivar la inversión en el sector eléctrico, afectando la seguridad energética del país y los altos precios de la energía podrían afectar negativamente la competitividad de las empresas colombianas en el mercado global.

La ciencia de datos emerge como una herramienta poderosa para abordar este complejo problema. A través del análisis de grandes conjuntos de datos históricos y actuales, se logró identificar patrones y relaciones entre las variables macroeconómicas, los componentes de la tarifa y el precio final de la energía. El empleo de modelos como el VAR (Vector Autoregresivo) permitió modelar simultáneamente la relación entre las variables, capturando las interacciones endógenas y los efectos de retroalimentación entre ellas. Esto a su vez permitió comprender mejor la dinámica de la variabilidad tarifaria y en un futuro a quien interese le permitirá formular políticas públicas más efectivas que promuevan un sector energético sostenible, competitivo y accesible para todos los colombianos.

El desarrollo de este trabajo se enfocó en analizar el impacto de las variables macroeconómicas en el costo

de la tarifa de energía eléctrica en Colombia.

1.1.2. Formulación del problema

¿Cuáles son los factores macroeconómicos que más han influido en la variabilidad del costo de la tarifa de energía eléctrica para usuarios regulados en Colombia, y cómo podemos identificar y cuantificar su impacto utilizando técnicas de ciencia de datos?

1.1.3. Sistematización

¿Cómo se pueden utilizar las técnicas de ciencia de datos para identificar y cuantificar el impacto de las variables macroeconómicas en la variabilidad de las tarifas de energía eléctrica en Colombia?

¿Cuáles son las variables macroeconómicas más relevantes para analizar la variabilidad de las tarifas de energía eléctrica en Colombia?

¿Qué métodos de recolección de datos se utilizarán?

¿Qué técnicas de análisis de datos se utilizarán?

¿Qué fuentes de datos se utilizarán?

¿Cómo se recolectarán los datos?

¿Cómo se garantizará la calidad de los datos?

¿Cómo se interpretarán los resultados del análisis de datos?

¿Cuáles son las principales conclusiones de la investigación?

1.2. Objetivos

1.2.1. Objetivo General

Evaluar el impacto de factores macroeconómicos sobre el costo de la tarifa de energía eléctrica en Colombia mediante el uso de modelos de Vectores Autorregresivos (VAR) y técnicas de aprendizaje supervisado.

1.2.2. Objetivos Específicos

1. Recopilar y limpiar datos históricos de las tarifas de energía eléctrica y las variables macroeconómicas identificadas.
2. Utilizar técnicas de aprendizaje no supervisado, como análisis de componentes principales (PCA), para identificar y reducir la dimensionalidad de las variables macroeconómicas.
3. Diseñar y entrenar un modelo de Vectores Autorregresivos (VAR) utilizando las tarifas de energía eléctrica y los factores macroeconómicos identificados.
4. Evaluar el desempeño del modelo VAR en términos de precisión y confiabilidad.
5. Implementar el modelo VAR entrenado para predecir las tarifas de energía eléctrica en función de los distintos parámetros macroeconómicos identificados.

1.3. Marco de Referencia

1.3.1. Marco Teórico

El análisis de la tarifa de energía eléctrica en Colombia requiere la comprensión de un marco regulatorio complejo y la influencia de diversas variables macroeconómicas. Este proyecto se fundamentó en el estudio de los siguientes conceptos clave, que permitieron analizar la estructura de la tarifa y su relación con los indicadores macroeconómicos.

1.3.1.1. Aspectos Regulatorios y Tarifarios

La Ley 142 y la Ley 143 de 1995 establecieron el marco legal para el sector eléctrico colombiano, definiendo los roles de los agentes y los mecanismos de regulación. En este proyecto, la comprensión de estas leyes fue fundamental para analizar cómo la regulación influye en la determinación de la tarifa de energía. Por ejemplo, se analizó cómo los mecanismos de control tarifario establecidos en estas leyes impactaron los costos finales para los usuarios.

La Resolución CREG 119 de 2007, que define la fórmula tarifaria, fue un elemento central en este análisis. La tarifa de energía eléctrica para usuarios regulados en Colombia se determina a partir del Costo Unitario de Prestación del Servicio (CU), que representa el costo económico eficiente de la cadena eléctrica, expresado en pesos por kilovatio hora [15]. Este CU se desglosa en componentes que reflejan los costos de compra de energía (G), transporte (T), distribución (D), comercialización (Cv), restricciones (R) y pérdidas (PR) [16]. La tarifa final se obtiene al aplicar al CU los factores de subsidio o contribución correspondientes según el estrato del usuario, o se iguala al CU para usuarios no subsidiados. Este desglose permite comprender cómo los costos asociados a cada etapa de la prestación del servicio eléctrico contribuyen al valor final que paga el usuario.

Los conceptos de usuario regulado y no regulado también fueron importantes para delimitar el alcance del análisis. Este proyecto se enfocó en usuarios regulados, cuyas tarifas están sujetas a la regulación de la CREG. [17]

1.3.1.2. Aspectos Del Sector Eléctrico

El sector eléctrico se basa en la potencia y la energía eléctrica, medidas fundamentales para su funcionamiento [9] [10]. La generación de energía se realiza a través de diversas tecnologías, incluyendo la térmica, que utiliza la combustión de materiales orgánicos para generar calor, la hidroeléctrica, que aprovecha la energía del agua [12], entre otras. Estas tecnologías tienen características y costos que influyen en la tarifa de energía. La transmisión, distribución y comercialización son etapas esenciales para llevar la energía desde las plantas de generación hasta los usuarios finales [13] [14], y cada una de estas etapas contribuye a la estructura de costos del sector y, por ende, a la tarifa final.

1.3.1.3. Aspectos Económicos

Las variaciones en las diferentes variables macroeconómicas afectaron las tarifas de energía eléctrica en Colombia. El Índice de Precios del Productor (IPP) que refleja la variación de los precios de los bienes en la etapa de producción [19], y el Índice de la Tasa de Cambio Real (ITCR) que compara el valor de una canasta de bienes en dos países, expresados en la misma moneda [18] fueron variables macroeconómicas clave en este análisis.

1.3.1.4. Aspectos Metodológico

La ciencia de datos se aplicó para extraer información significativa de grandes volúmenes de datos, procesarlos y finalmente crear un modelo capaz de pronosticar algunas variables de la tarifa de energía. Se utilizaron técnicas de análisis para identificar patrones y tendencias en las variables estudiadas. El modelo VAR fue una herramienta fundamental para analizar la interrelación entre las variables económicas y su impacto en la tarifa de energía. Este modelo permitió examinar cómo los cambios en una variable afectaron a la tarifa. [8] y [11]

1.3.2. Antecedentes

El estudio del impacto de las variables macroeconómicas en el sector energético ha sido abordado desde diversas perspectivas por investigadores en el campo. A continuación, se presenta una síntesis de cuatro trabajos relevantes que identifican variables macroeconómicas con influencia en la tarifa de energía:

1.3.2.1. Trabajo 1

Corresponde a Jhon Alexis Angarita Donado y Walter Pérez Bermúdez (2021), quienes realizaron la investigación del “Comportamiento tarifario de energías renovables tipo eólica y solar en la generación de energía eléctrica en Colombia” [22], esto con el fin de identificar si las tarifas con Fuentes No Convencionales de Energía Renovables (FNCER) son competitivas con respecto a las fuentes de energía convencionales.

la investigación se dividió en dos partes “... Primero, se presenta el marco legal que actualmente tiene el país en el ámbito de la promoción de las FNCER. Segundo, se modela el comportamiento histórico de las tarifas de energía eléctrica en la bolsa nacional para proyectar precios futuros. Finalmente, se comparan los resultados obtenidos de la proyección de precios futuros de la bolsa nacional con respecto a las tarifas del compromiso de energía establecido en la subasta de la Unidad de Planeación Minero-Energética (UPME) con FNCER tipo eólica y solar.”. El estudio arrojó resultados concluyentes que confirman que la generación eólica y solar no solo presenta precios inferiores al precio de bolsa de energía, sino que también contribuye a mitigar la volatilidad de esta.

Este trabajo se relaciona con la investigación en curso, ya que investiga el precio de compra de contratos y bolsa de energía, las cuales impactan directamente en el cálculo de la componente G de la tarifa de energía.

1.3.2.2. Trabajo 2

Javier Eduardo Duarte Aunta (2023), buscaba "Determinar la variabilidad de la demanda de energía eléctrica que permita evaluar el potencial uso de un esquema tarifario diferencial" [23] esto con el fin de estimar la flexibilidad del consumo energético en Colombia.

El trabajo se desarrolló en tres fases la primera se basa en un pre-procesamiento de datos, centrado en la organización y limpieza de registros individuales de consumo. Seguidamente, se realiza un procesamiento y clasificación de los datos mediante técnicas de análisis de variabilidad y clustering, para terminar con un análisis de resultados. El desarrollo del trabajo permitió identificar que la demanda se comporta diferente según el tipo de día, existe un comportamiento para los días ordinarios (de lunes a viernes), para los sábados y para los domingos y festivos, adicionalmente se encontró que en los días ordinarios “Las primeras horas de la mañana, particularmente entre las 6 y las 9 son las más adecuadas para la

implementar tarifas de Time-of-Use, ya que en este intervalo se observaron altos picos de variabilidad en los clústeres con mayor número de usuarios.”

Este trabajo se relaciona con la investigación en curso, ya que esta evaluación “...es clave para la posible implementación de esquemas tarifarios diferenciados, en particular tarifas Time-of-use (ToU).”

1.3.2.3. Trabajo 3

La subdirección de hidrocarburos del Ministerio de Minas y Energía de la República de Colombia (2023), realizó una investigación que buscaba encontrar la “Proyección de precios de los energéticos para generación eléctrica” [24], el trabajo se enfocó en realizar la proyección de los precios del combustible que utilizan las plantas térmicas para su generación, entre ellas esta Diesel, JET-A1, FuelOil, gasolina, gas natural, carbón y GLP. Es importante aclarar que en épocas de fenómeno del niño en donde la generación hidráulica es escasa, el precio de compra de energía en bolsa se cierra con generación de plantas terminas, que, aunque son más costosas a la hora de producir, tienen la ventaja de contar con combustible en casi cualquier momento del año.

La realización de los estimados “...se calculan en función de diversas variables que afectan directamente el resultado final, entre las que se encuentran las expectativas del mercado, la oferta y demanda, cambios tecnológicos, ciclos económicos, así como, la normatividad y regulación vigente de carácter nacional para cada uno de los energéticos analizados, entre otros.”. una vez obtenidos los estimados realizaron un análisis a corto, mediano y largo plazo. La conclusión que se puede obtener del documento es que los combustibles más económicos en el corto y largo plazo son el Diesel nacional, con un precio al largo plazo de \$6,1/MBTU y el carbón con un precio que oscila entre \$6.5/MBTU y \$2.6/MBTU según la planta.

Este trabajo se relaciona con la investigación en curso, ya que investiga el precio del combustible que utilizan las plantas terminar y que como se explicaba anteriormente afecta directamente el cálculo del precio de la bolsa de energía, impactando directamente en el cálculo de la componente G de la tarifa de energía.

1.3.2.4. Trabajo 4

Barrientos J, Rodas E, Velilla E, Lopera M y Villada F [25] en 2012 realizaron una investigación que mediante el uso de ciencia de datos diseñaron un modelo para el pronóstico del precio de compra de energía eléctrica en bolsa en Colombia, las variables utilizadas fueron el precio histórico de la energía, demanda de energía, oferta de energía, nivel de los embalses, el PIB y el crecimiento de la demanda, como modelos utilizaron dos tipos un primer modelo basado en redes neuronales artificiales (RNA) y un segundo modelo econométrico. En las conclusiones de destaca que ambos modelos (RNA y econométricos) encontraron una relación inversa entre el precio de la energía y el nivel de los embalses, el modelo econométrico tuvo un mejor desempeño al modelar el precio en función del PIB, el nivel de embalses y la demanda, el modelo neuronal tuvo un mejor desempeño al pronosticar los precios medios para el período 2010-2017, oscilando entre 100 \$/kWh y 150 \$/kWh y los precios proyectados por ambos modelos son consistentes con estudios recientes de la UPME.

En general, este estudio representa una valiosa contribución a la comprensión del mercado energético colombiano y proporciona herramientas útiles para pronosticar el precio de bolsa de la energía eléctrica en el futuro.

1.3.2.5. Análisis De Antecedentes

En general, los cuatro trabajos analizados resaltan la importancia de considerar diversas variables, incluyendo el precio de compra de energía, la variabilidad de la demanda, la transición energética, el contexto internacional, el impacto macroeconómico y el uso de ciencia de datos, para comprender el complejo panorama del sector energético y la determinación de las tarifas de energía eléctrica.

A continuación, se presenta una tabla resumen que sintetiza las relaciones entre los trabajos:

<i>Trabajo</i>	<i>Precio de compra de energía</i>	<i>Variabilidad de la demanda</i>	<i>Transición energética</i>	<i>Contexto internacional</i>	<i>Impacto macroeconómico</i>	<i>Ciencia de datos</i>
1	Si	No	Si	Mencionado	No	No
2	No	Si	No	No	No	No
3	Si	No	No	No	No	No
4	Si	Si	No	No	Si	Si

Ilustración 2 Relación entre antecedentes

Si bien los cuatro trabajos revisados aportan valiosas perspectivas sobre el sector energético en Colombia, ninguno de ellos explora específicamente el tema central de esta investigación: el análisis del Impacto macroeconómico en tarifas energéticas colombianas: análisis con modelo de ciencia de datos

Esta brecha en la investigación existente resalta la necesidad de un estudio que aborde de manera integral la influencia de las variables macroeconómicas en la variabilidad de las tarifas de energía eléctrica. Este trabajo, al enfocarse en este tema, tiene el potencial de llenar un vacío importante en el conocimiento y contribuir a la comprensión de los factores que determinan el precio de la energía en Colombia.

2. RECOLECCIÓN Y LIMPIEZA DE DATOS

2.1. Base de datos de variables macroeconómicas colombianas

Se utilizó una base de datos que provenino de los estudios “Exploring the asymmetric relationship between macroeconomic factors and corporate profitability in the MSCI Colombia index” [19] y el estudio “Macrofinancial interconnections in the Pacific Alliance: a quantile approach of stock markets and macroeconomic factors” [20]. Esta base, con frecuencia mensual desde febrero de 2000 hasta septiembre de 2023, contiene 153 variables macroeconómicas. Estas variables se encuentran organizadas en diversas categorías que buscan cubrir un amplio espectro de la actividad económica colombiana, incluyendo encuestas y pronósticos económicos, indicadores de precios al consumidor y al productor, cuentas nacionales (como el producto interno bruto y sus componentes), datos monetarios y financieros (tasas de interés, agregados monetarios, crédito), estadísticas del mercado laboral (empleo, desempleo), indicadores de diversos sectores industriales (manufactura, construcción, comercio, etc.), entre otros.

Dentro del conjunto de variables disponibles, y más allá de las mencionadas en los estudios originales, se encuentran indicadores relevantes para el análisis macroeconómico colombiano, tales como:

- Recibos fiscales: Desagregados en diferentes tipos de impuestos como el impuesto nacional, el impuesto al valor agregado (IVA), impuestos sobre activos financieros, ingresos de impuesto de timbre, impuestos sobre comercio exterior y otros impuestos.
- Saldos de cambio de divisas: Ofreciendo una perspectiva sobre la dinámica del mercado cambiario.
- Índices de precios: Incluyendo el Índice de Precios al Consumidor (IPC) en diferentes desagregaciones (Core 15, regulados, excluyendo alimentos y regulados, alimentos, bienes sin alimentos ni regulados, servicios sin alimentos ni regulados) y el Índice de Precios al Productor (IPP) para el mercado interno e importado, así como índices de precios de exportación e importación.
- Mercado laboral: Además de la tasa de desempleo y el empleo total, se incluyen datos de empleo por principales áreas metropolitanas y por sectores de actividad económica.
- Sector de la construcción: Índices de costos de construcción residencial, diferenciando por tipo de vivienda (unifamiliar, multifamiliar, asequible) y permisos de construcción.
- Indicadores de confianza: Como el índice de incertidumbre de políticas económicas y la confianza de los consumidores.
- Sector financiero: Datos sobre préstamos no rentables, provisiones y tasas de interés para diferentes tipos de crédito.
- Comercio internacional: Balanza comercial, exportaciones e importaciones totales y desagregadas por tipo de bien.
- Deuda externa: Desagregada por sector público y privado.
- Variables monetarias y financieras: Oferta monetaria, reservas monetarias, crédito al sector público y privado, tasas de cambio (nominal y real), entre otras.

La riqueza y la amplitud de esta base de datos permiten explorar las complejas interrelaciones entre diversos aspectos de la economía colombiana y su potencial influencia en las tarifas de energía eléctrica, que es el objetivo central de esta investigación. Las variables específicas para incluir en el modelo se tomarán de los resultados del análisis de componentes principales (PCA).

El estudio titulado Exploring the asymmetric relationship between macroeconomic factors and corporate profitability in the MSCI Colombia index analiza cómo los factores macroeconómicos como la inflación, el comercio exterior, el empleo y la producción influyen en el rendimiento del índice bursátil MSCI Colombia. Utilizando técnicas estadísticas avanzadas, se identificó que estos factores tienen un impacto significativo y a menudo asimétrico en el mercado de valores colombiano. Por ejemplo, un aumento en la inflación puede tener efectos positivos o negativos en el corto y largo plazo, dependiendo de otros factores económicos. Asimismo, el estudio muestra que el desempeño del mercado está estrechamente vinculado al ciclo económico del país, reflejándose en la sensibilidad del índice a variables como el empleo y la producción.

El estudio titulado Macrofinancial interconnections in the Pacific Alliance: a quantile approach of stock markets and macroeconomic factors analiza cómo los mercados bursátiles de los países de la Alianza del Pacífico (Chile, Colombia, México y Perú) están conectados entre sí a través de factores macroeconómicos. Utilizando técnicas avanzadas de análisis de datos, los investigadores descubrieron que existe una fuerte interdependencia entre estos mercados, y que los shocks económicos en un país pueden tener un impacto significativo en los demás. Además, el estudio muestra que la creación de la Alianza del Pacífico ha intensificado estas conexiones, haciendo que los mercados sean más sensibles a los cambios económicos en la región. Los resultados de esta investigación son relevantes para inversores, policymakers y académicos interesados en entender las dinámicas de los mercados financieros en América Latina.

2.2. Base de datos de tarifas de energía

Se construyó una nueva base de datos con información de las tarifas de 16 de las principales comercializadoras de energía en Colombia, desde enero de 2009 hasta junio de 2024. Los datos fueron recopilados de las publicaciones oficiales de cada comercializadora en sus respectivas páginas web. Se seleccionaron únicamente las tarifas para usuarios con nivel de tensión 1 y propiedad del activo del 100% de la empresa, correspondientes al departamento de operación de cada comercializadora. Es importante destacar que la base de datos presenta algunos datos nulos, esto debido a la complejidad de extraer datos de plataformas web con enlaces dañados o estructuras variables.

Para abordar los valores faltantes en los datos, se implementó una estrategia de imputación basada en el promedio simple de los valores adyacentes correspondientes al mismo comercializador. En aquellos casos donde faltaba un dato intermedio (por ejemplo, marzo de 2023), se calculó el promedio entre el valor del mes anterior (febrero) y el siguiente (abril). Sin embargo, para los primeros años de la serie temporal donde no existían datos previos, se optó por eliminar las observaciones incompletas, dado que la imputación en estos casos podría introducir un sesgo significativo en el análisis. Se reconoce que este método de imputación, basado en el promedio simple, no es necesariamente el óptimo. No obstante, la evaluación de diversas técnicas de imputación excede el alcance del presente trabajo de tesis. Por consiguiente, se opta por este método sencillo, partiendo de la base de que las tarifas de energía tienden al aumento, por lo que imputar con un valor superior al mes siguiente podría resultar menos representativo de la realidad.

