

Evaluation of the Economic Impact of the Oil Sector in Peru through an Autoregressive Model (VAR)

Anthony Cruzado¹, Nhayrin Ruiz¹, Joseph Sinchitullo¹

¹Universidad Nacional de Ingeniería, Perú

acruzados@fip.uni.edu.pe, nruizn@fip.uni.edu.pe, jsinchitullo@uni.pe

Abstract– This study examines the interrelationship between macroeconomic and oil sector variables using a Vector Autoregressive (VAR) model implemented in RStudio, with the aim of identifying patterns and dependencies that can contribute to a better understanding of the energy market. Data were collected and processed from April 2014 to December 2019, allowing key variables related to oil production, inflation, and fuel sales to be modeled and predicted based on Gross Domestic Product (GDP). Through the analysis of coefficients estimated in the VAR model, direct and inverse relationships between economic and energy variables were identified, evidencing their impact on economic stability and strategic planning of the sector. The results highlight the importance of considering these factors in the formulation of public policies and business strategies, promoting a data-driven approach to decision-making in highly volatile environments. This study provides a valuable quantitative tool to assess the effects of fluctuations in the oil sector on the national economy and suggests future lines of research to improve the accuracy of predictive models in uncertain scenarios.

Keywords: VAR, variables, relationship, coefficients, analysis.

Evaluación del Impacto Económico del Sector Petrolero en el Perú mediante un Modelo Autoregresivo (VAR)

Anthony Cruzado¹, Nhayrin Ruiz¹, Joseph Sinchitullo¹

¹Universidad Nacional de Ingeniería, Perú

acruzados@fip.uni.edu.pe, nruizn@fip.uni.edu.pe, jsinchitullo@uni.pe

Resumen— Este estudio examina la interrelación entre variables macroeconómicas y del sector petrolero mediante un modelo Vector Autorregresivo (VAR) implementado en RStudio, con el objetivo de identificar patrones y dependencias que puedan contribuir a una mejor comprensión del mercado energético. Se recopilaron y procesaron datos desde abril de 2014 hasta diciembre de 2019, lo que permitió modelar y predecir variables clave relacionadas con la producción petrolera, la inflación y las ventas de combustibles en función del Producto Bruto Interno (PBI). A través del análisis de coeficientes estimados en el modelo VAR, se identificaron relaciones directas e inversas entre las variables económicas y energéticas, evidenciando su impacto en la estabilidad económica y la planificación estratégica del sector. Los resultados destacan la importancia de considerar estos factores en la formulación de políticas públicas y estrategias empresariales, promoviendo un enfoque basado en datos para la toma de decisiones en entornos de alta volatilidad. Este estudio proporciona una herramienta cuantitativa valiosa para evaluar los efectos de las fluctuaciones del sector petrolero en la economía nacional y sugiere futuras líneas de investigación para mejorar la precisión de los modelos predictivos en escenarios de incertidumbre.

Palabras clave: VAR, variables, relación, coeficientes, análisis.

I. INTRODUCCIÓN

El petróleo desempeña un papel fundamental en la economía global, tanto por su importancia como fuente de energía como por su influencia en la industria y el comercio internacional. Su precio y disponibilidad afectan directamente diversas actividades económicas, influyendo en la inflación, la balanza comercial y el crecimiento económico de los países productores y consumidores. Dada esta interconexión, comprender la dinámica entre las variables del sector petrolero y las variables macroeconómicas resulta crucial para el diseño de políticas eficientes y estrategias de mitigación de riesgos.

Este estudio aplica un modelo autorregresivo vectorial (VAR) para analizar las relaciones dinámicas entre indicadores del sector petrolero y variables macroeconómicas en el contexto peruano, considerando datos mensuales del período comprendido entre abril de 2014 y diciembre de 2019. La elección del modelo VAR se fundamenta en su capacidad para capturar la interdependencia temporal entre múltiples variables sin necesidad de imponer relaciones causales previas, a diferencia de los modelos univariados (como ARIMA) que se centran en una sola serie temporal, o los modelos estructurales que requieren supuestos teóricos

restrictivos sobre la dirección de las relaciones [12].

Además de contribuir al entendimiento del comportamiento del sector en un país altamente dependiente de la renta de recursos naturales, este estudio busca proporcionar una herramienta cuantitativa para entender los impactos de las fluctuaciones del sector petrolero en la economía nacional y apoyar la toma de decisiones en materia energética y económica.

