Pronóstico del Índice de Producción Industrial del sector de extracción de petróleo crudo y gas natural en Colombia mediante modelos ARIMA y BSTS

Diego Ruiz-Solarte[1], y Heivar Rodríguez-Pinzón[1]

[1]Fundación Universitaria Los Libertadores

Aceptado: 2025

Resumen

Contexto. El Índice de Producción Industrial (IPI) del sector de extracción de petróleo crudo y gas natural representa el 15 % del IPI total colombiano y ha exhibido alta volatilidad histórica ante choques internacionales como la caída de precios de 2014-2015 y la pandemia de COVID-19. En junio de 2025, el sector registró una variación anual de $-5.9\,\%$, evidenciando una tendencia contractiva sostenida. Este contexto hace fundamental la disponibilidad de modelos de pronóstico precisos para anticipar el comportamiento del sector e informar la planificación fiscal y decisiones de inversión. **Propósito.** Este trabajo desarrolla modelos de pronóstico para el Índice de Producción Industrial del sector de extracción de petróleo crudo y gas natural en Colombia. **Metodología.** Se utilizan datos mensuales del DANE para el periodo de enero de 2014 a junio de 2025. Se implementan dos enfoques metodológicos: (i) modelos ARIMA mediante la metodología Box-Jenkins y (ii) modelos bayesianos de series temporales estructurales (BSTS). **Resultados.** Los pronósticos a 6 meses de ambos modelos confirman una contracción estructural persistente del sector, con niveles de producción proyectados aproximadamente 12-14 % por debajo del promedio histórico y sin señales de recuperación en el corto plazo. **Conclusiones.** Estos hallazgos tienen implicaciones importantes para la formulación de política económica, planificación presupuestal y toma de decisiones de inversión en el sector extractivo colombiano.

Palabras clave— Índice de Producción Industrial, series de tiempo, ARIMA, BSTS, sector petrolero colombiano.

Abstract

Context. The Industrial Production Index (IPI) for the crude oil and natural gas extraction sector represents 15% of Colombia's total IPI and has exhibited high historical volatility in response to international shocks such as the 2014-2015 price collapse and the COVID-19 pandemic. In June 2025, the sector registered an annual variation of -5.9%, evidencing a sustained contractionary trend. This context makes the availability of accurate forecasting models essential to anticipate sector behavior and inform fiscal planning and investment decisions. **Purpose.** This work develops forecasting models for the Industrial Production Index of the crude oil and natural gas extraction sector in Colombia. **Methodology.** Monthly data from DANE (Colombia's National Administrative Department of Statistics) for the period from January 2014 to June 2025 are used. Two methodological approaches are implemented: (i) ARIMA models using the Box-Jenkins methodology and (ii) Bayesian Structural Time Series (BSTS) models. **Results.** Six-month forecasts from both models confirm a persistent structural contraction of the sector, with production levels projected approximately 12-14% below the historical average and no signs of recovery in the short term. **Conclusions.** These findings have important implications for economic policy formulation, budget planning, and investment decision-making in Colombia's extractive sector.

Keywords— Industrial Production Index, time series, ARIMA, BSTS, Colombian oil sector.

1 Introducción

El Índice de Producción Industrial (IPI) es un indicador estadístico que permite comprender la evolución del sector industrial colombiano en el corto plazo. Calculado mensualmente por el DANE, el IPI mide los volúmenes de producción real con base en información que suministran los agentes industriales a través de la Encuesta Mensual Manufacturera con Enfoque Territorial (EMMET) y registros administrativos (DANE, 2020b). En 2018, el DANE amplió su alcance para incluir cuatro sectores: explotación de minas y canteras, industrias manufactureras, suministro de electricidad y gas, y captación y distribución de agua (DANE, 2020a, 2022). El sector de extracción de petróleo crudo y gas natural (identificado con el código CIIU 0610) representa el 15 % del IPI total, constituyendo uno de sus componentes más significativos (DANE, 2020a).