3. REDUCCIÓN DE LA DIMENSIONALIDAD DE LAS VARIABLES.

3.1. Base de datos de variables macroeconómicas colombianas

Para las variables macroeconómicas se decidió utilizar un análisis de componentes principales (PCA) para reducir la dimensionalidad. El análisis de componentes principales (PCA) es un método estadístico que permite simplificar la complejidad de espacios muestrales con muchas dimensiones a la vez que conserva su información. Supóngase que existe una muestra con n individuos cada uno con p variables (X_1, X_2, \dots, X_p), es decir, el espacio muestral tiene p dimensiones. PCA permite encontrar un número de factores subyacentes ($z < p$) que explican aproximadamente lo mismo que las p variables originales. Donde antes se necesitaban p valores para caracterizar a cada individuo, ahora bastan z valores. Cada una de estas z nuevas variables recibe el nombre de componente principal. [21]

Matemáticamente, una componente principal (PC_j) se expresa como una combinación lineal de las variables originales:

$$PC_j = a_{j1}X_1 + a_{j2}X_2 + \dots + a_{jp}X_p$$

Ilustración 3. PCA

donde a_{j1}, \dots, a_{jp} son los coeficientes de carga (o pesos) que indican la contribución de cada variable original a la componente principal j . Estos coeficientes se derivan de los vectores propios de la matriz de covarianza de los datos originales, y las componentes se construyen para ser ortogonales entre sí y para maximizar la varianza explicada.

Se aplicó un Análisis de Componentes Principales (PCA) utilizando el software R (empleando la función `prcomp()`) con el objetivo primordial de condensar la información contenida en un conjunto extenso de variables macroeconómicas en un número menor de factores sintéticos, denominados componentes principales. La justificación para seleccionar específicamente los primeros cinco componentes principales radica en su capacidad colectiva para capturar una porción sustancial de la variabilidad total presente en los datos originales. Específicamente, estos cinco primeros componentes explican el 77,77% de la variación total. Este umbral se demostró adecuado para retener una cantidad significativa de la información esencial de las variables macroeconómicas, minimizando al mismo tiempo la pérdida de información inherente a cualquier técnica de reducción de dimensionalidad. La decisión de detener la selección en el quinto componente se basó en un equilibrio entre la parsimonia del modelo (reducir el número de variables para evitar la sobreparametrización en modelos posteriores) y la necesidad de preservar una representación suficientemente rica de las dinámicas macroeconómicas relevantes para la predicción de las tarifas de energía eléctrica.

Finalmente, se identificaron las variables con mayores cargas en cada componente para proporcionar una interpretación más profunda de los factores obtenidos.

- **Factor 1 Factor Monetario:** mayormente influenciado por la Oferta monetaria, Población y los Precios del consumidor.

Variable	Dim.1
Peso colombiano	1.16366533
Población (Miles)	1.15983908
Services Ex Food and Regulated	1.15637505
Regulado	1.15565935
Core 15	1.15511158
Todo elemento	1.15291693
Todo elemento	1.15291433
Excluding Food and Regulated	1.15021563
Sector público y privado	1.14853963
Índice de costes de construcción residencial	1.14758013

Ilustración 4. Variables que más afectan la dimensión 1

- **Factor 2 Exposición Global:** mayormente influenciado por el Empleo, Tipos de cambio y el Crédito vivienda.

Variable	Dim.2
Construcción	3.0739957
Tipos de cambio efectivos (real)	3.03400691
Fin de próximo año	2.89110162
12 meses	2.79680488
Préstamos con mal desempeño	2.71797874
Fin de año	2.60704292
Tipo de cambio esperado	2.52490287
Transporte, almacenamiento y comunicaciones	2.41371332
Acciones (USD, Millones)	2.3568005
Fin de mes en curso	2.3214039

Ilustración 5.. Variables que más afectan la dimensión 2

- **Factor 3 Apertura comercial:** mayormente influenciado por Índices de precios de importación, Oferta monetaria, Términos de intercambio y Expectativas inflación.

Variable	Dim.3
Índices de precios de importación	5.17388814
Índices de precios de importación	5.17388814
Índices de precios de exportación	5.10002181
Índices de precios de exportación	5.10002181
Crédito neto doméstico al sector público	3.39448802
Términos de intercambio	3.27150176
Fin de año	3.21044943
Todo elemento	2.79111234
Cerraduras de hierro o acero	2.73394193
Otros	2.44208245

Ilustración 6.. Variables que más afectan la dimensión 3

- **Factor 4 Competitividad Externa:** mayormente influenciado por Préstamos, Índices de precios de exportación y Tipos de política.

Variable	Dim.4
Tipo de crédito de consumo, preferencia, ordinaria y Tesoro	4.85263225
Índices de precios de exportación	4.10197226
Índices de precios de exportación	4.10197226
Índices de precios de exportación	4.10197226
Tipos de política	3.9551086
Tipo aj por día laboral de crédito de consumo, preferencia, ordinaria y Tesoro	3.76299207
Índices de precios de exportación	3.40018514
Índices de precios de exportación	3.40018514
Tipo sobre créditos de consumo, preferencial, préstamos ordinarios ex. Tesorería	3.21526802
Todo elemento	3.12169298

Ilustración 7. Variables que más afectan la dimensión 4

- **Factor 5 Sector Inmobiliario y de construcción:** mayormente influenciado por Costes y precios de construcción, Expectativas inflación y Precios del consumidor.

Variable	Dim.5
Viviendas asequibles	12.4449813
Viviendas unifamiliares	12.3738473
Índice de costes de construcción residencial	12.3679867
Viviendas multifamiliares	12.2199594
Fin de mes en curso	9.25271903
Todo elemento	5.33136816
Media cortada 20%	3.99650145
Información y comunicación	3.23793658
Préstamos brutos (incluyendo arrendamiento) (COP, Mil millones)	1.44807452
Indicador de crédito (COP, Mil millones)	1.44807452

Ilustración 8. Variables que más afectan la dimensión 5

Las componentes principales no solo ofrecen una representación concisa de la información macroeconómica subyacente, sino que también optimizan la estructura de los datos, volviendo el modelo global menos pesado de ejecutar y más robusto en su formulación. La exploración de estas dimensiones latentes fue crucial para obtener una visión agregada de la influencia macroeconómica, complementando el análisis detallado de la influencia de variables individuales que se presentará en secciones posteriores para capturar relaciones específicas con las tarifas de energía.

3.2 Base de datos de tarifas de energía

Dado que el objetivo principal de este análisis es comprender el comportamiento general de las tarifas de energía a nivel nacional, se consideró adecuado utilizar un promedio simple para agrupar las tarifas de todos los comercializadores para cada mes del periodo analizado. Esta metodología permite obtener una serie de tiempo homogénea que refleja las tendencias promedio del mercado. Además, como se evidencia en la ilustración 4. Frecuencia de comercializadores., la base de datos original presenta inconsistencias en la cantidad de datos disponibles para cada comercializador, lo que dificulta un análisis comparativo directo. El cálculo del promedio simple mitiga este problema al generar una serie de tiempo con el mismo número de observaciones para cada mes. Si bien se reconoce que existen diferencias significativas entre las tarifas de los distintos comercializadores y regiones, el uso del promedio simple simplifica el análisis y facilita la identificación de patrones a gran escala. Análisis más detallados a nivel regional pueden realizarse posteriormente utilizando la base de datos original.

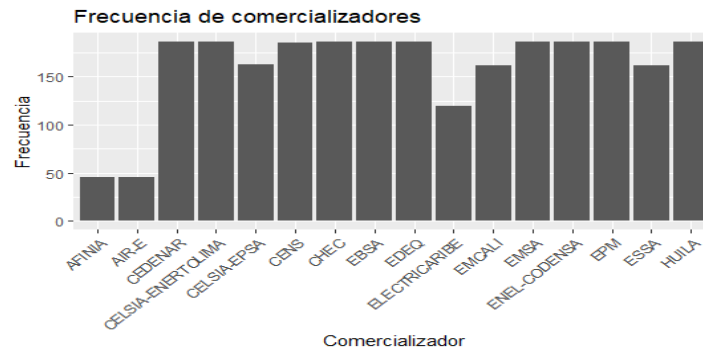


Ilustración 9 Frecuencia de comercializadores

Para una mejor comprensión de los datos, se realizaron histogramas para cada componente tarifario. Estos gráficos revelaron que la mayoría de las distribuciones son unimodales, concentrándose los valores alrededor de un punto central. Sin embargo, se observó cierta asimetría positiva en algunas distribuciones, indicando la presencia de valores atípicos que podrían sesgar el promedio hacia arriba. La dispersión de los datos también varió entre los componentes, siendo algunos más variables que otros. Esta heterogeneidad en la distribución y dispersión sugiere que los componentes con mayor variabilidad y valores promedio más altos tienen una influencia más significativa en el costo unitario total (CUcal).

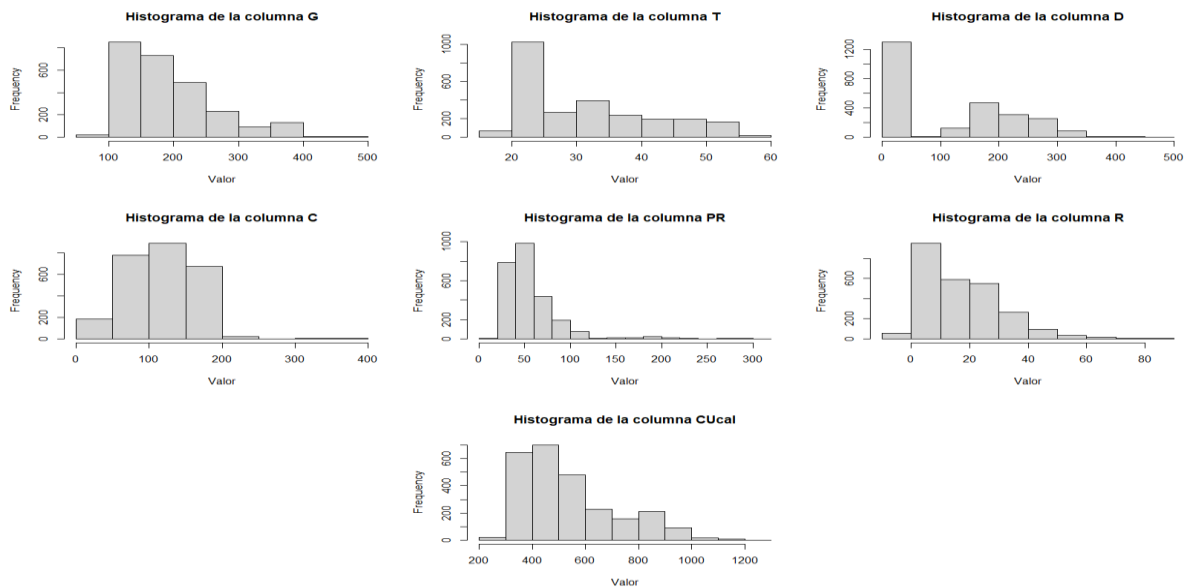


Ilustración 10 Histogramas por componentes

3.2.1. Análisis de varianza y correlación

Para analizar los datos de tarifas, es crucial comprender el significado de los estadísticos descriptivos utilizados. En este caso, el análisis se centra en la varianza, la desviación estándar y la correlación.

- **Varianza**

La varianza mide qué tan dispersos están los valores de un conjunto de datos con respecto a su media. En otras palabras, cuantifica la variabilidad de los datos. Una varianza alta indica que los valores están muy dispersos, mientras que una varianza baja sugiere que los valores tienden a estar cerca de la media.

En el contexto de las tarifas, la varianza permite entender la volatilidad o la inestabilidad de los precios. Una alta varianza en una tarifa específica (por ejemplo, la tarifa 'G') indica que esa tarifa ha fluctuado significativamente a lo largo del tiempo, lo que podría implicar un mayor riesgo o incertidumbre para los consumidores y las empresas. Por el contrario, una baja varianza en otra tarifa (por ejemplo, la tarifa 'T') señala una mayor estabilidad en los precios.

- **Desviación Estándar**

La desviación estándar es la raíz cuadrada de la varianza. Al igual que la varianza, mide la dispersión de los datos con respecto a la media. La principal ventaja de la desviación estándar es que se expresa en las mismas unidades que los datos originales, lo que facilita su interpretación.

La desviación estándar proporciona una medida de la dispersión de las tarifas en términos de las unidades originales (para este caso, pesos por kWh-mes). Esto facilita la comprensión de cuánto varían las tarifas típicas con respecto al promedio. Por ejemplo, una desviación estándar de 70.10 para la tarifa 'G' significa que los precios de 'G' típicamente se desvían en 70.10 unidades de su precio promedio.

- **Análisis de los Resultados**

Variable	Varianza	Desviacion_Estandar
G	4,913.84	70.10
T	108.11	10.40
D	10,227.20	101.13
C	2,181.29	46.70
PR	1,040.39	32.26
R	176.85	13.30
CUcal	31,474.85	177.41

Ilustración 11. Varianza y desviación estándar tarifaria

- **Tarifas G y D:**

Estas tarifas muestran las varianzas y desviaciones estándar más altas, lo que indica que sus precios son más volátiles en comparación con las otras tarifas. La tarifa D es la más volátil de todas, con una varianza de 10,227.20 (\$/kWh-mes)² y una desviación estándar de 101.13 \$/kWh-mes. Esto significa que los precios de la tarifa D tienden a desviarse en promedio 101.13\$/kWh-mes de su valor medio, mostrando una gran dispersión.

- **Tarifa C y PR:**

Estas tarifas muestran una varianza y desviación estándar medias. La tarifa C muestra una varianza de $2,181.29 (\$/kWh\text{-mes})^2$ y una desviación estándar de $46.70 \$/kWh\text{-mes}$. Esto indica una volatilidad moderada en comparación con las tarifas G y D.

- **Tarifas T y R:**

Estas tarifas muestran una varianza y desviación estándar bajas, lo que indica que sus precios son relativamente estables. tarifa T presenta la menor volatilidad con una varianza de $108.11(/kWh\text{-mes})^2$ y una desviación estándar de $10.40 \$/kWh\text{-mes}$, lo que sugiere que sus precios tienden a fluctuar menos.

- **CUcal:**

Tiene la varianza y la desviación estándar más altas, lo que refleja la variabilidad total de la variable que representa. Su desviación estándar de $177.41 \$/kWh\text{-mes}$ da una idea de la magnitud típica de las fluctuaciones de esta variable. Dado que CUcal es la sumatoria de las demás variables, su alta volatilidad puede explicarse por la combinación de las variabilidades de sus componentes. En particular, las tarifas G y D, que presentan las mayores volatilidades individuales, contribuyen significativamente a la variabilidad de CUcal. Es decir, al ser CUcal el resultado de agregar estas tarifas más volátiles, hereda y amplifica estas fluctuaciones, resultando en una mayor varianza y desviación estándar.

3.3. Base de datos final

Para desarrollar este proyecto, se consolidó una base de datos única a partir de dos fuentes distintas. Esta base de datos contiene una serie de tiempo completa que abarca las componentes tarifarias individuales, el cálculo de la tarifa final y los cinco componentes principales obtenidos a través del análisis de componentes principales (PCA). Esta base de datos unificada fue el insumo principal para todos los análisis y modelos desarrollados en este estudio.

El siguiente diagrama de flujo ofrece una representación visual del proceso metodológico seguido para la creación de la base de datos final. Se detallan las distintas etapas, comenzando con las fuentes de datos iniciales (variables macroeconómicas colombianas y tarifas de energía), pasando por las transformaciones y análisis aplicados (como el Análisis de Componentes Principales), hasta la consolidación en una base de datos unificada que sirvió como insumo principal para el desarrollo de este proyecto.

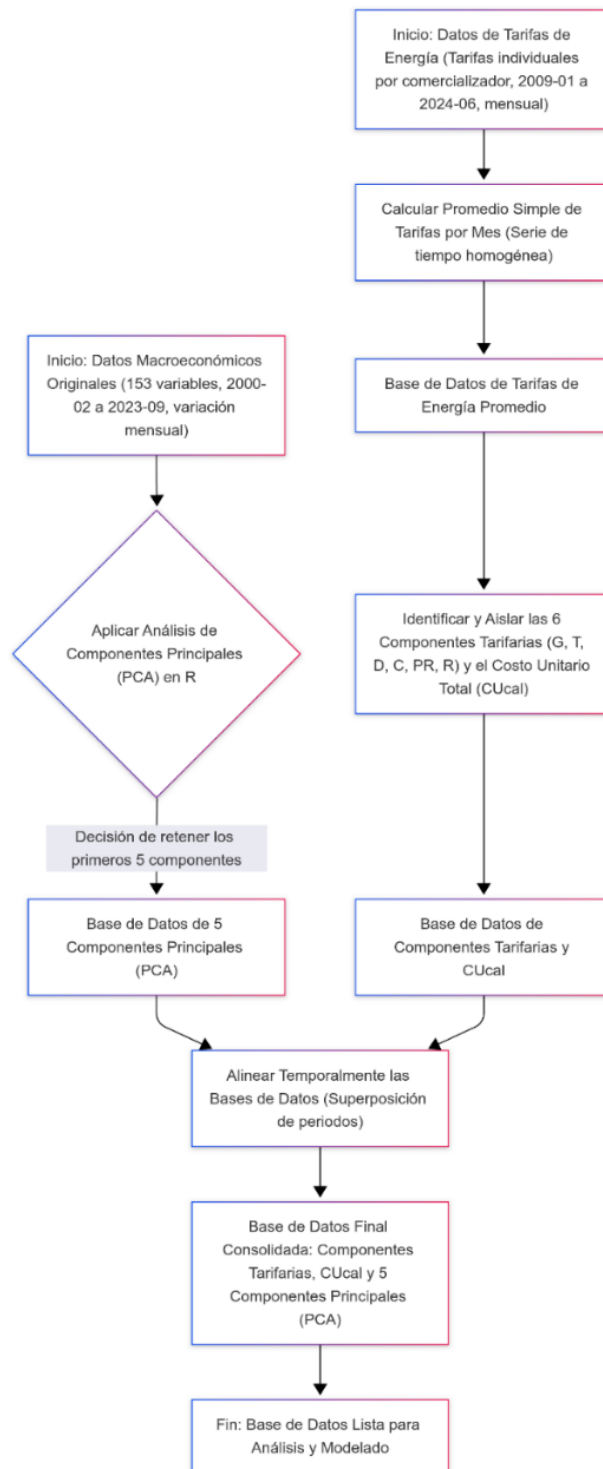


Ilustración 12. Flujo para la creación de la base de datos final

4. DISEÑAR Y ENTRENAR UN MODELO DE VECTORES AUTORREGRESIVOS (VAR) UTILIZANDO LAS TARIFAS DE ENERGÍA ELÉCTRICA Y LOS FACTORES MACROECONÓMICOS IDENTIFICADOS.

Como primera premisa para poder aplicar cualquier modelo VAR, todas las variables incluidas en el modelo deben ser variables estacionarias. Para validar la estacionariedad de las series temporales, se aplicaron la prueba de Augmented Dickey-Fuller (ADF), la prueba de Phillips-Perron (PP) y la prueba Kwiatkowski-Phillips-Schmidt-Shin (KPSS). La prueba ADF evalúa la presencia de una raíz unitaria en una serie temporal, considerando la autocorrelación de los errores mediante la inclusión de rezagos de la variable dependiente. Por otro lado, la prueba PP es una alternativa no paramétrica que corrige la autocorrelación y la heterocedasticidad de los errores sin necesidad de incluir rezagos. A diferencia de las pruebas ADF y PP, que tienen como hipótesis nula la presencia de una raíz unitaria (no estacionariedad), la prueba KPSS tiene como hipótesis nula la estacionariedad alrededor de una tendencia determinista. Es decir, la prueba KPSS verifica si una serie temporal es estacionaria alrededor de una tendencia. Adicionalmente, se analizaron las gráficas de las series temporales para identificar visualmente posibles tendencias o patrones estacionales. Finalmente, las variables que no superaron las pruebas de estacionariedad fueron transformadas mediante la aplicación del logaritmo para estabilizar su varianza.