II. METODOLOGÍA

Se recopilaron los datos de las variables seleccionadas desde abril del 2014 hasta diciembre del 2021. Datos extraídos del Banco central de reservas del Perú (BCRP) [1], del Ministerio de Energías y Minas (Minem) [2] y del Banco de datos de Perupetro [3]. Utilizando, para las predicciones, los datos desde abril del 2014 a diciembre del 2019. Las variables seleccionadas fueron Productos Terminados, Producción Total de Combustibles, Producción de Petróleo, Inventarios de Petróleo, Producto Bruto Interno Real, Producto Bruto Interno del sector Hidrocarburos, Ingresos del Gobierno, Impuesto Selectivo al Consumo de Combustibles, Bienes de consumo Importados, Exportaciones de Gas Natural, Importaciones, Cargas Importadas de Crudo, Cargas Totales (Crudo Nacional e Importado y Combustibles en Proceso y Terminados) y Cargas de Derivados de Crudo en Proceso. Las cuales se analizaron mediante la aplicación del modelo Vector Autoregresivo (VAR) en el programa RStudio, el procedimiento para la realización del proyecto se encuentra representado en el diagrama de flujo (Anexo1), basándose en las siguientes pruebas.

1. Prueba de Normalidad

Para garantizar la validez del modelo VAR, es necesario verificar que las variables utilizadas cumplan con ciertos supuestos estadísticos, entre ellos la normalidad. En este sentido, se emplearon pruebas de bondad de ajuste con el propósito de evaluar si los datos observados provienen de una distribución normal. Estas pruebas permiten identificar la distribución que mejor se ajusta a los datos, lo cual es fundamental para seleccionar adecuadamente las pruebas estadísticas a utilizar, sean estas paramétricas o no paramétricas. Asegurar la normalidad de las variables es un paso crucial, ya que el modelo VAR requiere este supuesto para una estimación eficiente y una interpretación confiable de los resultados.

1.1. Prueba de Normalidad Shapiro-Wilks en RStudio

El método consiste en comenzar ordenando la muestra de menor a mayor valor, obteniendo el nuevo vector muestral. Se rechaza la hipótesis nula de normalidad si el estadístico Shapiro-Wilk es menor que el valor crítico proporcionado por la tabla elaborada por los autores para el tamaño y el nivel de significancia dado [4].

1.2. Transformación de BoxCox en RStudio

Si mediante la prueba de Shapiro-Wilk, se concluye que los datos no provienen de una distribución normal, estos se transforman a normales por la transformación de BoxCox.

2. Prueba de estacionariedad

La prueba de estacionariedad es una herramienta clave en el análisis de series temporales que permite verificar si una serie mantiene una media y una varianza constantes en el tiempo. Este concepto es fundamental para numerosos modelos, como el modelo VAR, y métodos de análisis de series temporales, ya que garantiza la coherencia de los resultados y posibilita el uso de técnicas estadísticas convencionales.

2.1. Prueba Kwiatkowski-Phillips-Schmidt-Shin

A diferencia de la prueba de Dickey-Fuller, que evalúa si una serie temporal contiene una raíz unitaria, la prueba KPSS se emplea para determinar si una serie es estacionaria en torno a una tendencia determinista.

3. Prueba de Existencia de Raíz Unitaria

La ausencia de raíz unitaria en una serie temporal es crucial porque garantiza su estacionariedad, lo que significa que sus propiedades estadísticas, como la media y la varianza, permanecen constantes en el tiempo. Esto es fundamental para evitar relaciones espurias entre variables, asegurar la validez de las inferencias estadísticas y permitir modelos como ARIMA o VAR, que requieren estacionariedad para ser efectivos. Además, la estacionariedad facilita predicciones más confiables, ya que las series con raíz unitaria suelen ser impredecibles y sensibles a los shocks.

3.1. Dickey-Fuller

Partimos del Random Walk with Drift

$$Y_t = \rho Y_{t-1} + u \quad ; \quad -1 < \rho < 1 \quad (3)$$

donde:

u: Término de error de ruido blanco.

Si $\rho=1$ la ecuación se convierte en RWM. Si ρ es de hecho 1, nos enfrentamos a lo que se conoce como problema de raíz unitaria, es decir, una situación de no estacionariedad, ya sabemos que en este caso la varianza Y , no es estacionaria. Sin embargo, si $\rho < 1$ es decir, si el valor absoluto de ρ es menor que uno, entonces se puede demostrar que la serie temporal Y , es estacionaria en el sentido que la hemos definido.

4. Prueba de Causalidad de Granger

La prueba de causalidad en el sentido de Granger se aplica para analizar la relación dinámica entre variables temporales. Esta prueba evalúa si los rezagos de una variable explicativa,

denotada como z , contribuyen de manera significativa a la predicción de otra variable dependiente, y . Bajo la hipótesis nula, se asume que z no causa a y en el sentido de Granger. En el contexto del modelo VAR, se requiere que las variables incluidas mantengan una relación de causalidad mutua. Por ello, esta prueba se emplea como criterio fundamental para la selección de variables, asegurando que todas las series consideradas aporten información relevante al sistema dinámico.