El sector petrolero colombiano ha exhibido alta volatilidad histórica ante choques internacionales. La caída de precios de 2014-2015, cuando el Brent pasó de USD 112 en junio de 2014 a cerca de USD 45 por barril en enero de 2015 debido al boom del shale oil estadounidense y la decisión de la OPEP de mantener alta producción, generó presiones sobre la rentabilidad del sector (Chacón Cruz

& Riaño Amaya, 2020; Marzo Carpio, 2015). Más recientemente, la pandemia de COVID-19 en 2020 provocó una contracción abrupta de la demanda global de energía, reflejándose en una caída del IPI del sector. A nivel nacional, cambios en la política energética hacia la transición renovable y la reforma tributaria de 2022, que introdujo sobretasas de 5-15 % y prohibió la deducción de regalías, generaron incertidumbre adicional (Asociación Colombiana del Petróleo y Gas, 2023). En junio de 2025, el sector registró una variación anual de -5.9 %, contribuyendo con -0.8 puntos porcentuales a la variación total del IPI (DANE, 2025). Este contexto de volatilidad y cambios estructurales hace que la disponibilidad de modelos de pronóstico precisos del IPI del sector de extracción de petróleo crudo y gas natural sea fundamental para anticipar el comportamiento del sector, informar la planificación fiscal del gobierno, orientar decisiones de inversión empresarial y evaluar la sostenibilidad de proyectos extractivos.

El desarrollo de herramientas de pronóstico confiables para series de tiempo del sector energético ha sido objeto de investigación académica, dando lugar a dos enfoques metodológicos complementarios que han demostrado efectividad en diferentes contextos. La literatura sobre pronóstico de series de tiempo en el sector energético se ha desarrollado principalmente a través de dos enfoques metodológicos complementarios. Por un lado, los modelos autorregresivos integrados de media móvil (ARIMA), propuestos por Box y Jenkins, han demostrado efectividad para pronósticos de corto plazo. Estudios en precios del petróleo (Nguyen et al., 2025; Suleiman et al., 2025; Zhao & Wang, 2014), gas natural (Akpinar & Yumusak, 2021) e índices de producción industrial (Ejaz & Iqbal, 2021; Franses & Ooms, 1997; Golinelli & Parigi, 2009) confirman su utilidad para horizontes de 1-12 meses.

Por otro lado, los modelos bayesianos de series de tiempo estructurales (BSTS), desarrollados por Scott y Varian (Scott & Varian, 2014), ofrecen una aproximación alternativa que permite descomponer la serie en componentes estructurales (tendencia, estacionalidad, regresión) e incorporar explícitamente la incertidumbre mediante distribuciones posteriores. En Lu et al., 2020, se aplicaron exitosamente modelos BSTS al análisis y pronóstico de precios del petróleo crudo, demostrando su capacidad para identificar puntos de quiebre estructurales, como la crisis financiera de 2008, y capturar la dinámica compleja del mercado petrolero mediante la integración de múltiples predictores. Más recientemente, Hamza y Mohammad, 2024 realizaron una comparación directa entre ARIMA y BSTS para el pronóstico de producción de carbón en Turquía, concluyendo que el modelo BSTS superó al ARIMA en precisión predictiva, particularmente para series con cambios estructurales evidentes.

A pesar de estos avances metodológicos, existe una brecha en la literatura nacional respecto a modelos de pronóstico específicos para el IPI del sector petrolero colombiano. Los trabajos previos se han concentrado en análisis fiscal (Chacón Cruz & Riaño Amaya, 2020), perspectivas de inversión (Asociación Colombiana del Petróleo y Gas, 2023), o comercio regional (Torres Hernández et al., 2020), pero no en modelos predictivos del IPI sectorial que incorporen tanto la metodología clásica de Box-Jenkins como los enfoques bayesianos estructurales.

El presente trabajo desarrolla modelos de pronóstico para el IPI del sector de extracción de petróleo crudo y gas natural en Colombia utilizando datos mensuales del DANE (enero 2014 - junio 2025). Se implementan dos enfoques metodológicos: (i) modelos ARIMA siguiendo la metodología Box-Jenkins, incluyendo evaluación de estacionariedad, identificación de órdenes mediante ACF/PACF, selección del mejor modelo mediante AIC y análisis de residuales; y (ii) modelos BSTS con descomposición estructural y estimación bayesiana mediante cadenas de Markov Monte Carlo (MCMC).