4.1. Prueba de Augmented Dickey-Fuller (ADF)

Variable Dimensión 5: Es la única variable que ha demostrado ser estacionaria. El valor p muy bajo (menor a 0.01) indica una fuerte evidencia en contra de la hipótesis nula de que la serie tiene una raíz unitaria. Esto significa que la dimensión 5 tiende a regresar a una media constante a largo plazo.

Resto de las variables: Las demás variables presentan valores p mayores a 0.05, lo que implica que no podemos rechazar la hipótesis nula. Por lo tanto, estas series se consideran no estacionarias. Esto significa que su media, varianza o autocovarianza cambian a lo largo del tiempo.

4.2. Prueba de Phillips-Perron (PP)

La prueba de Phillips-Perron es una alternativa a la prueba ADF, especialmente útil cuando hay dudas sobre la distribución de los errores. La interpretación de los resultados es similar a la prueba ADF

- Variables estacionarias:

Variable T

Variable R

Variable Dimensión 5

Estas variables tienen valores p muy pequeños (menores a 0.01), lo que indica una fuerte evidencia en contra de la hipótesis nula. Por lo tanto, podemos considerarlas estacionarias.

- Variables no estacionarias:

El resto de las variables tienen valores p mayores a 0.05, lo que sugiere que no podemos rechazar la hipótesis nula. Por lo tanto, estas series se consideran no estacionarias.

Hipótesis nula: La serie tiene una raíz unitaria (no es estacionaria).

Hipótesis alternativa: La serie es estacionaria.

4.3. Prueba Kwiatkowski-Phillips-Schmidt-Shin (KPSS)

Es una prueba no paramétrica que se utiliza para evaluar la estacionariedad de una serie de tiempo. A diferencia de la prueba ADF y PP, la hipótesis nula de la prueba KPSS es que la serie es estacionaria.

Se aplica en las variables T y R que fueron las que salieron diferentes en las 2 primeras pruebas:

En ambas pruebas, tanto para la variable T como para la R, el valor p es significativamente menor a 0.01. Esto significa que **se rechaza la hipótesis nula de estacionariedad. En otras palabras, ambas series son NO estacionarias.**

Los resultados anteriores y el análisis del análisis gráfico presentado a continuación revelan que ninguna de las series de tarifas de energía eléctrica exhibe estacionariedad. Esta falta de estacionariedad representa un desafío significativo para la estimación y aplicación del modelo VAR. La no estacionariedad puede generar problemas de multicolinealidad, raíces unitarias y modelos espurios, lo que compromete la validez de las predicciones y la interpretación de los coeficientes. Por lo tanto, antes de proceder con la estimación del VAR, será necesario transformar las series para garantizar su estacionariedad, a través de técnicas como la diferenciación o la transformación logarítmica.

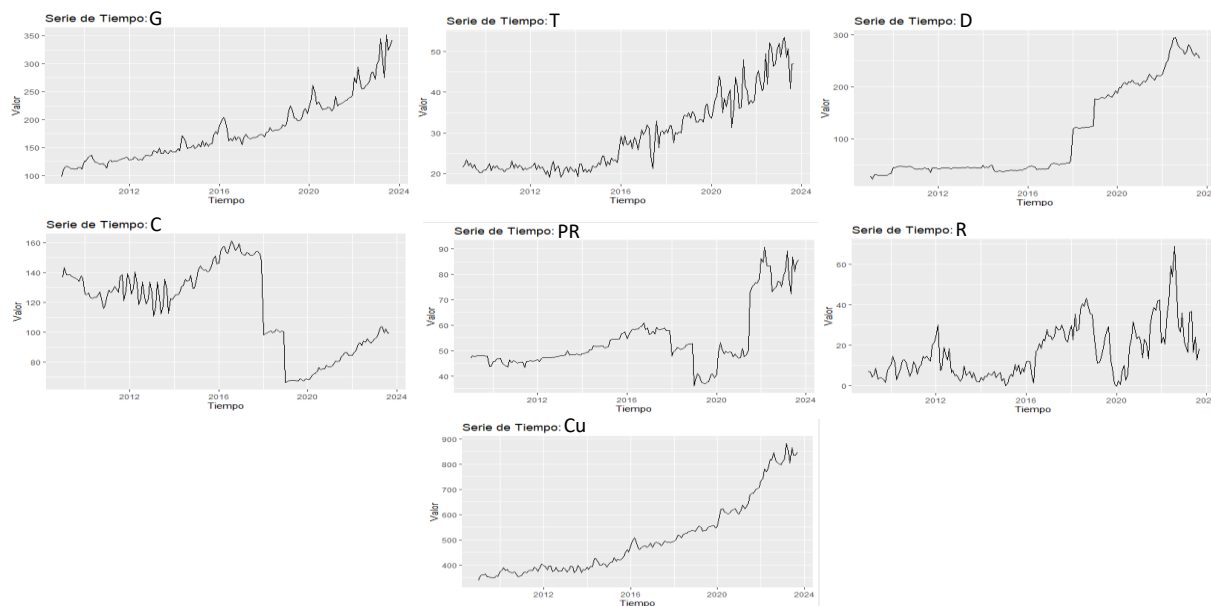


Ilustración 13 Series de tiempo no estacionarias

4.4. Convertir las series en estacionarias

Para cumplir con el supuesto de estacionariedad, necesario para la estimación de modelos VAR, se aplicó la transformación logarítmica a todas las series de las variables no estacionarias encontradas, siendo la dimensión 5 la única variable estacionaria confirmada. Esta transformación es ampliamente utilizada en

econometría para estabilizar la varianza y eliminar tendencias determinísticas. La estacionariedad de las series es fundamental para garantizar la validez de los resultados inferenciales y la capacidad predictiva del modelo. Los gráficos presentados a continuación muestran que la transformación logarítmica ha sido exitosa en lograr la estacionariedad de las series.

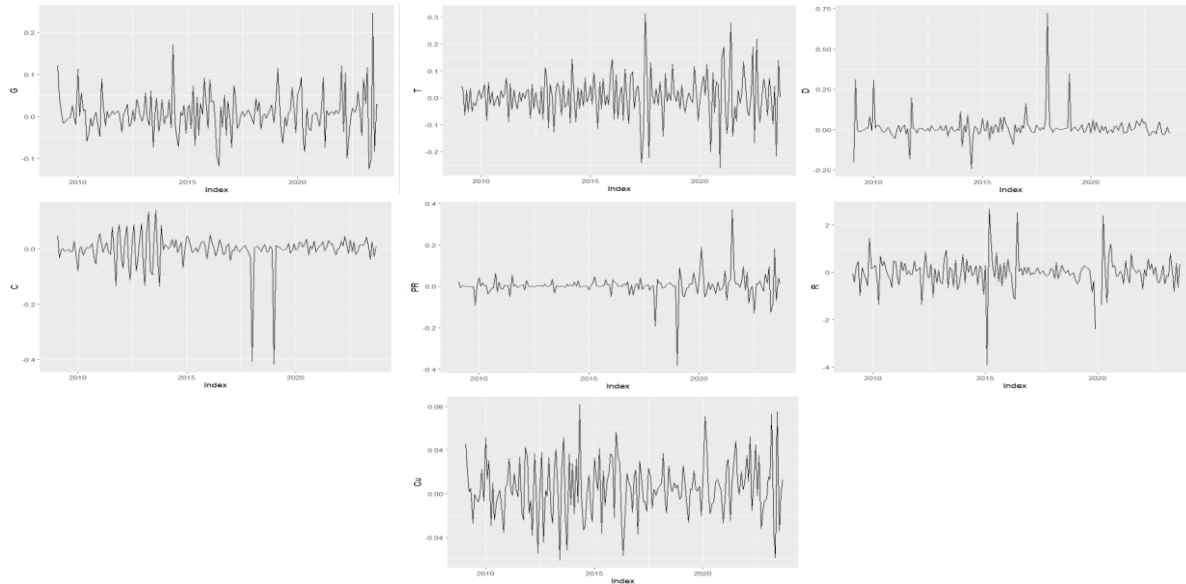


Ilustración 14 Series de componentes estacionarios

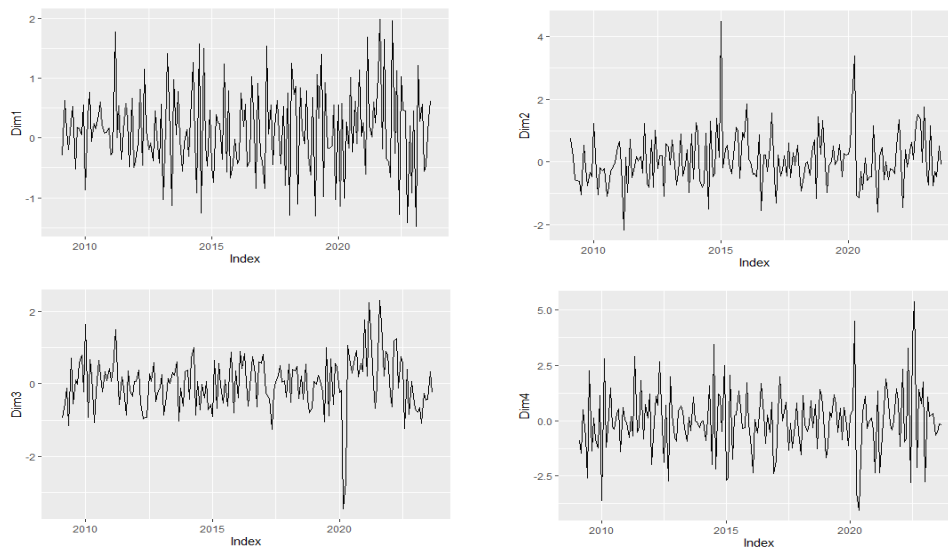


Ilustración 15 Series variables macroeconómicas estacionarias

4.5. Modelo de Vector AutoRegresivo (VAR)

Con el fin de analizar los choques producidos por los factores macroeconómicos en las componentes de la tarifa de energía, se implementó un modelo VAR (Vector AutoRegresivo). Este modelo estadístico permite examinar las interrelaciones dinámicas entre múltiples variables a lo largo del tiempo, proporcionando información valiosa sobre cómo las perturbaciones en una variable se propagan a través del sistema.

Un modelo VAR es una herramienta versátil que permite realizar pronósticos multivariados, anticipando el comportamiento futuro de múltiples variables de forma conjunta. Además, facilita el análisis de causalidad, revelando cómo las variables se influyen entre sí a lo largo del tiempo. También ofrece la posibilidad de analizar los impulsos de respuesta, examinando el impacto de una perturbación en una variable sobre el resto del sistema. Finalmente, permite descomponer la varianza de cada variable, identificando la contribución de las innovaciones de todas las variables del modelo.

En su forma reducida, un modelo VAR se puede expresar de la siguiente manera:

$$Y_t = A_0 + \sum_{i=1}^{\rho} A_i Y_{t-i} + \varepsilon_t$$

Ilustración 16. Expresión del modelo VAR

[26]

En donde:

- Y_t : Es un vector que contiene los valores de las K variables endógenas en el periodo t.
- A_0 : Es un vector de constantes.
- A_i : Son matrices de coeficientes que capturan el efecto de los rezagos de las variables endógenas sobre las variables actuales.
- Y_{t-i} : Son los valores de las variables endógenas en periodos anteriores (rezagos).
- ε_t : Es un vector de términos de error o innovaciones, que se asume que siguen una distribución normal con media cero y matriz de varianza-covarianza Σ .

Esta ecuación indica que el valor actual de cada variable endógena (Y_t) depende de sus propios valores pasados (Y_{t-i}) y de los valores pasados de las otras variables endógenas, además de un término de error. Los coeficientes A_i capturan la magnitud y el signo de la relación entre las variables.

5. EVALUAR EL DESEMPEÑO DEL MODELO VAR EN TÉRMINOS DE PRECISIÓN Y CONFIABILIDAD.

Una vez aplicado el modelo VAR para examinar las dinámicas entre las variables macroeconómicas y las componentes de la tarifa, se analizan los impulsos de respuesta generados por cada una de las cinco dimensiones consideradas. Este análisis permitirá evaluar la significancia y la magnitud de los efectos de cada dimensión sobre las variables de interés, específicamente cómo los cambios en los indicadores de los factores macroeconómicos generan impulsos de respuesta en las componentes de la tarifa. Para cada dimensión, se examinan los gráficos de impulsos de respuesta, prestando especial atención a los intervalos de confianza del 95% para determinar la significancia estadística de las respuestas. Un impulso se considera estadísticamente significativo cuando su línea de respuesta sale de los intervalos de confianza, lo que indica que el efecto observado no es producto del azar (es decir, el intervalo de confianza no contiene el valor cero). Además, se contextualizan los hallazgos dentro del marco del funcionamiento del mercado, explicando las razones detrás de la presencia o ausencia de impulsos significativos y su relevancia para la dinámica de las tarifas.

5.1. Dimensiones

5.1.1. Dimensión 1

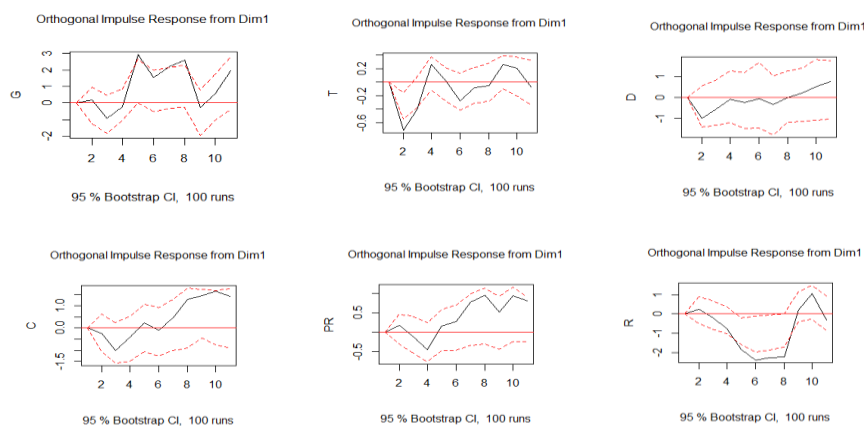


Ilustración 17. Impulsos de respuesta: Dimensión 1

El análisis de los impulsos generados por la Dimensión 1 (Factor Monetario) a través del modelo VAR revela patrones diversos en las respuestas de las variables G, T, D, C, PR y R. En términos generales, se observa que las respuestas iniciales propensas a ser negativas, seguidas de fluctuaciones y, en algunos casos, recuperaciones graduales. Sin embargo, al examinar los intervalos de confianza del 95%, se constata que en las variables G, D, C y PR, estos intervalos siempre incluyen el valor cero, lo que indica que las respuestas no son estadísticamente significativas. No obstante, en la variable T se aprecia un impulso significativamente negativo en el periodo 2, ya que su intervalo de confianza no incluye el cero. En el caso de la variable R, se aprecia un impulso negativo en el sexto período (mes), pero su magnitud es pequeña y su aparición tardía disminuye su relevancia. Por lo tanto, se puede concluir que, con la excepción de un impulso significativo en la variable T en el periodo 2, la Dimensión 1 no genera un impacto significativo en la mayoría de los componentes de la tarifa, ya que las respuestas observadas carecen de la robustez estadística necesaria para ser consideradas significativas.

5.1.2. Dimensión 2

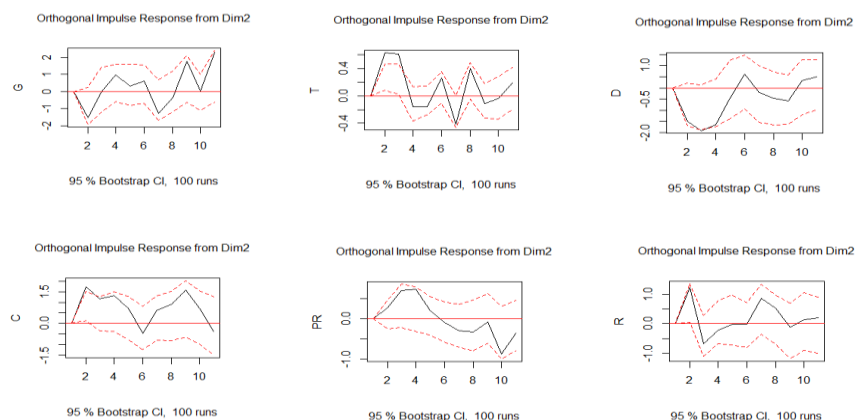


Ilustración 18. Impulsos de respuesta: Dimensión 2

El análisis de los impulsos generados por la Dimensión 2 (Factor de Exposición Global) revela patrones diversos en las respuestas de las variables. En general, se observa que las variables G, D, C, PR y R muestran fluctuaciones alrededor de cero, y sus intervalos de confianza siempre incluyen el valor cero, lo que indica que las respuestas no son estadísticamente significativas. Sin embargo, la variable T muestra un patrón diferente. Inicialmente, se observa un impacto positivo significativo, lo que sugiere una respuesta inmediata a la Dimensión 2. No obstante, a medida que avanzan los períodos, la variable T experimenta fluctuaciones considerables, alternando entre valores positivos y negativos. Esta volatilidad sugiere que, aunque la Dimensión 2 puede generar una respuesta inicial en T, el efecto no es sostenido y se disipa con el tiempo, dando lugar a una dinámica fluctuante. Por lo tanto, mientras que la Dimensión 2 no parece tener un impacto significativo en la mayoría de las componentes tarifarias, sí muestra una influencia inicial en la variable T, aunque esta influencia no se mantiene estable a lo largo del tiempo.

5.1.3. Dimensión 3

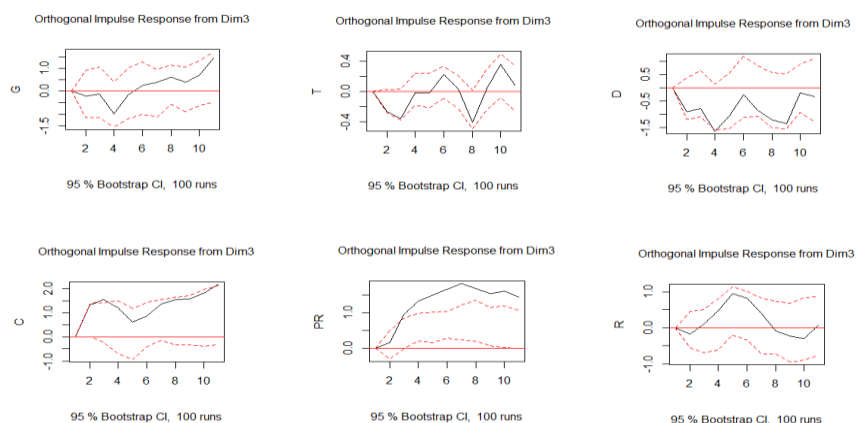


Ilustración 19. Impulsos de respuesta: Dimensión 3

El análisis de los impulsos generados por la Dimensión 3 (Factor de Apertura comercial) revela patrones mixtos en las respuestas de las variables. En general, se observa que las variables G, T, D y C muestran fluctuaciones alrededor de cero, y sus intervalos de confianza incluyen el valor cero, lo que indica que las respuestas no son estadísticamente significativas. Sin embargo, la variable PR muestra una clara tendencia positiva que supera los intervalos de confianza, lo que sugiere una respuesta significativa a la Dimensión 3. En particular, la variable PR muestra una tendencia positiva constante desde el primer período, lo que indica un impacto inmediato y sostenido de la Dimensión 3. En cuanto a la variable R, aunque se observa una tendencia positiva en el corto plazo, esta se revierte posteriormente, y los intervalos de confianza incluyen el cero, lo que disminuye su significancia estadística. Cuando la línea del modelo supera los intervalos de confianza, esto indica que la respuesta es estadísticamente significativa y no se debe al azar, lo que refuerza la validez del impacto observado. Por lo tanto, se puede concluir que la Dimensión 3 tiene un impacto significativo en la variable PR, mientras que su influencia en G, T, D, C y R es limitada o no significativa.