4.1. Hipótesis

H_0 : X no causa a Y en el sentido de Granger ($\gamma_j=0$ para todos los j)

H_a : X causa a Y en el sentido de Granger ($\gamma_j \neq 0$ para al menos un j)

4.2. Estadístico F

La prueba utiliza un estadístico F para comparar los 2 modelos (restringido y no restringido). Si el valor F calculado es menor al valor F crítico en el nivel de significancia elegido, rechazamos la hipótesis nula, indicando que X causa a Y.

5. Modelo VAR

En el análisis de regresión que involucra datos de series de tiempo, si el modelo de regresión incluye no solo los valores actuales sino también los valores rezagados de las variables explicativas, se denomina modelo de rezago distribuido. Si el modelo incluye uno o más valores rezagados de las variables dependientes entre sus variables explicativas, se denomina modelo autorregresivo [6]. Una vez seleccionadas las variables pertinentes mediante los criterios estadísticos correspondientes, se realiza la estimación del modelo VAR lo que permite analizar las interrelaciones dinámicas entre las series temporales seleccionadas.

6. Prueba de estabilidad del modelo VAR

La estabilidad es importante porque indica si las relaciones entre las variables son consistentes y si el modelo VAR es adecuado para hacer predicciones a largo plazo. Para realizar la prueba de estabilidad se debe identificar si las raíces del polinomio característico están dentro del circuito unitario en el plano complejo. Si estas están dentro del círculo unitario, se dice que el modelo VAR es estable.

7. Prueba de correlación serial del modelo VAR

La prueba de correlación serial en el contexto de un modelo VAR se utiliza para verificar si hay autocorrelación residual en los términos de error del modelo después de haber estimado el mismo. La presencia de autocorrelación serial en los residuos del modelo indica que aún queda información sistemática sin capturar en el modelo y puede afectar la precisión de las inferencias y predicciones realizadas con el mismo.

8. Prueba de Normalidad en los residuales

La prueba de normalidad evalúa si los residuos del modelo VAR se distribuyen de manera simétrica alrededor de su media y si tienen colas que se asemejan a la distribución normal estándar. Una distribución normal de los residuos

significa que los errores aleatorios del modelo no tienen ningún patrón sistemático y que el modelo es adecuado para capturar la variabilidad en los datos.

9. Prueba de Homocedasticidad

Se verifica si la varianza de los errores se mantiene constante a lo largo del tiempo mediante pruebas de homocedasticidad. Este supuesto es fundamental en la modelización estadística, ya que garantiza la validez de las inferencias y predicciones generadas por el modelo. Una vez que la ecuación VAR cumple con todas las pruebas diagnósticas requeridas, se procede a realizar la predicción de cada variable incluida en el modelo y al análisis de los coeficientes estimados en sus respectivas ecuaciones, lo que permite interpretar la dinámica entre las variables del sistema.

III. RESULTADOS Y DISCUSIONES

Se obtuvieron las ecuaciones en base al modelo VAR y las predicciones de las siguientes variables.

1. Resultados

1.1 Crudo Importado

TABLA I

RESULTADOS OBTENIDOS DE LA VARIABLE CRUDO IMPORTADO

Variable	Carga Importada	
Dependencia	Directa	Cargas de Derivados de Crudo en Proceso
	Inversa	ISC Combustibles
Valores	$p = 0.006356$ $R^2_{multiple} = 0.7996$ $R^2_{ajustado} = 0.5122$	



Fig. 1 Gráficas del Valor Real y Valores Predictivos de la variable Crudo Importado.

La TABLA I presenta la relación de la carga importada de crudo, mostrando una dependencia directa con las cargas de derivados en proceso e inversa con el Índice Selectivo al Consumo (ISC) de combustibles. Los valores estadísticos indican que el modelo tiene una capacidad moderada de explicación. En la Fig. 1, la serie real (línea azul) muestra

una caída significativa, mientras que la predicción (línea verde) mantiene una tendencia más estable.