El resto de este artículo está dividido como sigue. La Sección 2 describe ambas metodologías empleadas; la Sección 3 presenta los resultados y pronósticos de los modelos ARIMA y BSTS, incluyendo una comparación de su desempeño predictivo; finalmente, la Sección 4 concluye con las implicaciones de los hallazgos.

2 Métodos

Esta sección describe la metodología empleada para el desarrollo de modelos de pronóstico del IPI del sector petrolero colombiano, mediante los enfoques de modelos ARIMA mediante la metodología de Box-Jenkins, y modelos bayesianos de series temporales estructurales (BSTS). La sección se organiza presentando primero la descripción de los datos, seguido por la especificación de ambos modelos, el análisis de estacionariedad, la identificación y selección de modelos candidatos, y finalmente la validación mediante diagnóstico de residuales.

2.1 Descripción de los datos

Los datos utilizados provienen del Índice de Producción Industrial (IPI) del sector de extracción de petróleo crudo y gas natural (CIIU 0610), obtenidos de los boletines mensuales del DANE para el periodo enero 2014 - junio 2025 (DANE, 2020b, 2025).

El DANE construye este índice a partir de registros administrativos de volúmenes físicos compuesto por barriles de petróleo crudo extraídos diariamente (reportados por la Agencia Nacional de Hidrocarburos) y metros cúbicos de gas natural (reportados por el Gestor del Mercado del Gas). Estos volúmenes heterogéneos se agregan mediante ponderadores que reflejan la importancia relativa de cada producto según el valor agregado del Sistema de Cuentas Nacionales de 2017, transformándose en un índice estandarizado expresado en puntos (DANE, 2020b). Es decir, un valor de 100 puntos representa el nivel promedio de producción del año base 2018; así, un IPI de 120 indica producción 20 % superior a 2018, mientras que un IPI de 85 indica producción 15 % inferior al año base.

De acuerdo con este índice, se analizará una serie de tiempo que comprende 138 observaciones mensuales con una media de 97.68 puntos y mediana de 97.13. El rango oscila entre un mínimo de 80.31 (febrero 2022) y un máximo de 121.42 (enero 2015), evidenciando una amplitud de 41.11 puntos. El rango intercuartílico (IQR) de 14.81 puntos indica una dispersión moderada, con el 50 % central de

las observaciones concentradas entre 88.88 (Q1) y 103.69 (Q3).

La Figura 1 muestra cómo se distribuye el IPI en el sector analizado en cada año del periodo 2014-2025, permitiendo comparar visualmente la evolución de la producción a lo largo del tiempo. La línea horizontal punteada verde representa el promedio general del periodo (97.68 puntos). Los años 2014 y 2015 presentan los valores más altos, con medianas cercanas a 115 puntos, muy por encima del promedio histórico. Durante 2017-2019 la producción se estabiliza alrededor de 100-103 puntos, aunque la variabilidad aumenta y aparecen valores atípicos en 2018, 2019 y 2020. A partir de 2020, todos los años muestran niveles de producción por debajo del promedio histórico, con medianas entre 86 y 92 puntos. Los años 2021 y 2025 registran los valores más bajos (alrededor de 86-87 puntos), evidenciando una reducción sostenida de la producción. La reducción progresiva de los valores año tras año confirma la tendencia decreciente observada en la serie, reflejando el proceso de disminución de la producción del sector extractivo colombiano durante la última década.

2.2 Modelos a considerar

Los modelos ARIMA, propuestos por Box y Jenkins, 1970, combinan tres componentes. El componente autorregresivo AR(p) modela la dependencia lineal respecto a valores pasados mediante la ecuación

$$Y_t = c + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t,$$

donde $\phi_1, \ldots, \phi_p$ son los coeficientes autorregresivos que capturan cuánto influye cada valor pasado en el valor actual, c es el intercepto relacionado con el nivel promedio de la serie, y ε_t es el término de error aleatorio. El componente de media móvil MA(q), expresado como

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q},$$

donde $\theta_1, \ldots, \theta_q$ son los coeficientes de media móvil y μ es la media del proceso, modela la estructura de correlación de los errores, útil para capturar perturbaciones temporales persistentes. El componente de integración I(d) transforma la serie mediante diferenciación para alcanzar estacionariedad. La especificación del modelo ARIMA(p, d, q) se realiza mediante evaluación de estacionariedad con pruebas de raíz unitaria, identificación de órdenes mediante funciones de autocorrelación (ACF) y autocorrelación parcial (PACF), estimación por máxima verosimilitud, selección mediante el criterio de información de Akaike (AIC) y validación de residuales.