5.1.4. Dimensión 4

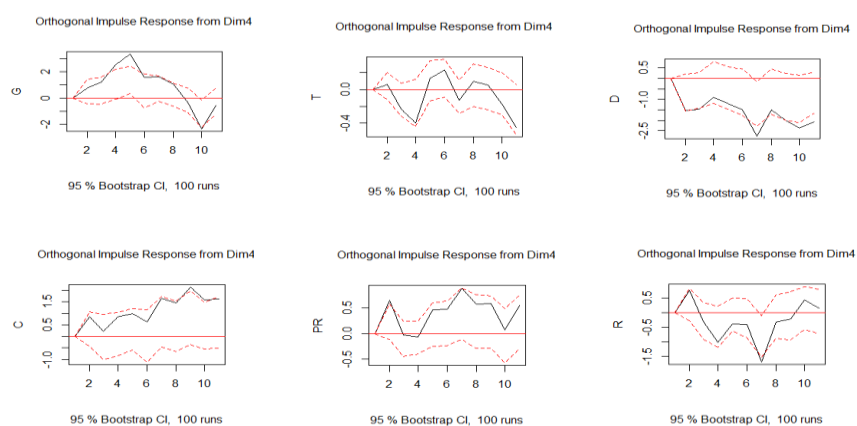


Ilustración 20. Impulsos de respuesta: Dimensión 4

El análisis de los impulsos generados por la Dimensión 4 (Factor de Competitividad Externa) revela patrones mayormente no significativos en las respuestas de las componentes tarifarias. En las variables G y R, se observa una respuesta inicialmente positiva, seguida de una reversión a valores negativos. Sin embargo, los intervalos de confianza solo excluyen el cero en el décimo y séptimo período, respectivamente, lo que disminuye la significancia de estas respuestas. De manera similar, la variable D muestra una respuesta negativa marcada y constante, pero el intervalo de confianza solo excluye el cero en el séptimo período, lo que también reduce su significancia. Dado que la mayoría de los intervalos de confianza incluyen el cero en la mayoría de los períodos para todas las variables (G, T, D, C, PR, R), se concluye que la Dimensión 4 no tiene un impacto significativo en las componentes de la tarifa. Las fluctuaciones observadas parecen ser aleatorias y no indican una tendencia clara o un impacto sostenido.

5.1.5. Dimensión 5

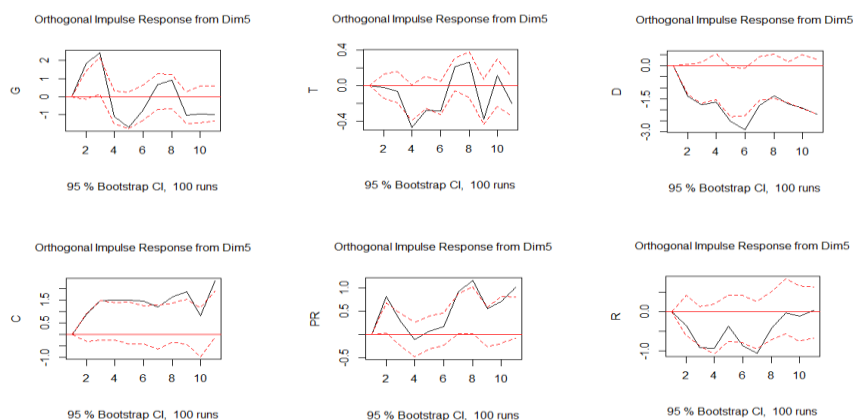


Ilustración 21. Impulsos de respuesta: Dimensión 5

El análisis de los impulsos generados por la Dimensión 5 (Factor del Sector Inmobiliario y de construcción) muestra que, para la mayoría de las componentes tarifarias (G, T, C, PR, R) y en todos los períodos, los intervalos de confianza incluyen el valor cero. Esto indica que las respuestas a los impulsos de la Dimensión 5 no son estadísticamente significativas para estas variables. Sin embargo, la variable D presenta un patrón diferente. Se observa una tendencia negativa marcada que supera los intervalos de confianza en el sexto período. Además, para este mismo período, los intervalos de confianza no incluyen el cero, lo que refuerza la significancia de esta respuesta. Por lo tanto, se concluye que la Dimensión 5 tiene un impacto significativo en la variable D en el sexto período, mientras que su influencia en las demás componentes de la tarifa es limitada o no significativa.

5.1.6. Impulso en la tarifa total calculada (Cu)

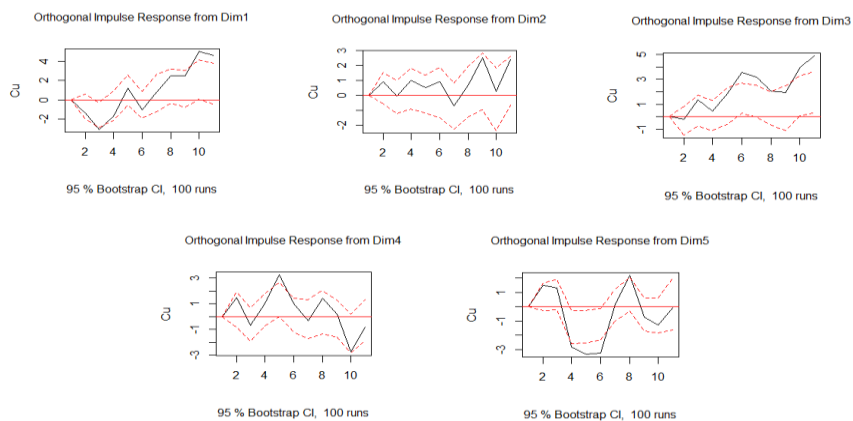


Ilustración 22. Impulsos de respuesta: Todas las dimensiones en el CU

El análisis de los impulsos de respuesta para la tarifa de energía unitaria (CU) en las cinco dimensiones revela patrones variados. La Dimensión 3 muestra un impacto positivo significativo, aunque los intervalos de confianza solo excluyen el cero en los períodos 6 y 10, lo que aún sugiere una respuesta relevante. La

Dimensión 5 muestra un impacto significativo en el sexto período, donde la respuesta supera los intervalos de confianza y estos no incluyen el cero, indicando un efecto puntual pero sustancial; sin embargo, el impulso es muy variado, alternando entre valores positivos y negativos. En contraste, las Dimensiones 1, 2 y 4 no muestran impactos significativos en la tarifa de energía unitaria (CU), ya que sus respuestas fluctúan alrededor de cero y los intervalos de confianza incluyen el cero en la mayoría de los períodos. Por lo tanto, se concluye que la Dimensión 3 tiene un efecto positivo en la tarifa de energía unitaria, la Dimensión 5 tiene un efecto significativo puntual en el sexto período, y las otras dimensiones no tienen un impacto significativo.

5.2. Variables Macroeconómicas Individuales

Continuando con el análisis de los resultados obtenidos a partir del modelo VAR y la identificación de patrones de impulso en las variables tarifarias ante los factores macroeconómicos latentes (dimensiones), se procedió a realizar un análisis complementario utilizando un conjunto específico de variables macroeconómicas. Esta decisión se fundamenta en la necesidad de profundizar la comprensión de los mecanismos de transmisión de los choques macroeconómicos a las tarifas del sector energético, dada la relevancia de estas últimas para la economía y los consumidores. Las variables macroeconómicas seleccionadas para este análisis específico, presentadas en la Ilustración 16. Variables Macroeconómicas Individuales, abarcan diversos aspectos de la economía colombiana, incluyendo indicadores fiscales, precios al productor, empleo, tipo de cambio, crédito, deuda externa y población. La elección de estas variables se basa en su reconocida influencia potencial sobre los costos y la dinámica del sector energético. Además, y reconociendo la naturaleza intrínsecamente ligada de la oferta y la demanda en el sector energético, se incorporó la variable "Oferta y demanda (Demanda, Generación Hidráulica y Gen Fósil)". Si bien esta variable describe una dinámica propia del sector, su comportamiento agregado refleja condiciones macroeconómicas subyacentes y ejerce una influencia directa en la formación de precios y, por ende, en las tarifas. Al incluirla como una variable macroeconómica específica, se busca capturar de manera más directa los efectos de los desequilibrios entre la oferta y la demanda agregada de energía en los componentes tarifarios, complementando la perspectiva ofrecida por los factores macroeconómicos latentes identificados previamente.

COD	V1	V2	Fuente
X2	Recibos fiscales	Impuesto nacional	DIAN
X127	Precios al productor (precios de producción)	Nacional	Departamento Administrativo Nacional de Estadística (DANE)
X128	Precios al productor (precios de producción)	Interno	Departamento Administrativo Nacional de Estadística (DANE)
X141	Empleo (encuesta de fuerza laboral)	Abastecimiento de electricidad, gas y agua	Departamento Administrativo Nacional de Estadística (DANE)
X320	Tipo de cambio (COP)	Fin de mes en curso	Banco Central de la República de Colombia (SENC)
X335	Crédito a negocios en circulación	Sectores	Fedesarrollo
X497	Deuda externa	Deuda externa (USD, Millones)	Banco Central de la República de Colombia (SENC)
X538	Población	Población (Miles)	Departamento Administrativo Nacional de Estadística (DANE)
DEM	Oferta y demanda	Demanda, Generación Hidraulica y Gen Fossil	XM

Ilustración 23. Variables macroeconómicas individuales

5.2.1. Recibos Fiscales

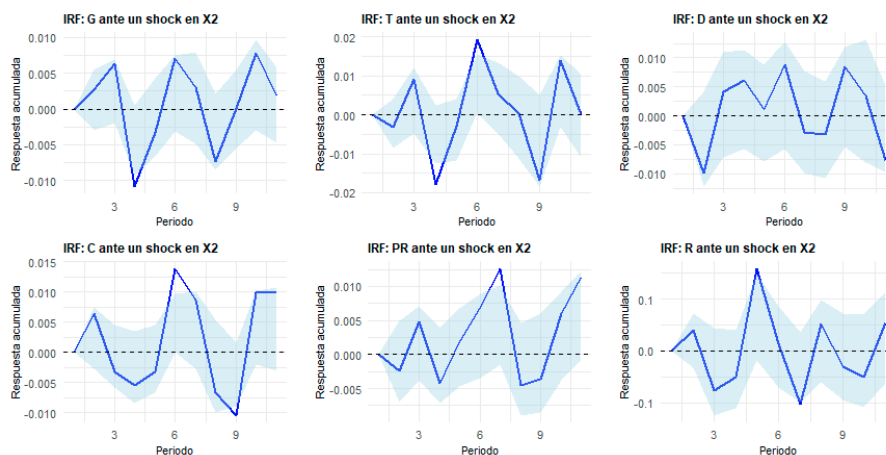


Ilustración 24. Impulsos de respuesta: Recibos fiscales

El análisis de los impulsos generados por la variable macroeconómica "Recibos Fiscales" (X2) revela que, para todas las componentes tarifarias analizadas (G, T, D, C, PR, R), las respuestas a lo largo del horizonte temporal presentan intervalos de confianza que consistentemente incluyen el valor cero. Esta observación generalizada indica que, estadísticamente, un choque en los Recibos Fiscales no genera un impacto significativo en ninguno de los componentes de la tarifa de energía. Si bien se observan algunas fluctuaciones en las respuestas de las variables, la inclusión del cero dentro de los intervalos de confianza sugiere que estas variaciones no son lo suficientemente robustas como para ser consideradas significativas desde un punto de vista estadístico. Por lo tanto, se concluye que, en el período analizado, la variable macroeconómica "Recibos Fiscales" (X2) no ejerce una influencia significativa en la dinámica de los componentes de la tarifa de energía.

5.2.2. Precios al productor Nacional

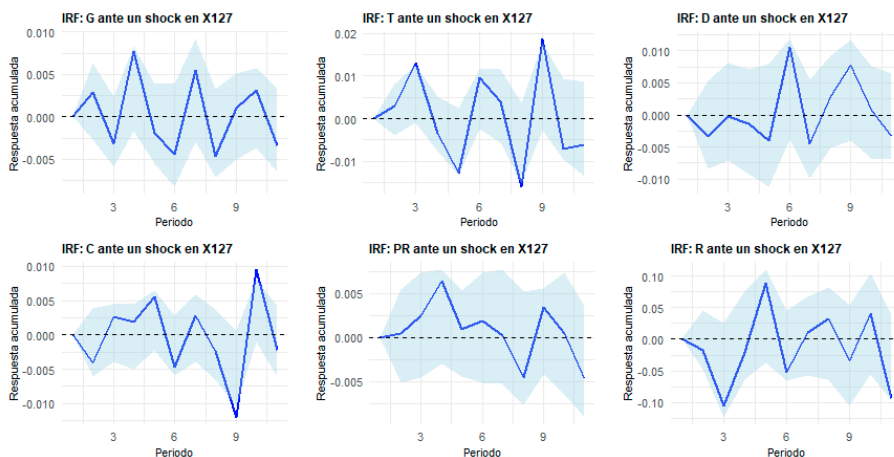


Ilustración 25. Impulsos de respuesta: Precios al productor nacional

El análisis de los impulsos generados por la variable macroeconómica "Precios al productor Nacional" (X127) revela que, para la totalidad de las componentes tarifarias consideradas (G, T, D, C, PR, R), las respuestas a lo largo del horizonte temporal muestran intervalos de confianza que consistentemente contienen el valor cero. Este patrón uniforme sugiere que, desde una perspectiva estadística, un choque en los Precios al productor Nacional no induce un impacto significativo en ninguno de los componentes de la tarifa de energía.

5.2.3. Precios al productor Interno

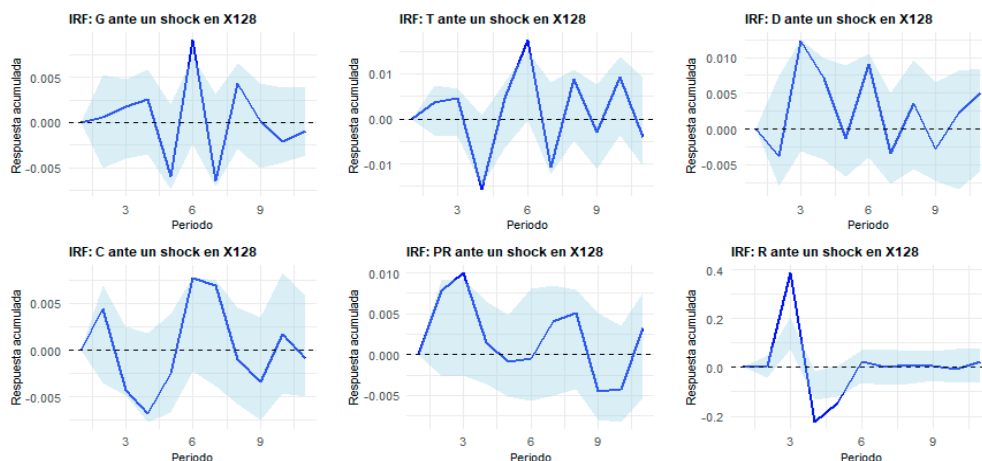


Ilustración 26. Impulsos de respuesta: Precios al productor interno

De acuerdo con el análisis de los impulsos generados por la variable macroeconómica "Precios al productor Interno" (X128), se observa un patrón general donde los intervalos de confianza para la mayoría de las componentes tarifarias (G, T, D, C, PR) incluyen el valor cero a lo largo del horizonte temporal. Sin embargo, la respuesta de la variable R presenta una excepción notable. A partir del período 3 (mes), se evidencia que los intervalos de confianza no incluyen el valor cero, lo que indica un impacto estadísticamente significativo de un choque en los Precios al productor Interno sobre esta componente tarifaria.} Se observa un pico

pronunciado en la respuesta de R alrededor de este período, seguido de una disminución y estabilización posterior.

Por lo tanto, se concluye que, si bien la variable macroeconómica "Precios al productor Interno" (X128) no genera un impacto significativo en la mayoría de los componentes de la tarifa de energía, sí ejerce una influencia estadísticamente significativa en la variable R a partir del período 3. Este hallazgo sugiere una sensibilidad particular del componente R a las variaciones en los Precios al productor Interno.

5.2.4. Empleo en el sector de abastecimiento de electricidad, gas y agua

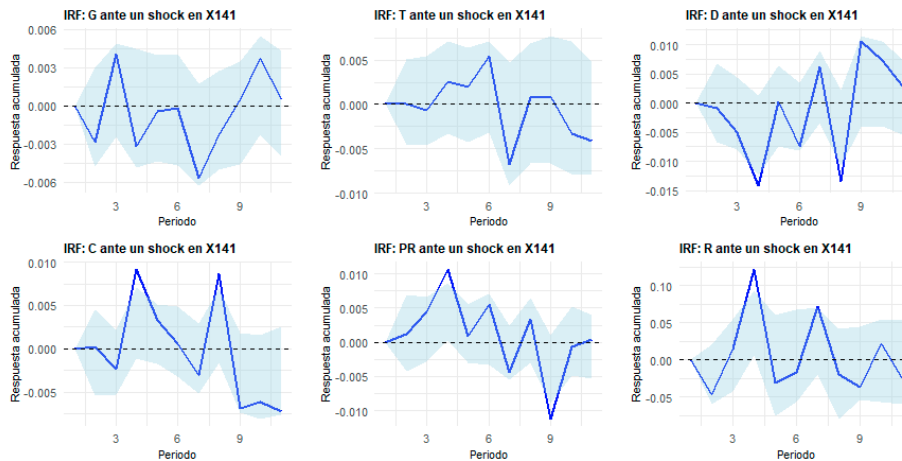


Ilustración 27. Impulsos de respuesta: Empleo en el sector eléctrico, gas y agua

El análisis de los impulsos generados por la variable macroeconómica "Empleo en el sector de abastecimiento de electricidad, gas y agua" (X141) revela que, para la totalidad de las componentes tarifarias consideradas (G, T, D, C, PR, R), las respuestas a lo largo del horizonte temporal muestran intervalos de confianza que consistentemente contienen el valor cero. Este patrón uniforme indica que, desde una perspectiva estadística, un choque en el nivel de empleo dentro del propio sector energético no induce un impacto significativo en ninguno de los componentes de la tarifa de energía.