1.2 Cargas Totales

TABLA II
RESULTADOS OBTENIDOS DE LA VARIABLE CARGAS TOTALES

Variable	Cargas Totales	
Dependencia	Directa	Producción Total de Combustible
	Inversa	Importaciones
Valores	$p = 0.000666$ $R^2_{multiple} = 0.7979$ $R^2_{ajustado} = 0.5734$	

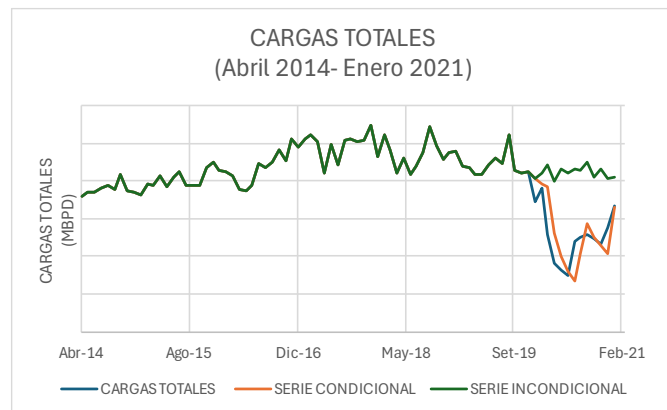


Fig. 2 Gráficas del Valor Real y Valores Predictivos de la variable Cargas Totales.

La TABLA II presenta la relación de las cargas totales, mostrando una dependencia directa con la producción total de combustible e inversa con las importaciones. Los valores estadísticos indican que el modelo tiene una capacidad moderada de explicación. En la Fig. 2, la serie real (línea azul) muestra una caída significativa, mientras que la predicción (línea verde) mantiene una tendencia más estable.

1.3 Ventas Totales de Combustibles

TABLA III
RESULTADOS OBTENIDOS DE LA VARIABLE VENTAS TOTALES

Variable	Ventas Totales	
Dependencia	Directa	Bienes de Consumo
		Importaciones
		PBI Real
Valores	$p = 8.991 \times 10^{-5}$ $R^2_{multiple} = 0.7196$ $R^2_{ajustado} = 0.5378$	

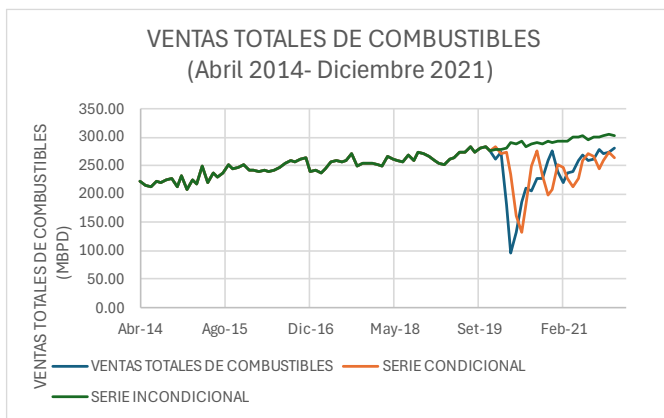


Fig. 3 Gráficas del Valor Real y Valores Predictivos de la variable Ventas Totales de Combustibles.

La TABLA III presenta la relación de las ventas totales de combustibles, mostrando una dependencia directa con los bienes de consumo, las importaciones y el PBI real. Los valores estadísticos indican que el modelo tiene una capacidad moderada de explicación. En la Fig. 3, la serie real (línea azul) muestra una caída significativa, mientras que la predicción (línea verde) presenta un incremento en su trayectoria.

1.4 Producción de Petróleo

TABLA IV
RESULTADOS OBTENIDOS DE LA
VARIABLE PRODUCCIÓN DE PETRÓLEO

Variable	Productos de Petróleo	
Dependencia	Inversa	Inflación
Valores	$p = 0.002757$ $R^2_{multiple} = 0.6933$ $R^2_{ajustado} = 0.4558$	

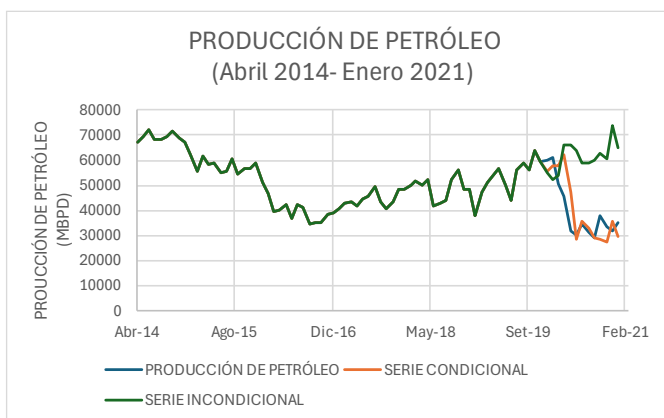


Fig. 4 Gráficas del Valor Real y Valores Predictivos de la variable Producción de Petróleo.