Por su parte, los modelos Bayesian Structural Time Series (BSTS) descomponen la serie temporal en componentes estructurales (tendencia y estacionalidad) utilizando inferencia bayesiana (Scott & Varian, 2014). El modelo se estima mediante simulación MCMC (Markov Chain Monte Carlo), generando distribuciones posteriores que cuantifican la incertidumbre asociada a los parámetros y pronósticos. A diferencia del enfoque frecuentista de ARIMA, los modelos BSTS resultan especialmente útiles para series que experimentan cambios estructurales abruptos, como los observados en el sector petrolero durante crisis económicas, choques de oferta o cambios regulatorios súbitos. La aproximación bayesiana permite incorporar explícitamente la incertidumbre sobre estos puntos de quiebre sin necesidad de especificar su ubicación temporal de manera predeterminada, adaptándose de forma flexible a la evolución de la serie. Esta característica convierte a los modelos BSTS en una herramienta complementaria para analizar sectores económicos sujetos a alta volatilidad estructural y eventos disruptivos recurrentes.

2.3 Serie de tiempo y pruebas de estacionariedad

La Figura 2 presenta la evolución mensual del IPI del sector a analizar desde enero de 2014 hasta junio de 2025. La serie muestra una tendencia decreciente a lo largo del periodo analizado, con una contracción de aproximadamente 35 puntos desde los máximos de 2015 hasta los niveles recientes. Note que no se evidencian patrones estacionales muy marcados. Además, dicha figura sugiere la presencia de no estacionariedad, evidenciada por la tendencia decreciente y los cambios de nivel a lo largo del tiempo. Para evaluar formalmente esta característica, se aplica la prueba Aumentada de Dickey-Fuller (ADF) de raíz unitaria, considerando las siguientes hipótesis:

H_0 : La serie tiene raíz unitaria (no es estacionaria).

H_1 : La serie no tiene raíz unitaria (es estacionaria).

La prueba ADF arroja un valor p de 0.4211, por lo que no se rechaza la hipótesis nula de raíz unitaria al nivel de significancia $\alpha = 0.05$ (ver Tabla 1). Este resultado, junto con el comportamiento de la serie ilustrado en la Figura 2, confirma la presencia de no estacionariedad. Por lo tanto, se aplica diferenciación de orden $d = 1$ para transformar la serie y lograr estacionariedad.

Tabla 1: Resultados de pruebas de estacionariedad para la serie original.

Estadístico/Parámetro	**Valor**
Estadístico ADF	-2.4013
Orden de rezagos	5
Valor p	0.4095

Los resultados sobre la serie diferenciada se muestran en la Tabla 2. Con una diferenciación $d = 1$, la prueba ADF rechaza la hipótesis nula H_0, confirmando que la serie diferenciada no tiene raíz unitaria. La Figura 3 muestra la serie diferenciada, donde se evidencia una ausencia de tendencia clara y la estabilidad visual de la varianza, confirmando que la diferenciación ha logrado transformar la serie en un proceso estacionario. Esta transformación permite proceder con el análisis de modelos ARIMA o SARIMA, donde el orden de integración $d = 1$ captura la no estacionariedad observada en la serie original.

Figura 1: Diagramas de caja del IPI del sector de extracción de petróleo crudo y gas natural por año (2014-2025).