5.2.5. Tipo de cambio (COP)

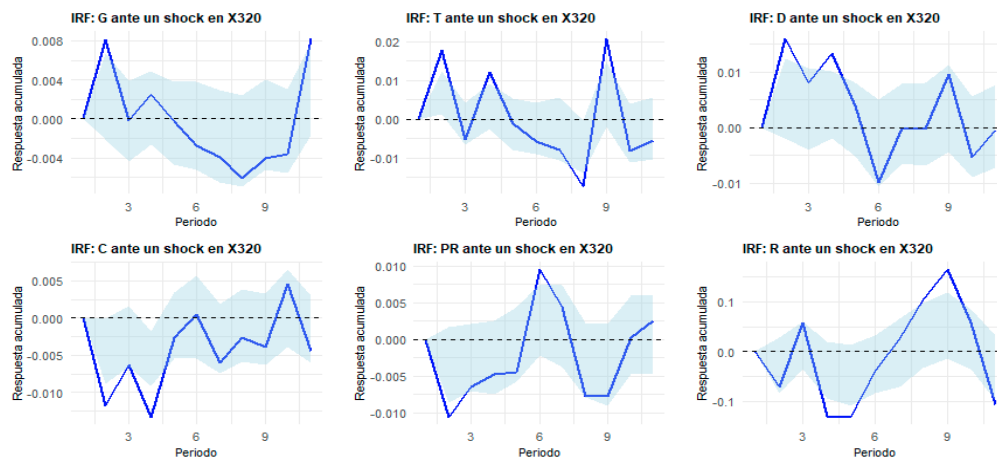


Ilustración 28. Impulsos de respuesta: Tipo de cambio (COP)

El análisis de los impulsos generados por la variable macroeconómica "Tipo de cambio (COP)" (X320) revela que, para la mayoría de las componentes tarifarias (G, D, C, PR, R), los intervalos de confianza incluyen el valor cero a lo largo del horizonte temporal. Esto sugiere que, en general, un choque en el Tipo de cambio (COP) no genera un impacto estadísticamente significativo en estas variables. Sin embargo, la respuesta de la variable T presenta una excepción específica en el período 2. En este período particular, se evidencia que el intervalo de confianza no incluye el valor cero, lo que indica un impacto estadísticamente significativo de un choque en el Tipo de cambio (COP) sobre este componente tarifario en ese momento. El hecho de que este impacto significativo se presente en el período 2 sugiere que la respuesta de la variable T al choque del Tipo de cambio es casi inmediato. Para los demás períodos analizados, el intervalo de confianza de la respuesta de T sí incluye el cero, sugiriendo una falta de significancia estadística en esos momentos.

5.2.6. Crédito a negocios en circulación

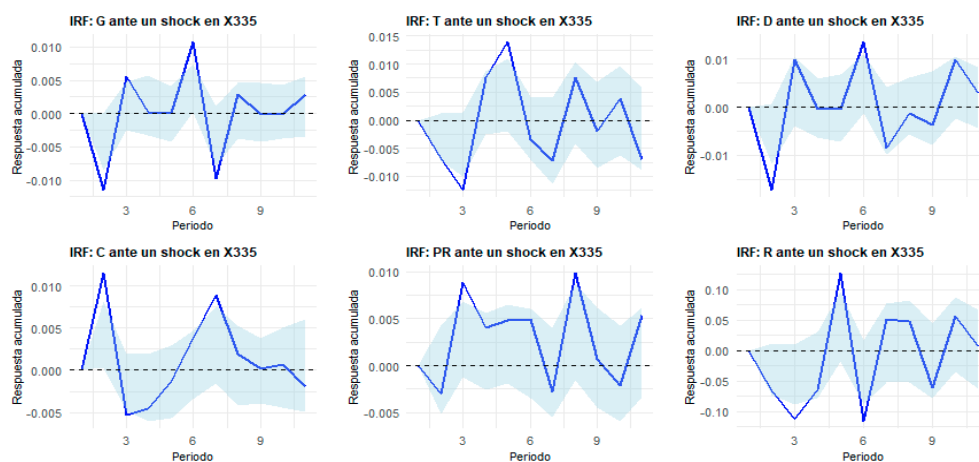


Ilustración 29. Impulsos de respuesta: Crédito a negocios en circulación

El análisis de los impulsos generados por la variable macroeconómica "Crédito a negocios en circulación" (X335) revela que, para la mayoría de las componentes tarifarias (G, T, D, PR, R), los intervalos de confianza incluyen el valor cero a lo largo del horizonte temporal. Esto sugiere que, en general, un choque en el Crédito a negocios en circulación no genera un impacto estadísticamente significativo en estas variables. Sin embargo, la respuesta de la variable C presenta una excepción notable en el período 2. Esto a período particular, se evidencia que el intervalo de confianza no incluye el valor cero, lo que indica un impacto estadísticamente significativo de un choque en el Crédito a negocios en circulación sobre este componente tarifario en ese momento. Este impacto significativo en el período 2 sugiere una respuesta casi inmediata de la variable C al choque del crédito. Posteriormente, a partir del período 3, la respuesta de C tiende a fluctuar dentro de los intervalos de confianza que incluyen el cero, lo que indica una estabilización del efecto.

5.2.7. Deuda externa

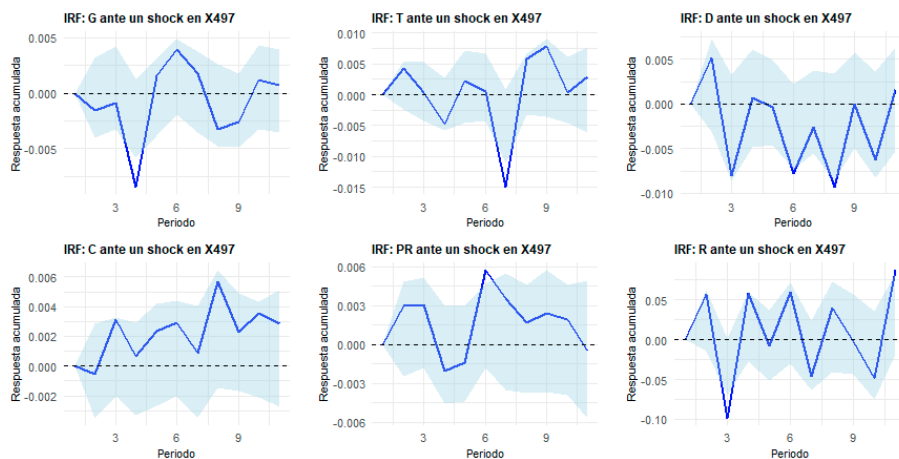


Ilustración 30. Impulsos de respuesta: Deuda externa

El análisis de los impulsos generados por la variable macroeconómica "Deuda externa" (X497) revela que, para la totalidad de las componentes tarifarias consideradas (G, T, D, C, PR, R), las respuestas a lo largo del horizonte temporal muestran intervalos de confianza que consistentemente contienen el valor cero. Este patrón uniforme indica que, desde una perspectiva estadística, un choque en la Deuda externa no induce un impacto significativo en ninguno de los componentes de la tarifa de energía. Si bien se pueden apreciar ciertas fluctuaciones en las líneas de respuesta de las variables, la superposición del cero dentro de las bandas de confianza sugiere que estas variaciones no son lo suficientemente robustas como para descartar la posibilidad de que sean producto del azar.

5.2.8. Población

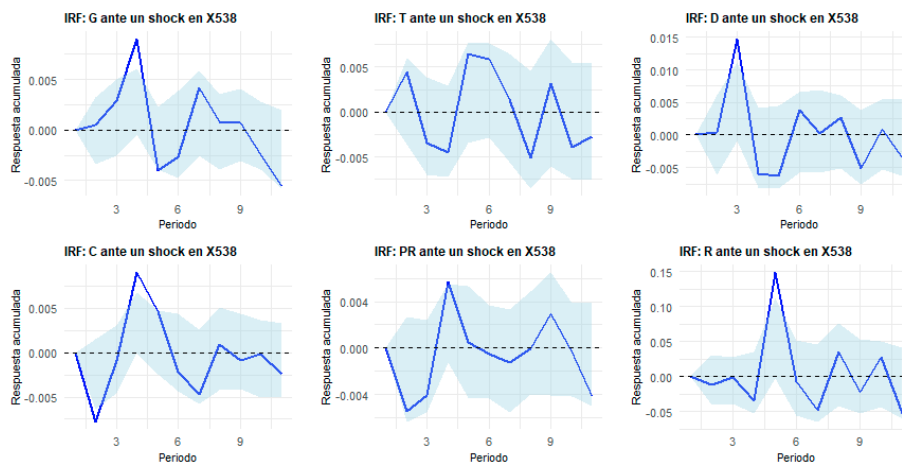


Ilustración 31. Impulsos de respuesta: Población

El análisis de los impulsos generados por la variable macroeconómica "Población" (X538) revela que, para la totalidad de las componentes tarifarias consideradas (G, T, D, C, PR, R), las respuestas a lo largo del horizonte temporal muestran intervalos de confianza que consistentemente contienen el valor cero. Si bien se observan algunas fluctuaciones en las líneas de respuesta de variables como D, C, PR y R, donde en

ciertos períodos (alrededor de los períodos 3, 4 y 5) los intervalos de confianza se acercan al cero, en ningún caso llegan a excluirlo completamente. Esta inclusión del cero dentro de las bandas de confianza indica que no se puede descartar la posibilidad de que estas variaciones sean producto del azar y, por lo tanto, no representan un impacto estadísticamente significativo de un choque en la Población sobre dichos componentes.

5.2.9. Demanda de energía

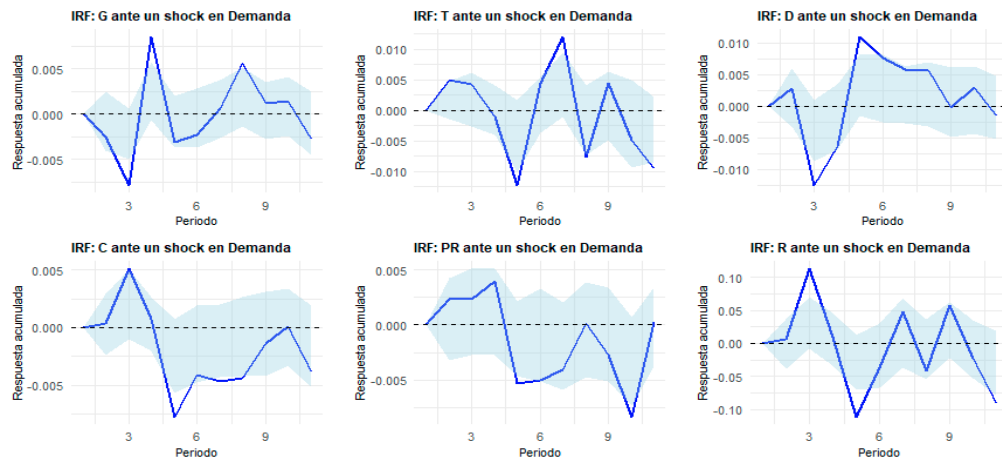


Ilustración 32. Impulsos de respuesta: Demanda de energía

El análisis de los impulsos generados por la variable "Demanda de energía" revela que, para la mayoría de los componentes tarifarios (G, T, C, PR, R), los intervalos de confianza incluyen el valor cero a lo largo del horizonte temporal. Esto sugiere que, en general, un choque en la demanda de energía no genera un impacto estadísticamente significativo en estas variables. Sin embargo, la respuesta de la variable D presenta una excepción notable en el período 3.

5.2.10. Generación hidráulica

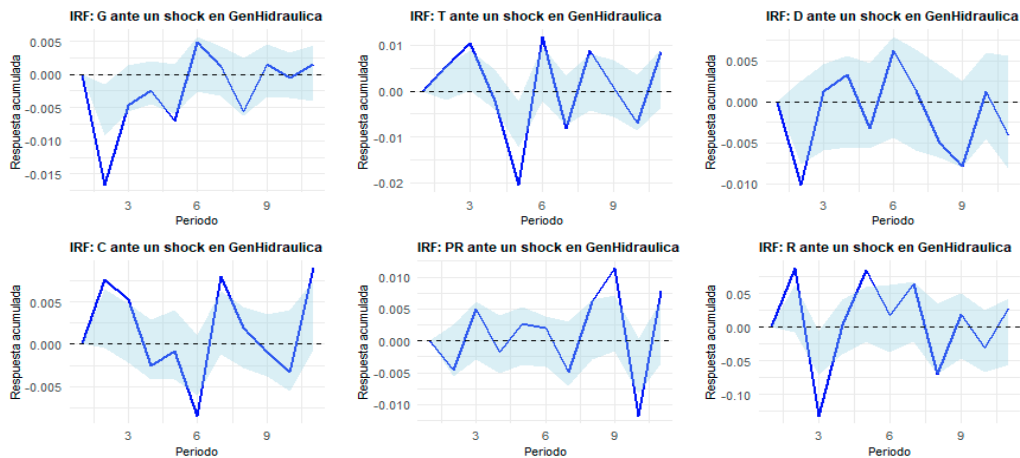


Ilustración 33. Impulsos de respuesta: Generación hidráulica

El análisis de los impulsos generados por la variable "Generación hidráulica" revela que, para la mayoría de las componentes tarifarias (T, D, C, PR), los intervalos de confianza incluyen el valor cero a lo largo del horizonte temporal. Esto sugiere que, en general, un choque en la generación hidráulica no genera un impacto estadísticamente significativo en estas variables. Sin embargo, las respuestas de las variables G y R presentan excepciones notables a partir del período 2. En ambas variables, se observa que los intervalos de confianza no incluyen el valor cero desde este punto inicial, lo que indica un impacto estadísticamente significativo de un choque en la generación hidráulica sobre estos componentes tarifarios de manera relativamente temprana y hasta alrededor del período 6, a partir del cual las respuestas tienden a estabilizarse.

5.2.11. Generación Fósil

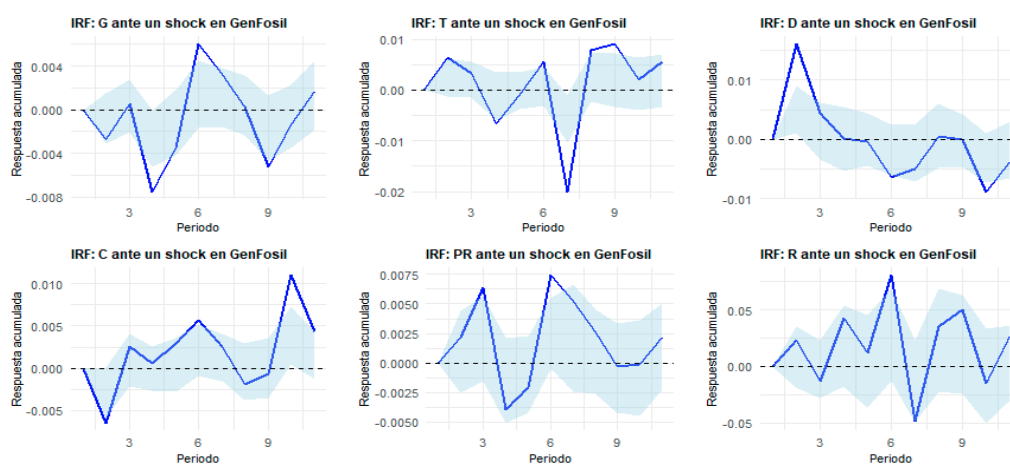


Ilustración 34. Impulsos de respuesta: Generación fósil

El análisis de los impulsos generados por la variable "Generación Fósil" revela impactos significativos y diferenciados en algunos componentes tarifarios. La variable G presenta un impacto estadísticamente significativo en el período 4, aunque sin una clara tendencia a la estabilización al final del análisis. La variable D exhibe un impacto significativo desde el primer período, con una respuesta inicial alza que se limita a la baja a partir del período 3. Por su parte, la variable C muestra un impacto inverso a la variable D exhibe un impacto significativo desde el primer período, con una respuesta inicial a la baja que se inclina al alza a partir del período 3. En contraste, las variables T, PR y R no presentan respuestas estadísticamente significativas ante un choque en la generación fósil, ya que sus intervalos de confianza incluyen el valor cero a lo largo de la mayoría o la totalidad del horizonte temporal.

5.3. Analisis final

El análisis de los impulsos de respuesta generados por las dimensiones macroeconómicas latentes y las variables macroeconómicas individuales sobre los componentes de la tarifa de energía revela una heterogeneidad en la transmisión de los choques.

En cuanto a las dimensiones macroeconómicas latentes, se identifican impactos en las tarifas, con las siguientes interpretaciones en la realidad:

- La Dimensión 1 (Factor Monetario), fuertemente ligada a variables como el Peso colombiano, la Población y diferentes índices de precios al consumidor (Core, Excluding Food and Regulated), generó un impacto significativo negativo en la variable T (Transmisión) en el periodo 2. Esto puede explicarse porque, ante un endurecimiento de las condiciones monetarias o una fortaleza del peso colombiano, los costos asociados a la infraestructura de transmisión, que pueden tener componentes importados o costos operativos sensibles a la liquidez y tasas de interés, tienden a reducirse, generando un alivio en esta parte de la tarifa.
- La Dimensión 2 (Factor de Exposición Global), caracterizada por variables como Construcción, Tipos de cambio efectivos (reales) y Préstamos con mal desempeño, mostró un impacto positivo significativo inicial en T (Transmisión), aunque no persistente. Un shock positivo en esta dimensión podría reflejar un aumento en la actividad global o una mayor exposición a mercados internacionales. Si bien esto podría impulsar inicialmente la inversión o los costos de equipos que afectan la transmisión, su volatilidad sugiere que estos efectos no se mantienen de forma sostenida, posiblemente debido a la rápida adaptación del mercado o la naturaleza transitoria de ciertos choques externos.
- La Dimensión 3 (Factor de Apertura Comercial), dominada por Índices de precios de importación y exportación, Crédito neto doméstico al sector público y Términos de intercambio, tuvo un impacto positivo significativo y sostenido en la variable PR (Pérdidas Reconocidas). Es lógico que un aumento en los precios de importación (por ejemplo, de insumos o equipos eléctricos) o una mayor exposición al comercio exterior impacte negativamente en las pérdidas reconocidas, ya sea por un aumento en el costo de los materiales para el mantenimiento de la red o por una mayor presión en los costos operativos asociados a la infraestructura.
- La Dimensión 5 (Factor del Sector Inmobiliario y de Construcción), compuesta por variables como Viviendas asequibles, Viviendas unifamiliares, Índice de costes de construcción residencial y Viviendas multifamiliares, generó un impacto negativo significativo en la variable D (Distribución) en el sexto período. Un auge o contracción en el sector inmobiliario y de construcción influye directamente en la demanda de infraestructura eléctrica y en los costos de desarrollo. Un impacto negativo en la distribución podría significar que un freno en la construcción reduce la inversión en nuevas redes o que se optimizan los costos asociados a la infraestructura existente, lo que alivia la componente D de la tarifa.

Al analizar las variables macroeconómicas individuales, la significancia estadística se concentra en casos específicos y se puede contextualizar de la siguiente manera:

- Los "Precios al productor Interno" (X128) impactaron significativamente a la variable R (Restricciones) a partir del período 3. Dado que R se asocia con costos por congestión o falta de infraestructura, un aumento en los precios de producción internos podría reflejar presiones inflacionarias que encarecen el desarrollo de proyectos de infraestructura o el mantenimiento, llevando a mayores costos por restricciones en el sistema.
- El "Tipo de cambio (COP)" (X320) tuvo un impacto significativo en la variable T (Transmisión) en el período 2, sugiriendo una respuesta casi inmediata. La infraestructura de transmisión a menudo

depende de equipos importados y deudas en moneda extranjera. Una devaluación del peso colombiano (aumento del tipo de cambio) encarece directamente estas importaciones y el servicio de la deuda, lo que se traduce rápidamente en un incremento de los costos de transmisión que se reflejan en la tarifa.