La TABLA IV presenta los resultados del modelo para la producción de petróleo, indicando una relación inversa con la inflación. Los valores estadísticos sugieren una capacidad moderada del modelo para explicar la variabilidad. Por otro lado, la

Fig. 4 muestra que la producción de petróleo (línea azul) ha disminuido, mientras que la proyección del modelo (línea verde) indica un incremento.

2. Discusiones

Se analizaron los coeficientes estimados para cada variable en el modelo VAR, observando la presencia de signos tanto positivos como negativos, lo que refleja diferentes tipos de relaciones entre las variables. Un coeficiente positivo indica una relación directa, un incremento en la variable explicativa en periodos anteriores tiende a provocar un aumento en la variable dependiente en el periodo actual. Este tipo de relación sugiere que las variables se refuerzan mutuamente, donde el crecimiento en una variable impulsa el crecimiento en otra. Por el contrario, un coeficiente negativo señala una relación inversa o indirecta; es decir, un aumento en la variable explicativa en periodos anteriores conduce a una disminución en la variable dependiente en el periodo actual. Esto indica que las variables se comportan de manera opuesta, como podría ser el caso de la relación entre las variables Ventas Totales de Combustible e Importaciones; donde, si las ventas de combustible aumentan debido a una mayor producción interna, se reduce la necesidad de importar combustible.

2.1 Ventas totales de combustible

A partir de la ecuación VAR obtenida para la variable Ventas Totales de Combustible se analizaron los coeficientes estimados de las variables relacionadas.

TABLA V
COEFICIENTES ESTIMADOS DE LA
ECUACIÓN DE LA VARIABLE VENTAS
TOTALES

	Ventas Totales de combustible	PBI Real	Importaciones	Bienes de consumo
Ventas Totales de combustible	-1.2358268	0.381218	1.98872723	1.80376043
PBI Real	0.42616373	-3.31438	25.1381543	2.76329641
Importaciones	0.01275631	0.012160	-0.0140527	0.23336209
Bienes de consumo	0.13237770	0.120739	-2.0523087	-1.0821234

El aumento en las ventas de combustibles impulsaría la actividad económica, incrementando el PBI real y la demanda de bienes importados, especialmente las importaciones de bienes de consumo. Dado que el Perú es importador de crudo y su producción de combustibles no abastece la demanda interna, este crecimiento también resultaría en mayores importaciones de petróleo y derivados.

Durante el primer semestre de 2020, el modelo refleja con claridad los efectos de la contracción económica

ocasionada por la pandemia. La fuerte caída en las ventas de combustibles —que alcanzaron su punto más bajo en abril (94.89 MBPD)— coincide con una contracción histórica del PBI (-39.9% en abril) y una reducción de las importaciones de bienes de consumo. Posteriormente, con la implementación progresiva del plan de reactivación, se observa una recuperación gradual tanto en las ventas de combustibles como en la actividad económica, lo cual confirma la sensibilidad de estas variables ante choques externos y medidas de política.

2.2 Producción de Petróleo

TABLA VI
COEFICIENTES ESTIMADOS DE LA
ECUACIÓN DE LA VARIABLE VENTAS
TOTALES

	Producción de petróleo	Inflación
Producción de petróleo	0.00397518	-0.025885
Inflación	-0.11547632	-1.7300134

Para la relación entre la producción de petróleo y la inflación se tiene una dependencia inversa. Cuando aumenta la producción de petróleo, la inflación tiende a disminuir. Esto se debe a que una mayor producción de petróleo permite abastecer a las refinerías, reduciendo la dependencia de las importaciones de combustibles.

En el caso peruano, al importar menos combustibles, disminuyen los costos logísticos, los aranceles y otros gastos asociados, lo que se traduce en precios más bajos para los consumidores.

A partir del modelo VAR estimado, se identifican relaciones significativas entre las ventas totales de combustibles, el PBI real, las importaciones y la producción nacional de petróleo. En particular, los coeficientes estimados muestran que un incremento en las ventas de combustibles está asociado a una expansión de la actividad económica y una mayor demanda de bienes importados. Este efecto es especialmente relevante en el contexto peruano, donde la producción interna no satisface la demanda, generando una alta dependencia de las importaciones de petróleo y derivados.

Durante el primer semestre de 2020, el modelo refleja con claridad los efectos de la contracción económica ocasionada por la pandemia. La fuerte caída en las ventas de combustibles, que alcanzaron su punto más bajo en abril (94.89 MBPD), coincide con una contracción histórica del PBI (-39.9% en abril) y una reducción de las importaciones de bienes de consumo. Posteriormente, con la implementación progresiva del plan de reactivación, se observa una recuperación gradual tanto en las ventas de combustibles como en la actividad económica, lo cual confirma la sensibilidad de estas variables ante choques externos y medidas de política.