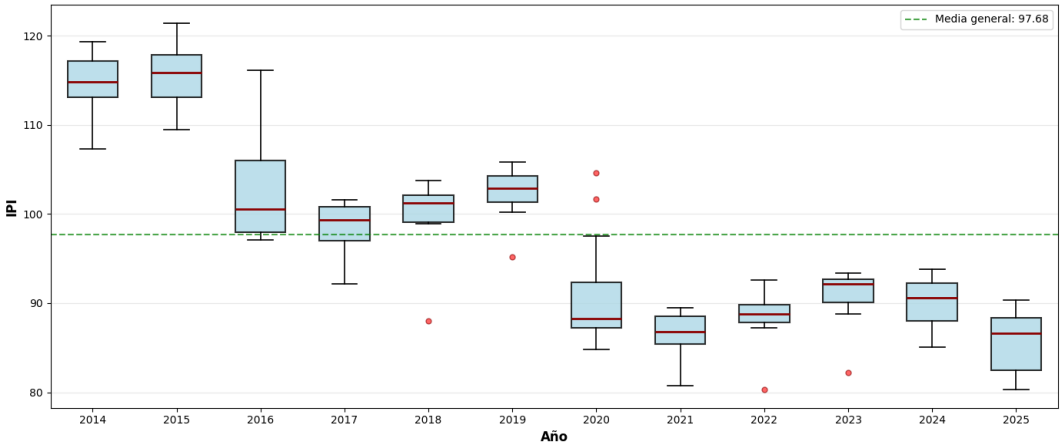


Figura 2: Serie de tiempo del IPI del sector de extracción de petróleo crudo y gas natural (2014-2025).

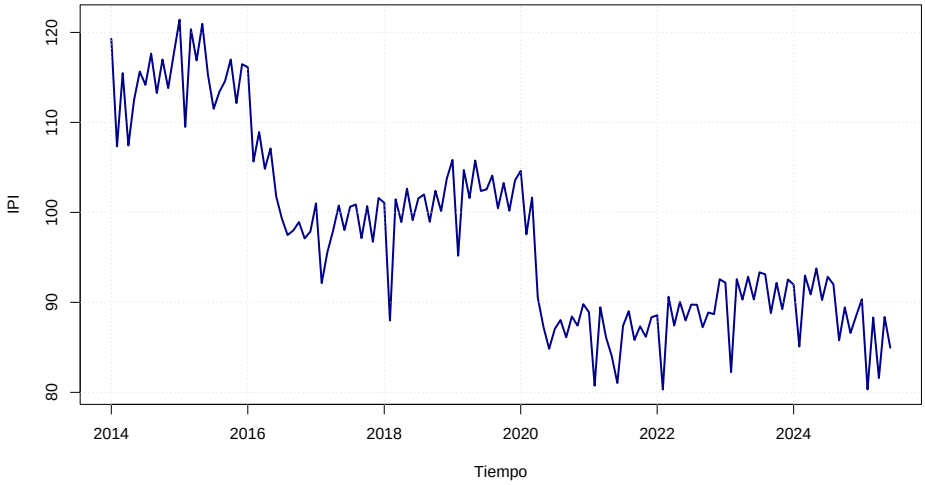


Tabla 2: Resultados de pruebas de estacionariedad luego de la primera diferenciación.

Estadístico/Parámetro	Valor
Estadístico ADF	-5.2925
Orden de rezagos	5
Valor p	0.01

2.4 Autocorrelación y autocorrelación parcial

Una vez establecido que la serie diferenciada es estacionaria, el siguiente paso consiste en examinar la estructura de dependencia temporal mediante las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF). Dado el comportamiento observado, un modelo ARIMA(p, d, q) es apropiado para capturar la dinámica temporal de los datos.

La Figura 4a muestra la función ACF de la serie diferenciada. Los rezagos significativos identificados son los lag 1, 2, 5, 6, 7, 11 y 12. El lag 1 presenta una autocorrelación negativa muy pronunciada (aproximadamente -0.6), indicando fuerte correlación inversa entre observaciones consecutivas. Los lags 5, 6 y 12 muestran autocorrelaciones positivas significativas, mientras que el lag 7 presenta autocorrelación negativa. Los valores candidatos para el componente de media móvil son:

$$q \in \{0, 1, 2, 5, 6, 7, 11, 12\}.$$

La Figura 4b muestra la función PACF de la serie diferenciada. Los rezagos significativos identificados son los lag 1, 2, 4, 5, 6, 10, 11 y 12. El lag 1 muestra una autocorrelación parcial negativa pronunciada (aproximadamente -0.6), mientras que el lag 12 presenta autocorrelación parcial positiva significativa (aproximadamente 0.4), sugiriendo dependencia con observaciones del mismo mes del año anterior. Los valores candidatos para el orden del componente autorregresivo son:

$$p \in \{0, 1, 2, 4, 5, 6, 10, 11, 12\}.$$

La persistencia de autocorrelaciones significativas hasta el lag 12, combinada con el patrón observado en el ACF y PACF, confirma la presencia de una estructura de dependencia en el tiempo que incluye dinámicas de corto plazo y dependencia en lag 12. La especificación final del modelo ARIMA$(p, 1, q)$ se determinará evaluando diferentes