- El "Crédito a negocios en circulación" (X335) afectó significativamente a la variable C (Comercialización) en el período 2 con una respuesta inicial. Un aumento en el crédito disponible para negocios podría generar mayor actividad económica y, en consecuencia, mayor consumo de energía. Esto podría impactar los costos de comercialización al requerir más recursos para la facturación, recaudo y atención al cliente, o al permitir a las empresas distribuidoras acceder a financiación más barata para sus operaciones.
- La "Demanda de energía" influyó significativamente en la variable D (Distribución) en el período 3. Un incremento en la demanda de energía implica una mayor presión sobre la red de distribución. Esto puede llevar a la necesidad de más inversiones en infraestructura, mantenimiento y expansión de la red, lo que eleva los costos de distribución y, por ende, la tarifa.
- La "Generación hidráulica" impactó significativamente a las variables G (Generación) y R (Restricciones) desde el mes 2, con una tendencia a la estabilización posterior. Un choque en la generación hidráulica, por ejemplo, debido a variaciones en el régimen de lluvias, afecta directamente la oferta de energía más barata. Menor generación hidráulica (aumento de R) puede llevar a mayores costos de generación al tener que recurrir a fuentes más caras (térmicas), y también a mayores restricciones en el sistema por la menor disponibilidad de energía, lo que se traslada a las tarifas G y R.
- Finalmente, la "Generación Fósil" mostró impactos significativos en G (Generación) en el período 4, en D (Distribución) de manera sostenida a la baja desde el inicio, y en C (Comercialización) de forma continua durante todo el período analizado. Un aumento en la generación fósil, que suele ser más costosa y dependiente de combustibles importados, naturalmente incrementa los costos de generación (G). Su impacto negativo en D y C es más complejo; podría deberse a que una mayor disponibilidad de energía fósil reduce la necesidad de inversiones en nuevas infraestructuras de distribución para atender picos de demanda, o que la mayor estabilidad en la oferta permite optimizar procesos de comercialización al haber menos incertidumbre en el suministro.

Las otras variables macroeconómicas individuales (Recibos Fiscales, Precios al productor Nacional, Empleo en el sector de abastecimiento, Deuda externa, y Población) no generaron impactos estadísticamente significativos en ninguno de los componentes tarifarios dentro del horizonte temporal considerado, lo que sugiere que su influencia es menos directa o menos robusta en el periodo analizado.

Los resultados consolidados de este análisis de impulsos evidencian una clara conexión y una influencia multidimensional de la macroeconomía sobre la conformación de las tarifas de energía en Colombia. Aunque los efectos varían en su magnitud y persistencia para cada componente tarifaria, y no todos los factores macroeconómicos ejercen el mismo nivel de impacto, los hallazgos en conjunto demuestran que variables y dimensiones macroeconómicas claves actúan como motores significativos en la dinámica de estos costos.

6. IMPLEMENTAR EL MODELO VAR ENTRENADO PARA PREDECIR LAS TARIFAS DE ENERGÍA ELÉCTRICA EN FUNCIÓN DE LOS DISTINTOS PARÁMETROS MACROECONÓMICOS IDENTIFICADOS.

El análisis de la influencia de los factores macroeconómicos sobre las tarifas energéticas es un paso fundamental para la comprensión de su dinámica. Sin embargo, para que esta comprensión se traduzca en valor práctico y herramientas de utilidad para la toma de decisiones, es imprescindible avanzar hacia la predicción de estas tarifas. Por ello, el presente apartado se enfoca en la implementación de modelos de pronóstico, cuyo objetivo principal es proyectar el comportamiento futuro de los componentes tarifarios (G, T, D, C, PR, R) con base en la información derivada de los análisis macroeconómicos previos.

Esta sección busca evaluar y comparar la capacidad predictiva de diferentes enfoques de modelado de series temporales, incluyendo el modelo de Vectores Autorregresivos (VAR), así como alternativas como el modelo Autorregresivo Integrado de Media Móvil con Explicativas (SARIMAX), el algoritmo de aprendizaje automático XGBoost, y una estrategia de modelado combinada que integra las capacidades del VAR con el potencial predictivo del XGBoost. Esta comparación es crucial para determinar cuál metodología ofrece el mejor desempeño en un contexto de variabilidad y complejidad como el del mercado energético colombiano. La finalidad es no solo pronosticar, sino también identificar el modelo más preciso y confiable, validando así la aplicación y relevancia de los hallazgos previos sobre las interacciones macroeconómicas. Para el entrenamiento de todos estos modelos, se utilizó el 90% de los datos disponibles, reservando el 10% restante para la evaluación de su desempeño y la realización de pruebas.

Para la construcción de los modelos de pronósticos, se utilizarán selectivamente aquellas variables macroeconómicas y dimensiones latentes que demostraron tener una significancia estadística en el análisis de impulsos para cada componente tarifaria. Por ejemplo, para la predicción de la variable T, se consideran principalmente las variables "X320" (Tipo de cambio), "Dimensión 1" (Factor Monetario) y "Dimensión 2" (Factor de Exposición Global), dado que fueron las que presentaron impactos significativos previamente.

Para cuantificar y comparar la capacidad predictiva de los modelos VAR, SARIMAX, XGBoost y la combinación VAR+XGBoost en la tarea de pronosticar los componentes tarifarios del sector energético se utilizarán adicional a análisis de las gráficas generadas, los indicadores MSE (Error Cuadrático Medio), RMSE (Raíz del Error Cuadrático Medio) y MAE (Error Absoluto Medio) que son métricas fundamentales utilizadas para evaluar el desempeño y la precisión de modelos de pronóstico. El MSE calcula el promedio de los cuadrados de las diferencias entre los valores pronosticados y los valores reales, penalizando en mayor medida los errores más grandes. El RMSE es simplemente la raíz cuadrada del MSE y proporciona una medida de la magnitud típica de los errores de predicción en las mismas unidades que la variable de interés, lo que facilita su interpretación. Por último, el MAE calcula el promedio de las diferencias absolutas entre las predicciones y los valores reales, ofreciendo una medida del error promedio sin considerar la dirección de la desviación.

6.1. Modelo de Vectores Autorregresivos (VAR)

Como se detalló en la sección 4.5. Modelo de Vector AutoRegresivo (VAR), el Modelo en su forma reducida es una herramienta robusta para analizar las interrelaciones dinámicas entre múltiples series temporales. Para el propósito de pronóstico, la ecuación del VAR se utiliza de la misma manera que para el análisis de impulso-respuesta. Una vez que el modelo ha sido estimado con datos históricos, los valores futuros de Y_t

se pueden pronosticar recursivamente. Para un horizonte de pronóstico h , el pronóstico para Y_{t+h} se obtiene sustituyendo Y_t y sus rezagos en la ecuación estimada. Para Y_{t+2} , se usa Y_{t+1} (el valor pronosticado) y así sucesivamente.

Hiperparámetros del Modelo VAR:

Orden de rezagos (p): El número de periodos pasados de las variables que se incluyen en el modelo para predecir los valores actuales. En este estudio, el orden de rezagos se determinó automáticamente utilizando el criterio de información de VARselect con un máximo de 12 rezagos, resultando en un valor óptimo de $p=3$.

6.2. Modelo Autorregresivo Integrado de Media Móvil con Explicativas (SARIMAX)

Es una extensión del modelo ARIMA que permite modelar series temporales con componentes estacionales y la inclusión de variables exógenas (predictores). La formulación general de un modelo SARIMAX(p,d,q)(P,D,Q,S) se define como:

$$\varphi(L)\vartheta(L^S)((1-L)^d)((1-L^S)^D)Y_t = c + \theta(L)\sigma(L^S)\epsilon_t + \beta X_t$$

Ilustración 35. Expresión del modelo SARIMAX

[27]

Donde:

- Y_t = Es la variable endógena en el periodo t (en este caso, una de las componentes tarifarias).
- X_t = Es un vector de variables exógenas (macroeconómicas y dimensiones latentes) en el periodo t .
- c = Es una constante
- $\varphi(L)$: $1 - \varphi_1(L) - \dots - \varphi_p(L^p)$ = Es el operador autorregresivo (AR) no estacional de orden p .
- $\vartheta(L^S)$: $1 - \vartheta_1(L^S) - \dots - \vartheta_p(L^{pS})$ = Es el operador autorregresivo estacional (SAR) de orden P con periodicidad S .
- $(1-L)^d$ = Es el operador de diferenciación no estacional de orden d , utilizado para lograr la estacionariedad.
- $(1-L^S)^D$ = Es el operador de diferenciación estacional de orden D con periodicidad S .
- $\theta(L)$: $1 + \theta_1(L) + \dots + \theta_q(L^q)$ = Es el operador de media móvil (MA) no estacional de orden q .
- $\sigma(L^S)$: $1 + \sigma_1(L^S) + \dots + \sigma_Q(L^{QS})$ =Es el operador de media móvil estacional (SMA) de orden Q con periodicidad S .
- L = Es el operador de rezago (es decir, $LY_t = Y_{t-1}$).
- ϵ_t = Es el término de error o ruido blanco, que se asume sigue una distribución normal con media cero y varianza constante.
- β = Son los coeficientes de las variables exógenas X_t .

Hiperparámetros del Modelo SARIMAX:

Los órdenes del modelo SARIMAX, tanto no estacionales como estacionales, se determinaron automáticamente utilizando el algoritmo `auto.arima()`. Órdenes no estacionales (p,d,q): (0,0,1). Órdenes estacionales (P,D,Q,S): (0,1,1,12). El valor de S=12 indica una periodicidad estacional mensual.

6.3. Algoritmo XGBoost (Extreme Gradient Boosting):

Es un algoritmo de aprendizaje automático basado en árboles de decisión que utiliza un enfoque de boosting (potenciación). A diferencia de los modelos econométricos tradicionales, XGBoost no se basa en supuestos sobre la distribución de los datos y es altamente efectivo para problemas de regresión y clasificación. Se construye de forma aditiva, donde cada nuevo árbol corrige los errores de los árboles anteriores.

La función objetivo que XGBoost minimiza durante el entrenamiento se compone de una función de pérdida (que mide la diferencia entre las predicciones y los valores reales) y un término de regularización (que penaliza la complejidad del modelo para evitar el sobreajuste). Para un problema de regresión como el pronóstico de tarifas, la función de pérdida comúnmente utilizada es el error cuadrático.

Dada una función objetivo $L(\varphi)$ a minimizar, donde φ representa el conjunto de árboles aditivos, la predicción \hat{y}_i para una instancia i se calcula como la suma de las predicciones de K árboles:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i)$$

Ilustración 36. Predicción Agregada del Modelo XGBoost

Donde $f_k(x_i)$ es la predicción del k -ésimo árbol para la entrada x_i . El algoritmo construye los árboles de manera iterativa, añadiendo un nuevo árbol f_t en cada paso t que minimiza la siguiente función objetivo:

$$L^t = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t)$$

Ilustración 37. Función Objetivo Minimizada por XGBoost

[28]

Donde L es la función de pérdida diferenciable (por ejemplo, error cuadrático) y $\Omega(f_t)$ es el término de regularización para el árbol f_t , que penaliza la complejidad del árbol para controlar el sobreajuste.

Hiperparámetros del Modelo XGBoost:

- `max_depth`: 3 (profundidad máxima de cada árbol). Controla la complejidad de los árboles, una mayor profundidad puede llevar a sobreajuste.
- `n_estimators`: 100 (número de árboles de potenciación). Determina cuántos árboles se construyen en el modelo.
- `eta`: 0.5 (tasa de aprendizaje). Controla el tamaño del paso en cada iteración de boosting. Un valor más bajo requiere más rounds pero puede mejorar la generalización.
- `objective`: "reg:squarederror" (función objetivo para regresión). Indica que el modelo está optimizado para minimizar el error cuadrático.

6.4. Modelo Combinado VAR + XGBoost

Busca aprovechar las fortalezas de ambos enfoques: la capacidad del VAR para capturar las interdependencias lineales y dinámicas entre las series temporales, y la habilidad de XGBoost para modelar relaciones no lineales y complejas en los residuos o pronósticos del VAR. Este enfoque híbrido se implementa en varias etapas:

1. Se entrena un modelo VAR para obtener un pronóstico inicial de las series temporales de interés. Este pronóstico captura la estructura lineal y las dependencias temporales entre las variables.
2. Se calculan los errores o residuos de pronóstico generados por el modelo VAR. Estos residuos representan la parte de la señal que el VAR no pudo explicar.
3. Se entrena un modelo XGBoost con el objetivo de predecir estos residuos del VAR. Para esto, XGBoost utiliza las variables macroeconómicas, las dimensiones latentes, y posiblemente también los rezagos de los propios residuos como predictores, buscando patrones no lineales que el VAR pasó por alto.
4. El pronóstico definitivo se obtiene sumando el pronóstico inicial del VAR y el pronóstico de los residuos generado por XGBoost. Esta combinación permite que el VAR maneje la estructura lineal subyacente y las dependencias temporales, mientras que XGBoost aprende patrones residuales complejos que el VAR no puede capturar, mejorando así la precisión general del pronóstico.

- Hiperparámetros del Modelo VAR + XGBoost:

Para la parte del modelo VAR: $P=3$ y para la parte del modelo XGBoost: $\text{max_depth} = 3$, $\text{nrounds} = 100$, $\text{eta} = 0.5$, $\text{objective} = \text{"reg:squarederror"}$.

6.5. Variable G

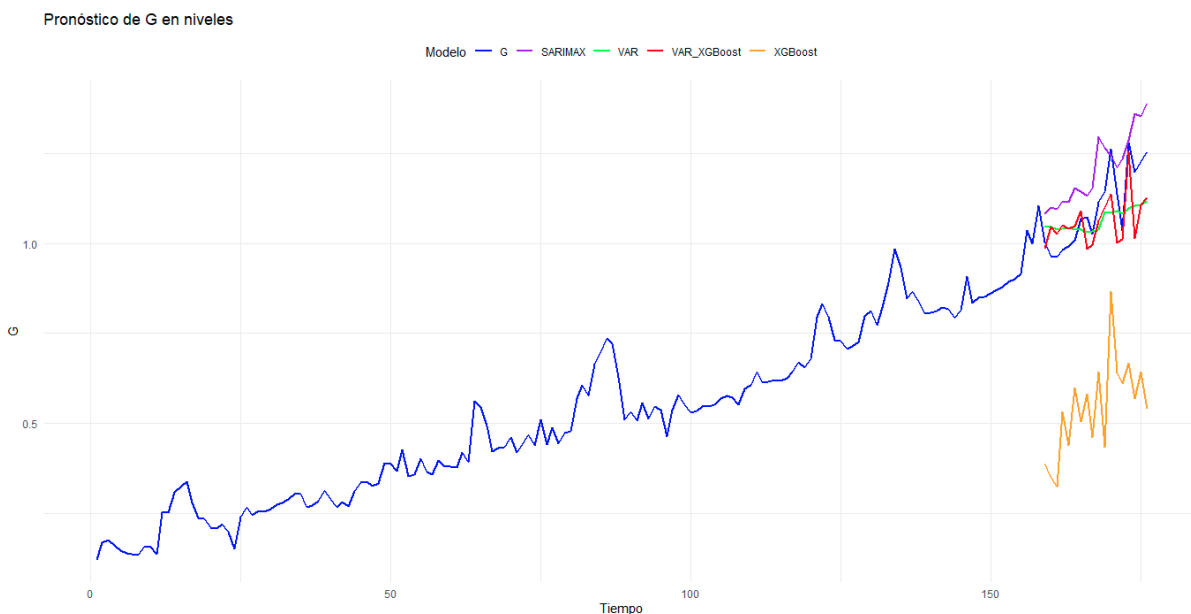


Ilustración 38. Pronóstico de la variable G

Los pronósticos de la variable objetivo "G" utilizando los predictores "G", "GenHidraulica" y "GenFosil" revelan un rendimiento diferenciado entre los modelos implementados. El modelo VAR+XGBoost exhibe las métricas de error más bajas, con un MSE de 0.00745, un RMSE de 0.08631 y un MAE de 0.07246, lo que sugiere una mayor precisión en la predicción en comparación con los otros modelos. El modelo VAR le sigue en rendimiento, con métricas ligeramente superiores (MSE: 0.00799, RMSE: 0.08938, MAE: 0.07521). En contraste, el modelo SARIMAX presenta un error mayor (MSE: 0.01531, RMSE: 0.12374, MAE: 0.11278), indicando una menor capacidad predictiva para esta variable específica. El modelo XGBoost muestra un rendimiento considerablemente inferior, con métricas de error sustancialmente más altas (MSE: 0.31446, RMSE: 0.56077, MAE: 0.55268), lo que sugiere que, en su forma pura y con los predictores definidos, no es el modelo más adecuado para pronosticar la variable "G".

Visualmente, la gráfica "Pronóstico de G en niveles" muestra cómo las predicciones del VAR y el VAR+XGBoost (líneas verde y roja respectivamente) tienden a seguir más de cerca la trayectoria observada de la variable "G" en el período de pronóstico. El modelo SARIMAX (línea morada) también muestra una tendencia similar, aunque con una mayor desviación en algunos puntos. La predicción del XGBoost (línea naranja) se desvía significativamente de los otros modelos y de la tendencia observada, lo que se alinea con sus métricas de error más elevadas. En general, la combinación del modelo VAR con XGBoost parece ofrecer una ligera mejora en la precisión del pronóstico para la variable "G", aunque el modelo VAR por sí solo también proporciona resultados competitivos.

6.6. Variable T

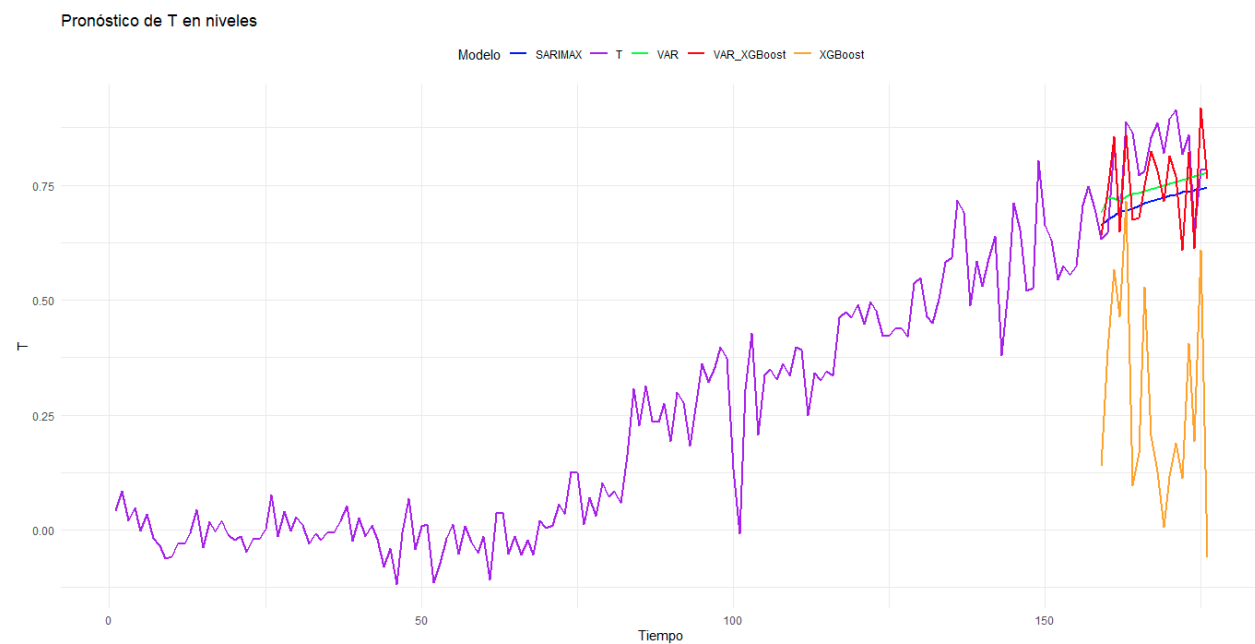


Ilustración 39. Pronóstico de la variable T

El pronóstico de la variable objetivo "T" utilizando los predictores "T", "X320" (Tipo de cambio), "Dim1" (Factor Monetario) y "Dim2" (Factor de Exposición Global) también presenta un rendimiento variado entre los modelos. Al igual que en el caso de la variable "G", la combinación VAR+XGBoost demuestra ser el modelo con el menor error, registrando un MSE de 0.00942, un RMSE de 0.09708 y un MAE de 0.07675. Esto sugiere una ligera mejora en la precisión predictiva al combinar ambos enfoques. El modelo VAR se

sitúa en segundo lugar en términos de rendimiento, con métricas de error ligeramente superiores (MSE: 0.01005, RMSE: 0.10023, MAE: 0.08802). El modelo SARIMAX exhibe un error mayor (MSE: 0.01401, RMSE: 0.11836, MAE: 0.10355), indicando una menor capacidad para capturar la dinámica de la variable "T" en el horizonte de pronóstico. De manera similar a lo observado con la variable "G", el modelo XGBoost presenta un rendimiento considerablemente inferior (MSE: 0.32849, RMSE: 0.57314, MAE: 0.52105), lo que sugiere que no es el método más adecuado para pronosticar la variable "T" con los predictores seleccionados.