En cuanto a la relación entre producción de petróleo e

inflación, el modelo evidencia una dependencia inversa: mayores niveles de producción tienden a contribuir a la reducción de la inflación, al disminuir la necesidad de importaciones y los costos logísticos asociados. No obstante, este comportamiento no se mantiene de forma constante en todos los periodos analizados. Factores como la interrupción de operaciones en lotes estratégicos (como el 95 y el 192), conflictos sociales y choques en los precios internacionales introducen variabilidad que afecta la robustez de esta relación.

A pesar de la tendencia general identificada, en varios momentos se observan desviaciones debidas a dinámicas no capturadas completamente por el modelo, como alteraciones estacionales, restricciones logísticas o shocks exógenos vinculados a precios internacionales y al tipo de cambio. Asimismo, la evolución de la inflación muestra influencia significativa del componente alimentario, lo cual puede mitigar o contrarrestar el efecto de la producción petrolera.

Estos hallazgos subrayan la utilidad del modelo VAR para analizar interrelaciones dinámicas entre variables macroeconómicas y del sector hidrocarburos, proporcionando una base cuantitativa para la toma de decisiones en política energética y económica. No obstante, el modelo presenta algunas limitaciones, como la sensibilidad a cambios estructurales en el periodo de estudio y la omisión de variables externas relevantes, como el tipo de cambio, la inversión pública o la política fiscal.

IV. CONCLUSIONES

Los resultados obtenidos mediante la aplicación del modelo VAR evidenciaron relaciones significativas entre las variables económicas y del sector petrolero. Se observó que las ventas totales de combustibles tienen una fuerte correlación con el PBI, lo que sugiere que el crecimiento económico impulsa la demanda de combustibles. Además, se identificó una relación inversa entre la producción de petróleo y la inflación, indicando que un aumento en la producción podría contribuir a la estabilidad de los precios al reducir la dependencia de importaciones energéticas.

El análisis de coeficientes del modelo mostró que las importaciones de petróleo y derivados tienen un impacto directo en la oferta total de combustibles, mientras que la variabilidad en los inventarios de petróleo afecta de manera significativa la estabilidad de los precios. La aplicación del modelo VAR permitió predecir con alta precisión la evolución de variables clave, proporcionando información valiosa para la toma de decisiones estratégicas en el sector energético.

Estos hallazgos resaltan la importancia de la planificación basada en modelos cuantitativos para mitigar el impacto de las fluctuaciones en el mercado petrolero y mejorar la formulación de políticas públicas. Se recomienda continuar con estudios que integren factores exógenos adicionales,

como eventos geopolíticos y políticas fiscales, para optimizar la capacidad predictiva del modelo y fortalecer la estabilidad del sector energético en el país.

Como línea futura, se sugiere ampliar el modelo integrando variables exógenas a través de enfoques como el VARX o incorporar mecanismos de corrección de error (VECM) si se confirma cointegración entre las series. También sería valioso explorar modelos estructurales o de series temporales no lineales para capturar mejor las interrupciones abruptas, como las provocadas por crisis sanitarias o conflictos sociales.

VII. RECOMENDACIONES

Para mejorar la precisión de las predicciones y la robustez del modelo VAR, se recomienda ampliar la cantidad de datos utilizados, tanto en términos de variables como de horizonte temporal. La recopilación de datos con frecuencia diaria, en lugar de mensual, permitiría un análisis más detallado de las fluctuaciones del mercado y su impacto económico. Además, incorporar variables como política fiscal, eventos geopolíticos y variaciones en los precios internacionales del petróleo fortalecería la capacidad predictiva del modelo. La metodología también podría adaptarse para evaluar la influencia de factores económicos en sectores estratégicos como la minería y la energía renovable. Con base en los resultados del estudio, se sugiere realizar simulaciones y evaluar políticas de subsidios y regulación del mercado de combustibles, garantizando así un abastecimiento eficiente y estable.

REFERENCIAS

- [1] BCRP, *Base de Datos de Estadísticas* (2024).
- [2] Ministerio de Energía y Minas. *Banco de datos*. Perú (2024).
- [3] Perupetro. *Banco de datos*, Perú (2024).
- [4] Alfonso Novales *Modelos Vectoriales Autorregresivos (VAR)* Universidad Complutense de Madrid (2014).
- [5] Pengfei Li *Box-Cox Transformations: An Overview* University of Connecticut (2005).
- [6] D. N. Gujarati y D. C. Porter, *Basic Econometrics*, 5th ed. New York: McGraw-Hill Education, 2009.
- [7] Banco Central de Reserva del Perú, *Reporte de inflación: Panorama actual y proyecciones macroeconómicas, 2021-2022*, marzo 2021.
- [8] Banco Central de Reserva del Perú, *Reporte de inflación: Panorama actual y proyecciones macroeconómicas 2020-2021*, junio 2020.
- [9] Banco Central de Reserva del Perú, "Impacto económico de la COVID-19 en las regiones del Perú", Nota de Estudios, 76-2020, 2020.
- [10] Banco Central de Reserva del Perú, "El impacto del gasto público en las regiones: Resultados recientes", Nota de Estudios, 02-2021, 2021.
- [11] J. H. Stock y M. M. Watson, *Introducción a la Econometría*, 3ª ed. México: Pearson Educación, 2011.
- [12] Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 48(1), 1-48