Figura 3: Primera diferenciación de la serie de tiempo del IPI del sector de extracción de petróleo crudo y gas natural (2014-2025).

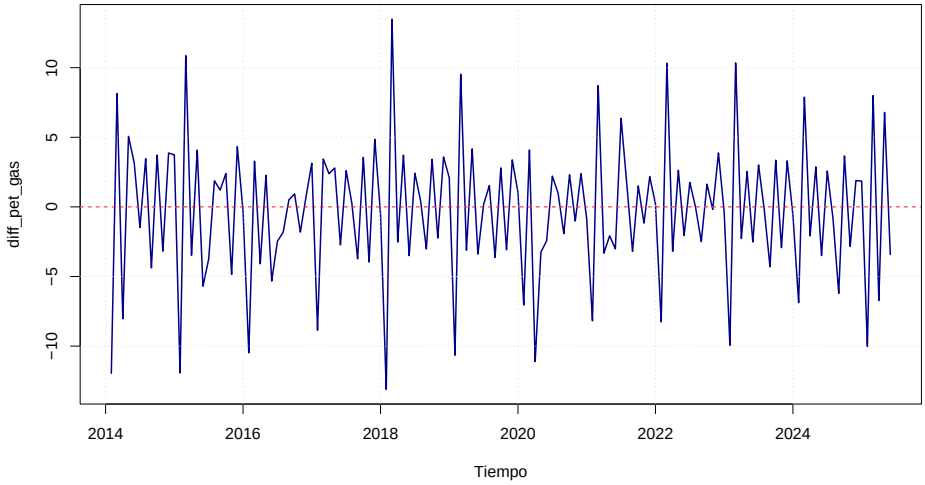
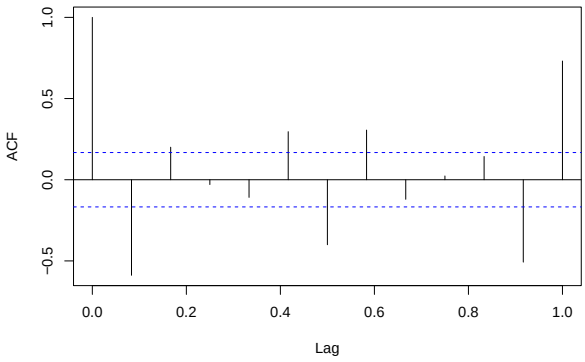
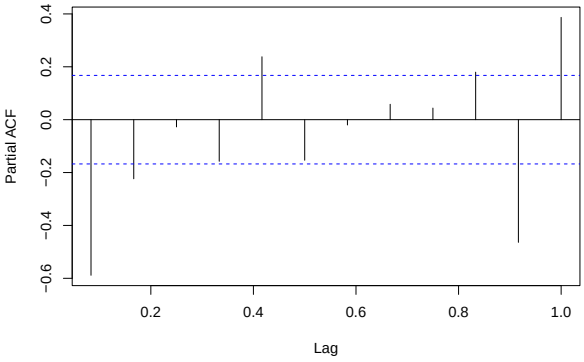


Figura 4: Funciones de autocorrelación (a) ACF y (b) PACF, de la serie diferenciada del sector de extracción de petróleo crudo y gas natural (2014-2025).



(a) Función de Autocorrelación (ACF)



(b) Función de Autocorrelación Parcial (PACF)

combinaciones de estos órdenes candidatos mediante el Criterio de Información de Akaike y análisis de residuales.

2.5 Especificación y selección de modelos

De acuerdo con los valores de p, el orden $d = 1$ y q establecidos en la Sección 2.4, se formaron 71 modelos candidatos correspondientes a las combinaciones posibles de p y q, eliminando la terna $(0, 1, 0)$ por carecer de estructura de autocorrelación.