Visualmente, la gráfica "Pronóstico de T en niveles" revela diferencias notables en la capacidad de los modelos para capturar la dinámica de la variable "T" en el período de pronósticos. Se observa que tanto el modelo VAR (línea verde) como el modelo SARIMAX (línea azul) tienden a proyectar líneas relativamente rectas, mostrando una capacidad limitada para replicar las fluctuaciones y la volatilidad que presenta la variable "T" en el período de predicción. Si bien el modelo VAR parece mantenerse ligeramente más cercano a la tendencia general en comparación con el SARIMAX, ambos contrastan con el comportamiento más dinámico que intenta capturar el modelo VAR+XGBoost (línea roja). Tal como refleja sus métricas, el modelo XGBoost (línea naranja) se desvía significativamente de la trayectoria observada y de las predicciones de los otros modelos. En general el modelo VAR+XGBoost se destaca como el que visualmente se ajusta mejor a la tendencia ya las variaciones del comportamiento de la variable "T" en el período de predicción, similar a lo observado en el pronóstico de la variable "G".

6.7. Variable D

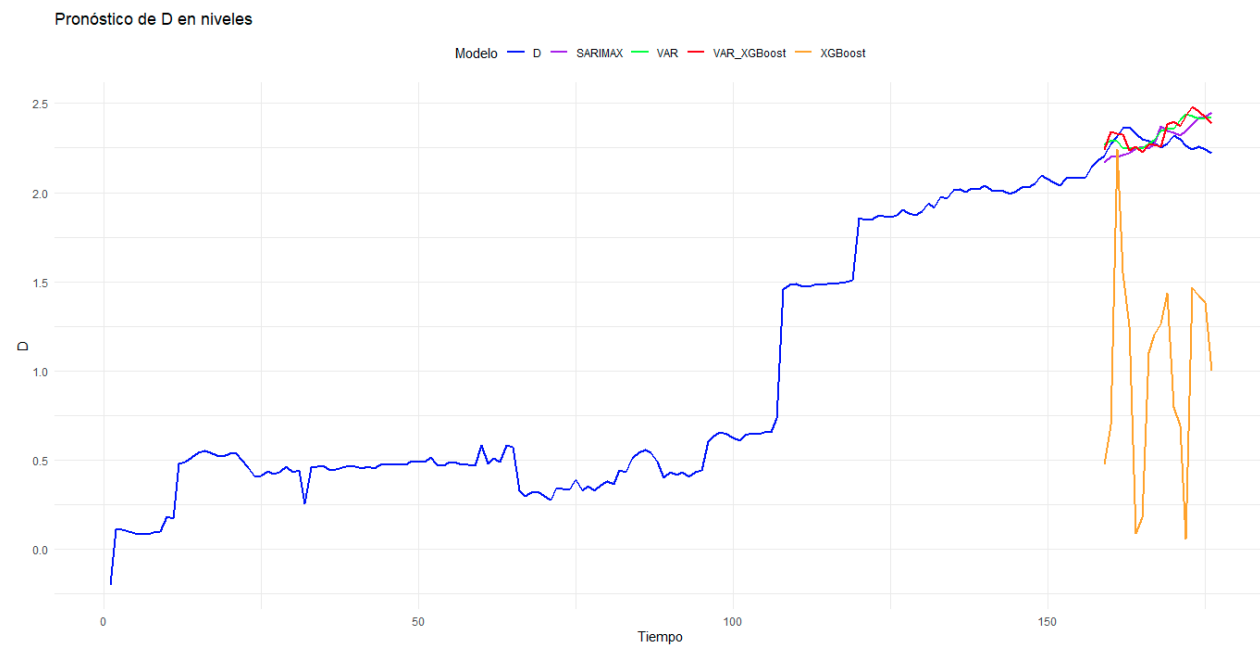


Ilustración 40. Pronóstico de la variable D

Los pronósticos de la variable objetivo "D" utilizando los predictores "D", "Dim5" (Factor del Sector Inmobiliario y de construcción), "Demanda" (Demanda de energía) y "GenFosil" (Generación Fosil) arroja resultados donde tres de los modelos muestran un rendimiento relativamente similar, mientras que el

XGBoost se destaca negativamente. En términos de métricas de error, el modelo SARIMAX presenta ligeramente el MSE más bajo (0.01292), seguido muy de cerca por el modelo VAR (MSE: 0.01299). El modelo VAR+XGBoost exhibe un MSE ligeramente superior (0.01361), aunque las diferencias en RMSE (0.1137 para SARIMAX, 0.1140 para VAR y 0.1167 para VAR+XGBoost) y MAE (0.0957 para SARIMAX, 0.0958 para VAR y 0.0938 para VAR+XGBoost) también son pequeñas entre estos tres modelos. Esto sugiere que, estadísticamente, el rendimiento predictivo de VAR, SARIMAX y VAR+XGBoost es comparable para la variable "D". Por otro lado, el modelo XGBoost vuelve a mostrar un rendimiento sustancialmente inferior, con métricas de error muy elevadas (MSE: 1.916, RMSE: 1.384, MAE: 1.266), lo que lo descarta como una opción adecuada para el pronóstico de esta variable con los predictores utilizados.

Visualmente, la gráfica del pronóstico de "D" en niveles confirma la similitud en el comportamiento predictivo de los modelos VAR, SARIMAX y VAR+XGBoost. Las líneas correspondientes a estos tres modelos (verde, morada y roja respectivamente) tienden a superponerse y a seguir la trayectoria observada de la variable "D" en el período de pronósticos de manera similar. Aunque las diferencias son sutiles, la línea del modelo VAR+XGBoost parece mostrar una ligera capacidad para ajustarse a los cambios de dirección de la variable real, tal como se apreció en el análisis de la variable "G", lo que podría explicar su MAE ligeramente inferior en comparación con el VAR. Una vez más, la línea naranja del modelo XGBoost se desvía considerablemente del comportamiento real y de las predicciones de los otros modelos, lo que refuerza su pobre desempeño para esta variable. En general, para los pronósticos de la variable "D", los modelos VAR, SARIMAX y la combinación VAR+XGBoost ofrecen resultados comparables, con una ligera ventaja visual para el modelo combinado en términos de adaptación a la dinámica de la serie.

6.8. Variable C

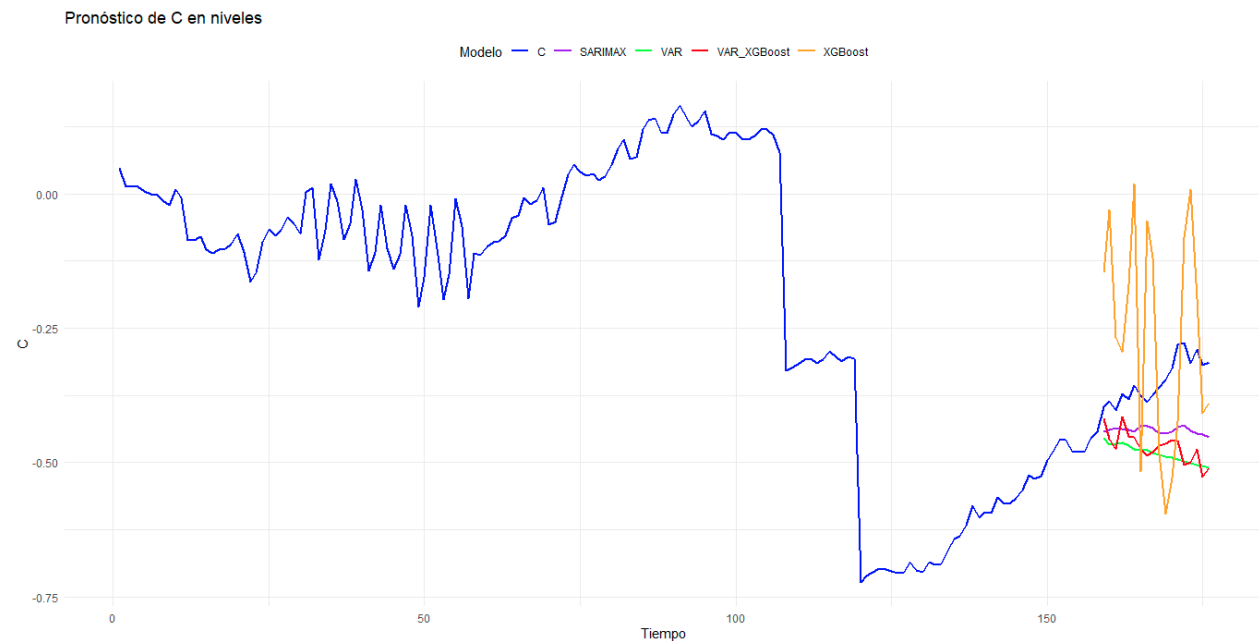


Ilustración 41. Pronóstico de la variable C

Los pronósticos de la variable objetivo "C" utilizando los predictores "C", "X335" (Crédito a negocios en circulación) y "GenFosil" (Generación Fosil) revelan que, en general, ninguno de los modelos logra capturar perfectamente la trayectoria del valor real. Sin embargo, al analizar las métricas de error, el modelo

SARIMAX presenta los valores más bajos (MSE: 0.01021, RMSE: 0.10105, MAE: 0.09220), lo que sugiere que, estadísticamente, es el modelo con el mejor rendimiento para esta variable y estos predictores. Le sigue el modelo VAR+XGBoost, que muestra métricas de error ligeramente superiores (MSE: 0.01857, RMSE: 0.13626, MAE: 0.12313), indicando una capacidad predictiva razonable. El modelo VAR exhibe un error aún mayor (MSE: 0.02140, RMSE: 0.14630, MAE: 0.13594), mientras que el modelo XGBoost presenta el peor desempeño, con métricas de error con considerables elevadas (MSE: 0.05032, RMSE: 0.22431, MAE: 0.20245).

Visualmente, la gráfica del pronóstico de "C" en niveles confirma que ningún modelo sigue con exactitud el valor real. No obstante, se observa que el modelo VAR+XGBoost (línea roja) es el único que intenta capturar las fluctuaciones, las alzas y bajas de la variable "C", aunque con cierta desfase y divergencia en la magnitud. Un patrón interesante es que, mientras el valor real de "C" muestra una tendencia al alza en la parte final del período de pronósticos, los modelos VAR+XGBoost, VAR (línea verde) y SARIMAX (línea morada) tienden a predecir una disminución. El modelo XGBoost (línea naranja) presenta una trayectoria errática y alejada del comportamiento real, lo que concuerda con sus altas métricas de error. En este caso específico, a pesar de no ofrecer una predicción perfecta, el modelo VAR+XGBoost parece ser el que mejor intenta replicar la dinámica de la variable "C". El modelo SARIMAX, si bien con métricas de error más bajas, ofrece una predicción más suavizada que no captura las fluctuaciones de la misma manera.

6.9. Variable PR

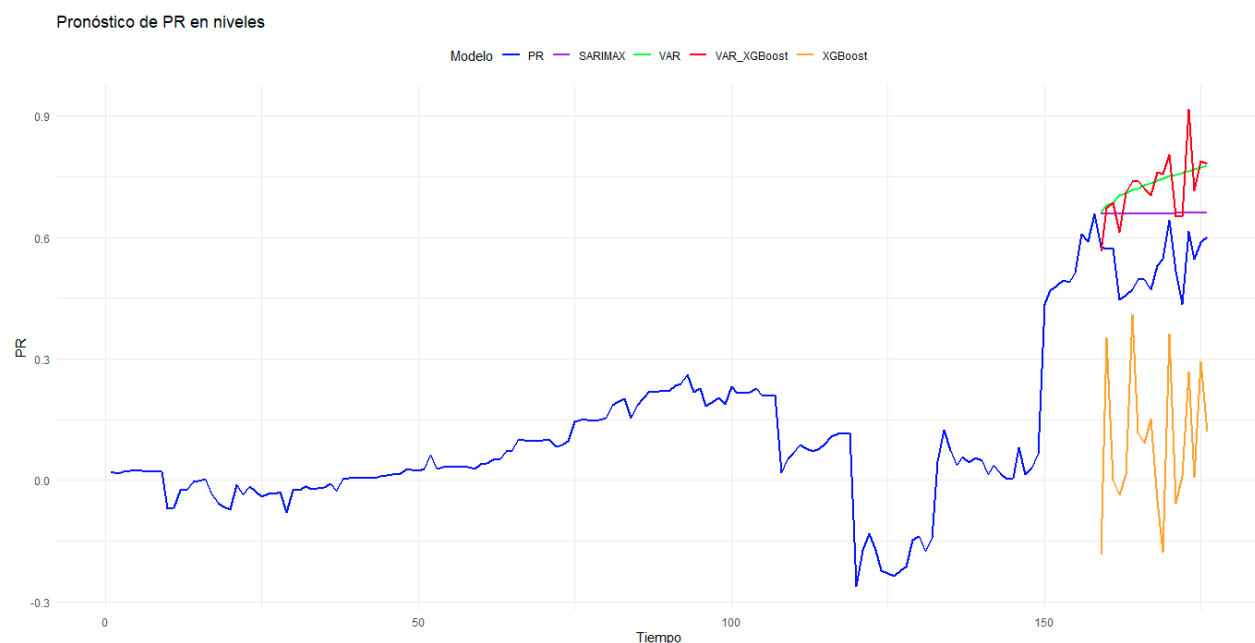


Ilustración 42. Pronóstico de la variable PR

Los pronósticos de la variable objetivo "PR" utilizando los predictores "PR" y "Dim3" (Factor de Apertura Comercial) revelan que, según las métricas de error, el modelo SARIMAX presenta el mejor rendimiento (MSE: 0.01967, RMSE: 0.14023, MAE: 0.12671), indicando la mayor precisión estadística entre los modelos. Le sigue el modelo VAR+XGBoost, con métricas ligeramente superiores (MSE: 0.04059, RMSE: 0.20146, MAE: 0.18988), sugiriendo una capacidad predictiva razonable. El modelo VAR muestra un error aún mayor

(MSE: 0.04398, RMSE: 0.20971, MAE: 0.19984), mientras que el modelo XGBoost presenta el peor desempeño, con métricas de error con considerables elevadas (MSE: 0.22081, RMSE: 0.46991, MAE: 0.43822).

Visualmente, la gráfica del pronóstico de "PR" en niveles muestra que el modelo VAR+XGBoost (línea roja) parece capturar mejor los movimientos de alzas y bajas de la variable real, aunque se mantiene consistentemente un poco más alto en su predicción. El modelo VAR (línea verde) proyecta una línea casi recta con una tendencia alza, lo que no refleja la variabilidad observada. El modelo SARIMAX (línea morada) ofrece un pronóstico como una línea completamente recta y horizontal, lo que sugiere que no logra capturar ninguna dinámica de la variable "PR". Una vez más, el modelo XGBoost (línea naranja) se desvía significativamente de las predicciones de los otros modelos y de la trayectoria real. En este caso, a pesar de que las métricas favorecen al SARIMAX, la inspección visual sugiere que el modelo VAR+XGBoost es el que mejor intenta replicar la dinámica de la variable "PR", siguiendo sus fluctuaciones. El modelo VAR simplifica la tendencia, y el SARIMAX ignora por completa la variabilidad de la serie.

6.10. Variable R

Pronóstico de R en niveles

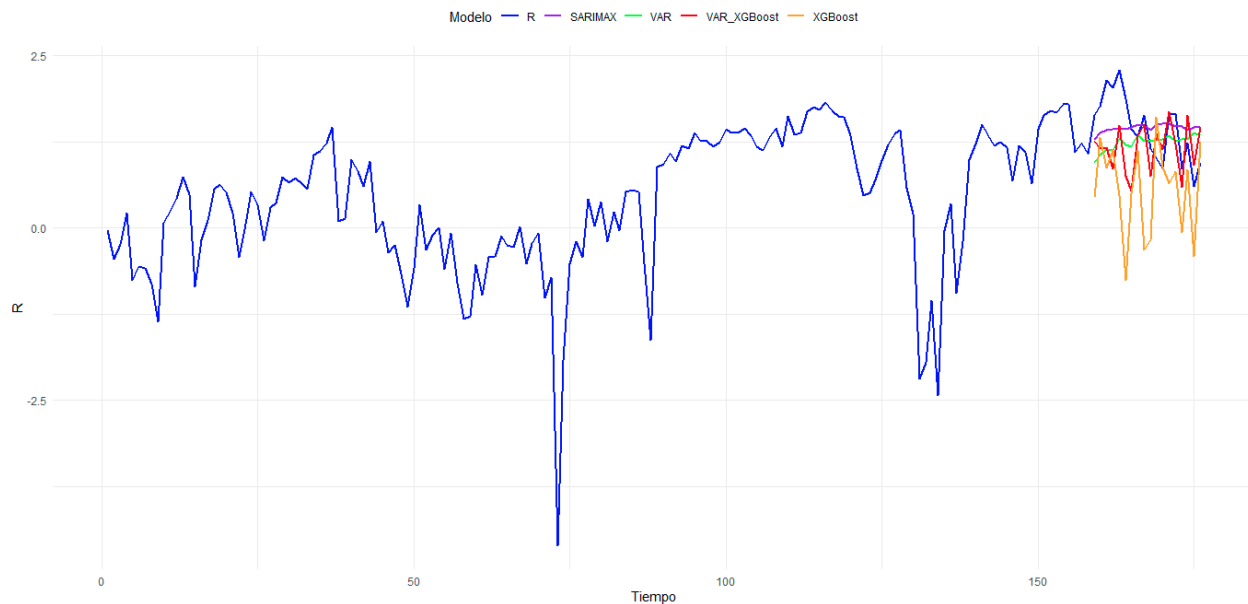


Ilustración 43. Pronóstico de la variable R

Los pronósticos de la variable objetivo "R" utilizando los predictores "R", "GenHidraulica" y "X128" (Precios al productor Interno) revelan que, según las métricas de error, el modelo SARIMAX presenta el mejor rendimiento (MSE: 0.2387, RMSE: 0.4885, MAE: 0.4213), indicando la mayor precisión estadística entre los modelos. Le sigue el modelo VAR, con métricas ligeramente superiores (MSE: 0.3278, RMSE: 0.5725, MAE: 0.4870). El modelo VAR+XGBoost exhibe un error aún mayor (MSE: 0.3716, RMSE: 0.6096, MAE: 0.5045), mientras que el modelo XGBoost presenta el peor desempeño, con métricas de error considerablemente elevadas (MSE: 1.380, RMSE: 1.175, MAE: 0.985).

Visualmente, la gráfica del pronóstico de "R" en niveles sugiere una interpretación diferente a lo que indican las métricas. Se observa que el modelo VAR+XGBoost (línea roja) parece capturar de manera notable la dinámica de la variable "R", siguiendo de cerca sus fluctuaciones, incluyendo los picos y valles, en el período de pronósticos. Los modelos VAR (línea verde) y SARIMAX (línea morada) muestran un comportamiento similar entre sí, ofreciendo una predicción más suavizada que se ajusta al promedio de la variable, pero que no logra replicar la volatilidad y los movimientos específicos que sí capta el modelo combinado. Una vez más, el modelo XGBoost (línea naranja) se desvía significativamente de la trayectoria real y de las predicciones de los otros modelos. En este caso, a pesar de que las métricas favorecen al SARIMAX, la inspección visual indica que el modelo VAR+XGBoost es el que mejor se adapta y predice la dinámica de la variable "R", mostrando una capacidad superior para seguir sus cambios en comparación con las predicciones más suavizadas del VAR y el SARIMAX.