Anexo 1. Diagrama de flujo



Anexo 2. Métodos utilizados.

1. Prueba de Normalidad

1.1. Prueba de Normalidad Shapiro-Wilks

Esta prueba se emplea para contrastar normalidad cuando el tamaño de la muestra es mayor a 3 y menor a 5000 observaciones. El método consiste en comenzar ordenando la muestra de menor a mayor valor, obteniendo el nuevo vector muestral. Se rechaza la hipótesis nula de normalidad si el estadístico Shapiro-Wilk es menor que el valor crítico proporcionado por la tabla elaborada por los autores para el tamaño y el nivel de significancia dado [4]. El estadístico de prueba es el siguiente:

$$W = \frac{\sum_{i=1}^n \alpha_i y_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (1)$$

donde:

n: Tamaño de la muestra.

y_i : Valores ordenados de la muestra.

\bar{y} : Media de la muestra.

α_i : Coeficientes específicos de la prueba de Shapiro-Wilks.

1.3. Transformación de Box-Cox

Si mediante la prueba de Shapiro-Wilk, se concluye que los datos no provienen de una distribución normal, estos se transforman a normales por la transformación de BoxCox. La fórmula de la transformación de BoxCox es la siguiente:

$$y(\lambda) = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \text{si } \lambda \neq 0 \\ \log \lambda, & \text{si } \lambda = 0 \end{cases} \quad (2)$$

donde:

y: Transformación de BoxCox.

λ : Parámetro para la transformación.

2. Prueba de estacionariedad

2.1. Prueba Kwiatkowski-Phillips-Schmidt-Shin

A diferencia de la prueba de Dickey-Fuller, que evalúa si una serie temporal contiene una raíz unitaria, la prueba KPSS se emplea para determinar si una serie es estacionaria en torno a una tendencia determinista.

3. Prueba de Existencia de Raíz Unitaria

3.1. Dickey-Fuller

Partimos del Random Walk with Drift

$$Y_t = \rho Y_{t-1} + u \quad ; \quad -1 < \rho < 1 \quad (3)$$

donde:

u: Término de error de ruido blanco.

Si $\rho=1$ la ecuación se convierte en RWM. Si ρ es de hecho 1, nos enfrentamos a lo que se conoce como problema de raíz unitaria, es decir, una situación de no estacionariedad, ya sabemos que en este caso la varianza Y, no es estacionaria. Sin embargo, si $\rho < 1$ es decir, si el valor absoluto de ρ es menor que uno, entonces se puede demostrar que la serie temporal Y, es estacionaria en el sentido que la hemos definido.

Restando Y_{t-1} a cada lado de la ecuación

$$Y_t - Y_{t-1} = \rho Y_{t-1} - Y_{t-1} + u$$

$$\Delta Y_t = (\rho - 1)Y_{t-1} + u$$

$$\Delta Y_t = \delta Y_{t-1} + u \quad (4)$$

donde $\delta = \rho - 1$.

Por lo tanto, en la práctica, en lugar de estimar la ecuación, estimamos la ecuación y probar la hipótesis (nula) de que $\delta=0$, siendo la hipótesis alternativa que $\delta < 0$. Si $\delta < 0$ entonces $\rho=1$, es decir, tenemos una raíz unitaria, lo que significa que la serie temporal considerada no es estacionaria [6].

Antes de proceder a estimar la ecuación se puede observar que si $\delta=0$, la ecuación se convertirá en

$$\Delta Y_t = (Y_t - Y_{t-1}) = u_t \quad (5)$$

Dado que u, es un término de error de ruido blanco, es estacionario, lo que significa que las primeras diferencias de las series temporales de paseo aleatorio son estacionarias, un punto que ya hemos señalado antes [6].

Dickey y Fuller han demostrado que bajo la hipótesis nula de que $\delta=0$, el valor t estimado del coeficiente de Y_{t-1}

en la ecuación sigue el estadístico τ (tau). Estos autores han calculado los valores críticos del estadístico tau basándose en simulaciones de Monte Carlo [6].

3.2. Dickey-Fuller Aumentado

Al realizar la prueba DF como en las ecuaciones, se supuso que el término de error u, no estaba correlacionado. Pero en caso de que las u, estén correlacionadas, Dickey y Fuller han desarrollado otra prueba, conocida como prueba Dickey-Fuller aumentada (ADF). Esta prueba se realiza "aumentando" las tres ecuaciones anteriores sumando los valores rezagados de la variable dependiente ΔY_t . Para ser específicos, supongamos que usamos la ecuación. La prueba ADF aquí consiste en estimar la siguiente regresión [6].

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \sum_{i=1}^m \alpha_i \Delta Y_{t-i} + \varepsilon_i \quad (6)$$

4. Prueba de Causalidad de Granger

Originalmente, Granger planteaba que una variable z se considera causal si la predicción de y basada en los valores pasados de ambas variables (y y z) es más precisa, es decir, tiene un menor error, que la predicción basada únicamente en los valores pasados de y. En este contexto, si añadir z no mejora las predicciones, se concluye que z no causa a y en el sentido de Granger.

$$E(y_t / y_{t-1}, y_{t-2}, \dots ; z_{t-1}, z_{t-2}, \dots) = E(y_t / y_{t-1}, y_{t-2}, \dots) \quad (7)$$

La fórmula para la causalidad de Granger es la siguiente:

Modelo de regresión restringido (sin la serie X_t como predictor):

$$Y_t = \alpha_0 + \sum_{i=1}^p \beta_i Y_{t-i} + \varepsilon_t \quad (8)$$

Modelo de regresión restringido (con la serie X_t como predictor adicional):

$$Y_t = \alpha_0 + \sum_{i=1}^p \beta_i Y_{t-i} + \sum_{j=1}^q \gamma_j X_{t-j} + \varepsilon_t \quad (9)$$

donde:

Y_t : variable dependiente

X_t : variable independiente

p y q: número de rezagos de Y y X, respectivamente

ε_t : término de error

4.1. Hipótesis

H_0 : X no causa a Y en el sentido de Granger ($\gamma_j=0$ para todos los j)

H_a : X causa a Y en el sentido de Granger ($\gamma_j \neq 0$ para al menos un j)

4.2. Estadístico F

La prueba utiliza un estadístico F para comparar los 2 modelos (restringido y no restringido)

$$F = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(T - k)} \quad (10)$$

donde:

RSS_R : Suma de residuos al cuadrado del modelo restringido

RSS_{UR} : Suma de residuos al cuadrado del modelo no restringido

q: Número de coeficientes restringidos (Coeficientes de X)

T: Número total de observaciones

k: Número total de parámetros estimados en el modelo no restringido

Si el valor F calculado es menor al valor F crítico en el nivel de significancia elegido, rechazamos la hipótesis nula, indicando que X causa a Y.

5. Modelo VAR

En el análisis de regresión que involucra datos de series de tiempo, si el modelo de regresión incluye no solo los valores actuales sino también los valores rezagados de las variables explicativas, se denomina modelo de rezago distribuido. Si el modelo incluye uno o más valores rezagados de las variables dependientes entre sus variables explicativas, se denomina modelo autorregresivo [6]. De este modo:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + u_t \quad (11)$$

Representa un modelo de retraso distribuido, mientras que

$$Y_t = \alpha + \beta X_t + \gamma Y_t + u_t \quad (12)$$

Es un ejemplo de modelo autorregresivo.

Un vector autorregresivo (VAR) es un conjunto de k regresiones de series temporales, en las que los regresores son los valores retardados de las k series. Un VAR extiende el modelo autorregresivo univariante a una lista, o «vector», de variables de series temporales. Cuando el número de retardos en cada una de las ecuaciones es el mismo y es igual a p, el sistema de ecuaciones se denomina VAR(p) (Stock and Watson, 2012).

$$Y_t = \beta_{10} + \beta_{11} Y_{t-1} + \dots + \beta_{1p} Y_{t-p} + \gamma_{11} X_{t-1} \dots + \gamma_{1p} X_{t-p} + u_{1t} \quad (13)$$

$$X_t = \beta_{20} + \beta_{21} Y_{t-1} + \dots + \beta_{2p} Y_{t-p} + \gamma_{21} X_{t-1} \dots + \gamma_{2p} X_{t-p} + u_{2t} \quad (14)$$

donde:

β, γ : Coeficientes desconocidos.

u_{1t}, u_{2t} : Términos de error.

En el programa RStudio, se realizó la modelación VAR gracias al comando VAR (), el cual nos muestra los coeficientes y la constante.