6.11. Análisis final

La evaluación del desempeño predictivo de los modelos VAR, SARIMAX, XGBoost y la combinación VAR+XGBoost sobre las diferentes componentes de las tarifas energéticas (G, T, D, C, PR y R) ha permitido identificar qué enfoque resulta más adecuado para predecir el comportamiento futuro de estas variables. Más allá de la mera comparación de métricas de error, resulta fundamental comprender la capacidad de cada modelo para capturar la verdadera dinámica de las tarifas.

Al analizar las métricas de error (MSE, RMSE, MAE), que cuantifican la proximidad de las predicciones a los valores reales, se observaron los siguientes resultados: el modelo VAR+XGBoost exhibió las métricas de error más bajas en dos de las componentes (G y T), lo cual indica una alta precisión cuantitativa en esos casos. Para las demás variables (D, C, PR y R), VAR+XGBoost se posicionó consistentemente como el segundo mejor modelo en términos de métricas, demostrando un rendimiento robusto. Los modelos VAR y SARIMAX, basados en técnicas estadísticas y econométricas tradicionales, presentaron un rendimiento competitivo; en algunas instancias, SARIMAX incluso logró métricas de error marginalmente superiores (para las variables D, C, PR y R), posicionándose como el mejor en esos casos. Por otro lado, XGBoost en su forma pura mostró consistentemente los errores más elevados y un desempeño deficiente en todas las variables. Esto sugiere que, para series de tiempo económicas con dependencias temporales como las tarifas, este algoritmo requiere ser complementado con modelos que incorporen de manera inherente la estructura temporal.

Un aspecto crucial en este análisis es que, dada la metodología de este proyecto, las predicciones se realizaron sobre bases de datos diferenciadas. Esto significa que los resultados de los pronósticos representan variaciones mensuales en las tarifas, más que el valor exacto para un mes determinado. Por lo tanto, la capacidad de los modelos para anticipar las subidas y bajadas que experimentan las tarifas se vuelve de suma relevancia. Esta característica es fundamental, ya que el análisis de las tarifas energéticas a menudo se centra en sus variaciones para comprender su comportamiento real y su impacto en la planificación. Si bien minimizar el error promedio es importante, para la toma de decisiones, entender la dirección y la magnitud del cambio en estas variaciones es de vital importancia.

La inspección visual de las gráficas de pronóstico revela un patrón distintivo. El modelo VAR+XGBoost demostró una habilidad superior para replicar la dinámica intrínseca de la serie temporal. Se ajustó con mayor fidelidad a las fluctuaciones, picos y valles observados en las tarifas mes a mes. Esta capacidad de "seguir el ritmo" de la variable real es particularmente notable en variables como G, T, C, PR y R, donde este modelo fue el único que intentó capturar las fluctuaciones reales, mientras que otros enfoques

ofrecían predicciones más "suaves" o líneas rectas que no reflejaban la volatilidad. En contraste, los modelos SARIMAX y VAR, si bien en ocasiones exhibieron métricas de error favorables, tendieron a suavizar excesivamente la serie. Sus pronósticos a menudo se presentaron como líneas más homogéneas o con menor variabilidad que la realidad, lo que implica una limitación para capturar la riqueza de las variaciones mensuales de interés. Por ejemplo, en las variables T, PR y R, SARIMAX y VAR predecían tendencias, pero no las oscilaciones críticas que el modelo híbrido sí intentaba replicar. La incapacidad de XGBoost en su forma pura para capturar la dinámica de las series de tiempo se evidenció tanto en sus elevadas métricas de error como en sus representaciones gráficas, donde sus predicciones se desviaron significativamente.

La superioridad del modelo híbrido VAR+XGBoost puede atribuirse a dos razones principales. Primero, su diseño combina las fortalezas de dos enfoques complementarios: el VAR es eficaz para identificar relaciones lineales e interdependencias temporales entre variables macroeconómicas y tarifas, mientras que XGBoost sobresale en la detección de patrones complejos y no lineales que los modelos econométricos tradicionales no pueden capturar. Esta integración permite que el modelo híbrido aproveche la estructura temporal del VAR y la flexibilidad de XGBoost para capturar tanto la tendencia general como las variaciones y cambios abruptos en las tarifas. Esto es particularmente relevante dado que el mercado energético colombiano presenta características de ruido y no linealidad, haciendo más valioso un modelo capaz de adaptarse a estas complejidades. Segundo, su relevancia para la toma de decisiones es significativa. En el sector energético, la precisión no solo se mide por un error promedio bajo; es crucial predecir el momento y la magnitud de las subidas o bajadas significativas en las tarifas. Si bien las métricas de error promedian todas las desviaciones y pueden no reflejar plenamente esta capacidad, la habilidad del VAR+XGBoost para replicar estas fluctuaciones lo posiciona como una herramienta más útil y confiable, al proporcionar una imagen más fiel de la probable evolución de las tarifas mes a mes.

En conclusión, aunque las métricas de error son un componente esencial de la evaluación, el análisis detallado, especialmente el cualitativo de las gráficas, subraya que el modelo híbrido VAR+XGBoost es el más adecuado para el pronóstico de las tarifas energéticas colombianas. Su principal fortaleza radica en su capacidad para predecir de manera efectiva las variaciones y fluctuaciones mensuales, un atributo fundamental para la planificación en un mercado tan dinámico.

7. CONCLUSIONES Y TRABAJOS FUTUROS

7.1. CONCLUSIONES

A través del desarrollo de la presente investigación, se lograron los siguientes resultados y se llegó a las siguientes conclusiones en relación con los objetivos planteados:

7.1.1. Conclusión general

Este trabajo ha logrado evaluar el impacto de factores macroeconómicos sobre el costo de la tarifa de energía eléctrica en Colombia mediante el uso de modelos de Vectores Autorregresivos (VAR) y técnicas de aprendizaje supervisado. El análisis de los impulsos de respuesta, piedra angular de la aplicación del modelo VAR, ha sido fundamental para desentrañar las complejas relaciones dinámicas entre el entorno macroeconómico y las diferentes componentes de la tarifa de energía eléctrica en Colombia.

Los resultados consolidados de este análisis permiten concluir que existe una relación importante entre la macroeconomía y la dinámica de las tarifas de energía en Colombia. Si bien no todos los factores macroeconómicos y no todas las variables mostraron impactos significativos en cada caso particular, la combinación de los hallazgos revela una conexión clara. Se ha evidenciado que dimensiones latentes y variables macroeconómicas específicas (como el tipo de cambio, el crédito, los precios al productor interno, y factores de demanda y generación) ejercen una influencia estadísticamente significativa sobre distintas componentes tarifarias. Por ejemplo, se observaron impactos relevantes de factores monetarios y de exposición global en la variable T, entre otros. Esta variabilidad en los impactos, donde no todas las variables se ven afectadas por todos los factores, subraya la complejidad y especificidad de los canales de transmisión de choques macroeconómicos hacia el sector energético.

Esta evidencia de impactos significativos en múltiples puntos de la estructura tarifaria implica un hallazgo crucial: la macroeconomía no es un elemento aislado en la formación de las tarifas de energía. Por el contrario, sus fluctuaciones y tendencias se propagan a través de diversos canales, afectando los costos de componentes específicos. Este entendimiento matizado, que va más allá de una simple correlación, es un aporte sustancial al conocimiento sobre los determinantes de las tarifas energéticas en Colombia, proporcionando una base empírica para el análisis de políticas y la toma de decisiones informada.

Adicionalmente, en el marco de este estudio, se desarrollaron modelos de pronóstico que exhibieron un desempeño prometedor. Particularmente, la estrategia combinada VAR+XGBoost se destacó por su capacidad para replicar la dinámica y las fluctuaciones mensuales de las series temporales, lo cual es de gran valor para la anticipación de cambios en un sector tan dinámico. Los resultados obtenidos en este estudio no solo enriquecen el conocimiento sobre la interacción entre la macroeconomía y el sector energético, sino que también proporcionan herramientas analíticas y predictivas valiosas para la toma de decisiones en este ámbito crucial de la economía colombiana.

7.1.2. Conclusiones específicas

- Se compiló una base de datos robusta que integra información tarifaria detallada con una amplia gama de variables macroeconómicas relevantes para la economía colombiana. La implementación de estrategias de limpieza e imputación permitió contar con un conjunto de datos de calidad adecuado para el análisis econométrico posterior.

- La aplicación del PCA fue exitosa en la identificación de cinco factores macroeconómicos latentes que explican una proporción significativa (77,77%) de la variación en el conjunto de datos originales. Estos factores, interpretados como dimensiones económicas fundamentales, no solo simplificaron la complejidad del análisis macroeconómico, sino que también facilitaron su incorporación eficiente en el modelo VAR, contribuyendo a la estabilidad y reducción de la multicolinealidad.
- Se implementó un modelo VAR para analizar las interrelaciones dinámicas entre los componentes tarifarios y los factores macroeconómicos. Si bien la mayoría de las series requirieron una transformación logarítmica para alcanzar la estacionariedad, se logró estimar un modelo VAR que permitió examinar la propagación de choques macroeconómicos a las tarifas de energía.
- El análisis de los impulsos reveló relaciones dinámicas específicas entre ciertas variables y dimensiones macroeconómicas y componentes tarifarias. Se identificaron múltiples casos de significancia estadística que demuestran una clara interdependencia entre la macroeconomía y las tarifas. Este análisis proporcionó información que permitió comprender la sensibilidad de las tarifas ante perturbaciones macroeconómicas y para la selección de predictores claves en la etapa de pronóstico, justificando la validez de los factores macroeconómicos que influyen en los costos de energía.
- Se desarrollaron y compararon cuatro modelos de pronóstico diferentes. La evaluación del desempeño mediante métricas de error y análisis visual sugiere que, en general, el modelo combinado VAR+XGBoost muestra una mayor capacidad para capturar la dinámica real de las series tarifarias, aunque en algunos casos las métricas favorecieron ligeramente a otros modelos. Este hallazgo resalta el potencial de la combinación de modelos lineales y no lineales para mejorar la precisión de los pronósticos en el sector energético.

7.2. TRABAJOS FUTUROS

A partir de los hallazgos y las limitaciones identificadas en el presente trabajo, se proponen las siguientes líneas de investigación y desarrollos futuros:

- **Análisis a Nivel Regional:** El presente estudio se centró en un promedio nacional de las tarifas de energía. Futuras investigaciones podrían desagregar el análisis a nivel regional, considerando las particularidades económicas y regulatorias de cada zona del país. Esto permitiría identificar relaciones más específicas y desarrollar modelos de pronósticos adaptados a las dinámicas regionales.
- **Incorporación de Variables Regulatorias:** El marco regulatorio del sector energético tiene un impacto significativo en la formación de las tarifas. Trabajos futuros podrían incluir variables regulatorias específicas (como cambios en la legislación, decisiones de la Comisión de Regulación de Energía y Gas - CREG, o políticas de subsidios) dentro del modelo VAR y los modelos de pronósticos para mejorar su capacidad explicativa y predictiva.
- **Modelado de No Linealidades:** Si bien el modelo VAR+XGBoost demostró un buen desempeño, la exploración de otros modelos no lineales o técnicas de aprendizaje automático más avanzadas (como redes neuronales recurrentes) podría capturar patrones no lineales más complejos en la relación entre las variables macroeconómicas y las tarifas de energía.
- **Evaluación de Escenarios de Política:** Los modelos de pronósticos desarrollados podrían utilizarse para evaluar el impacto potencial de diferentes escenarios de política económica o energética en

la evolución futura de las tarifas. Esto podría proporcionar información valiosa para la toma de decisiones por parte de los reguladores y los actores del sector.

- **Frecuencia de los Datos:** La disponibilidad de datos con mayor frecuencia (por ejemplo, semanal o diaria) podría permitir la captura de dinámicas de corto plazo más precisas y la mejora de los modelos de pronósticos, especialmente en respuesta a eventos macroeconómicos o regulatorios repentinos.
- **Explorar razones subyacentes al mejor desempeño visual del modelo VAR+XGBoost:** Esto podría incluir un análisis más detallado de los patrones no lineales específicos que el XGBoost está identificando y cómo estos interactúan con la estructura lineal del VAR. Además, la experimentación con diferentes configuraciones del modelo híbrido, la incorporación de características adicionales o la aplicación de técnicas de validación cruzada más robustas podrían ayudar a consolidar aún más su potencial predictivo y comprender mejor sus fortalezas y limitaciones en el contexto de la predicción de tarifas energéticas.

8. REFERENCIAS BIBLIOGRÁFICAS

[1] Ley 143 de 1994, "Por la cual se establece el régimen para la generación, interconexión, transmisión, distribución y comercialización de electricidad en el territorio nacional, se conceden unas autorizaciones y se dictan otras disposiciones en materia energética", Art. 11, 15 de diciembre de 1994.

[2] Portafolio, "Así se saca el precio promedio de la tarifa de la energía en el país", Portafolio, 19 de octubre de 2022. [En línea]. Disponible en: <https://www.portafolio.co/economia/asi-se-saca-el-precio-promedio-de-la-tarifa-de-la-energia-en-el-pais-570973>.

[3] JV Arenales, "Estas son las razones que componen el incremento del recibo de la energía en marzo", Diario La República, 5 de abril de 2023. [En línea]. Disponible en: <https://www.larepublica.co/economia/estas-son-las-razones-del-incremento-del-recibo-de-energia-en-marzo-3586908>.

[4] AZ Quinchía, "Recibo de la luz: kilovatio subió más de 30% en algunos estratos", El Colombiano, 7 de abril de 2024. [En línea]. Disponible en: <https://www.elcolombiano.com/negocios/impacto-del-fenomeno-de-el-nino-en-las-tarifas-de-energia-en-colombia-DA24174106>.

[5] D. Garzón, "Consumo de energía creció 12,06 % en marzo: con estas medidas buscan evitar el racionamiento", Infobae, 19 de abril de 2024. [En línea]. Disponible en: <https://www.infobae.com/colombia/2024/04/19/consumo-de-energia-crecio-1206-en-marzo-con-estas-medidas-buscan-evitar-el-rationamiento/>.

[6] AP Maestre, "Presidente Petro ordenó al MinMinas bajar tarifas de energía del país en menos de un mes", W Radio, 8 de mayo de 2024. [En línea]. Disponible en: <https://www.wradio.com.co/2024/05/08/presidente-petro-ordeno-al-minminas-bajar-tarifas-de-energia-del-pais-en-menos-de-un-mes/>.

[7] CAB Nieto y JC Robledo, "Relación a largo plazo entre consumo de energía y PIB en América Latina: Una evaluación empírica con datos panel", Investigaciones de Economía, vol. 32, núm. 78, págs. 13-37, 2012. [En línea]. Disponible en: <https://www.redalyc.org/articulo.oa?id=329027339004>.

[8] Amazon Web Services, Inc., "¿Qué es la ciencia de datos? - Explicación de la ciencia de datos", Amazon Web Services, Inc., sin fecha. [En línea]. Disponible en: <https://aws.amazon.com/es/what-is/data-science/>.

[9] TA Blanco, "¿Qué es y cómo calcular la potencia eléctrica?", BBVA Noticias, 7 de diciembre de 2023. [En línea]. Disponible en: <https://www.bbva.com/es/sostenibilidad/que-es-y-como-calcular-la-potencia-electrica/>.

[10] Electrificadora de Santander, "¿Qué es la energía eléctrica?", ESSA, sin fecha. [En línea]. Disponible en: <https://www.essa.com.co/site/comunidad/portal-kids/aprende-sobre-la-energia/-que-es-la-energia-electrica>.

[11] P. Flores y JP Londoño, "Efectos Macroeconómicos de la Política Fiscal en Ecuador 1993-2009", *Analítika, Revista de Análisis Estadístico*, vol. 9, núm. 1, art. 4, 2015.

[12] Equipo editorial, Etecé, "Generación Eléctrica - Qué es, cómo se produce, renovables", *Concepto*, 16 de julio de 2021. [En línea]. Disponible en: <https://concepto.de/generacion-electrica/>.

[13] Comisión de Regulación de Energía y Gas (CREG), "Distribución de Energía Eléctrica", Bogotá, Colombia, Tomo 1, p. 5, 2008.

[14] Sonia, "¿Qué es una comercializadora eléctrica?", *Audinfo System - Software Sector Energético*, 23 de marzo de 2023. [En línea]. Disponible en: <https://www.audinfosystem.es/todo-energia/que-es-una-comercializadora-de-energia/>.

[15] Comisión de Regulación de Energía y Gas (CREG), Resolución CREG 119 de 2007, *Diario Oficial*, no. 46881, 24 de enero de 2008.

[16] Comisión de Regulación de Energía y Gas (CREG), "Estructura tarifaria", CREG, 10 de julio de 2023. [En línea]. Disponible en: <https://creg.gov.co/publicaciones/7820/estructura-tarifaria/>.

[17] Comisión de Regulación de Energía y Gas (CREG), Resolución CREG 183 de 2009, *Diario Oficial*, no. 47576, 28 de diciembre de 2009.

[18] Banco de la República, "Índice de la tasa de cambio real (ITCR)", Banco de la República, sin fecha. [En línea]. Disponible en: <https://www.banrep.gov.co/es/estadisticas/indice-tasa-cambio-real-itcr>.

[19] O. Joaqui, B. Osorio, C. Ramírez y C. Ojeda, "Explorando la relación asimétrica entre factores macroeconómicos y rentabilidad corporativa en el índice MSCI Colombia", *Journal of Economics, Finance and Administrative Science*, 2024, Acceso anticipado. doi: 10.1108/JEFAS-01-2024-0010.

[20] O. Joaqui, J. Gómez y S. López, "Interconexiones macrofinancieras en la Alianza del Pacífico: un enfoque cuantil de los mercados bursátiles y los factores macroeconómicos", *Applied Economics Letters*, 2024, acceso anticipado. doi: 10.1080/13504851.2024.2343834.

[21] IBM, "¿Qué es el análisis de componentes principales (PCA)?", IBM, 2023. [En línea]. Disponible en: <https://www.ibm.com/es-es/topics/principal-component-analysis>.

[22] JA Angarita Donado y W. Pérez Bermúdez, "Comportamiento tarifario de energías renovables tipo eólica y solar en la generación de energía eléctrica en Colombia", Repositorio Institucional Universidad EAN, Bogotá, Colombia, 2021. Disponible en: <https://repository.universidadean.edu.co/bitstream/handle/10882/11402/PerezWalter2021.pdf?sequence=1>.

[23] JE Duarte Aunta, "Determinar la variabilidad de la demanda de energía eléctrica que permita evaluar el potencial uso de un esquema tarifario diferencial", Repositorio Institucional Universidad Nacional de Colombia, Bogotá, Colombia, 2023. Disponible en: <https://repositorio.unal.edu.co/handle/unal/86066?show=full>.

[24] Ministerio de Minas y Energía de la República de Colombia, Subdirección de Hidrocarburos, "Proyección de precios de los energéticos para generación eléctrica", Unidad de Planeación Minero-Energética (UPME), Bogotá, Colombia, 2023. Disponible en: https://www1.upme.gov.co/sipg/Publicaciones_SIPG/Proyeccion_precios_energeticos_I_semestre_2023_vf.pdf.

[25] J. Barrientos, E. Rodas, E. Velilla, M. Lopera y F. Villada, "Modelo para el pronóstico del precio de la energía eléctrica en Colombia", *Lecturas De Economía*, núm. 77, págs. 91-127, 2012. doi: 10.17533/udea.le.n0a14771.

[26] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*. Berlin, Heidelberg: Springer, 2005.

[27] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, y G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. Hoboken, NJ: John Wiley & Sons, 2015.

[28] T. Chen y C. Guestrin, "XGBoost: A Scalable Tree Boosting System," en *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 785–794.