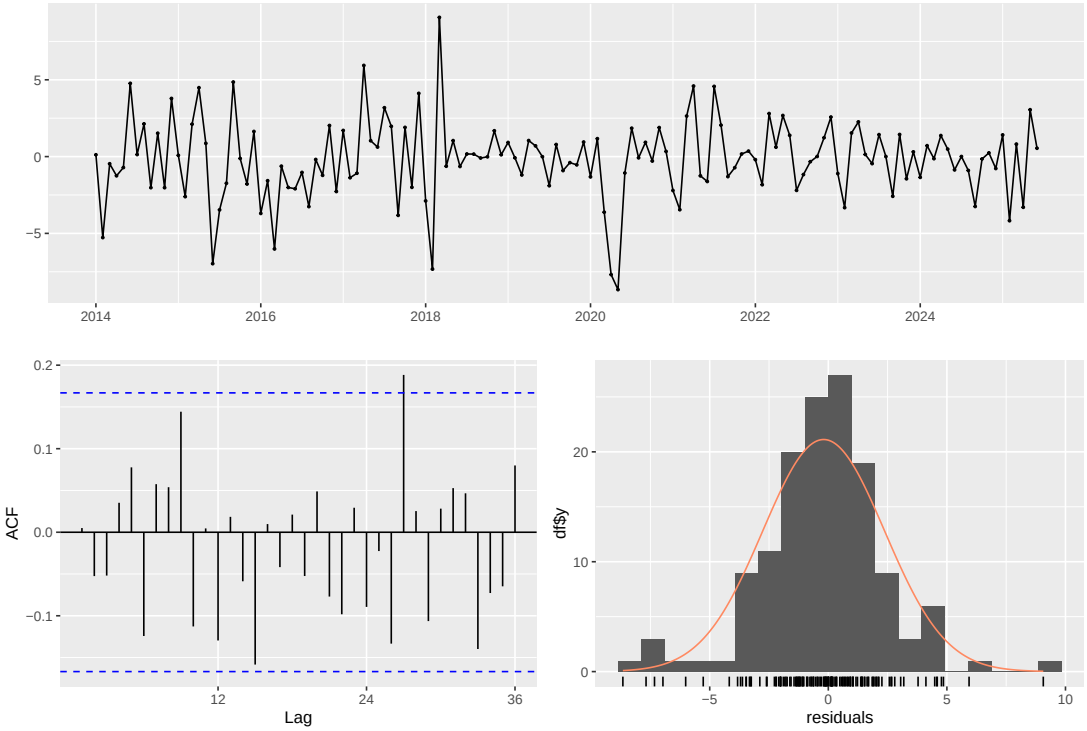
Para cada modelo candidato, se estimó la especificación completa evaluando la significancia estadística de cada coeficiente mediante pruebas z ($\alpha = 0.05$). Cuando un modelo presentaba coeficientes no significativos, se reestimó fijando dichos coeficientes en cero. Este proceso de simplificación se aplicó sistemáticamente a los 71 modelos candidatos.

La selección del mejor modelo se realizó mediante el Criterio de Información de Akaike (AIC), prefiriendo el modelo con menor valor. Del universo inicial de 71 modelos, se obtuvieron 36 finales tras aplicar el proceso de depuración, debido a problemas de convergencia y modelos duplicados. La Tabla 3 presenta los valores de AIC para los 10 mejores modelos. Estos resultados muestran que el modelo64 ARIMA$(12, 1, 0)$ presenta el menor valor de AIC (666.08). Antes de la selección final, es necesario verificar que el modelo cumple con los supuestos mediante diagnóstico de residuales.

Tabla 3: Valores AIC de los 10 mejores modelos ARIMA estimados.

Rank	Modelo	AIC
1	modelo64	666.08
2	modelo57	685.05
3	modelo68	688.14
4	modelo63	696.62
5	modelo47	704.03
6	modelo36	712.29
7	modelo23	716.38
8	modelo31	716.38
9	modelo56	720.47
10	modelo39	723.98

Figura 5: Diagnóstico completo de residuales del modelo ARIMA$(12, 1, 0)$ seleccionado. Panel superior: residuales estandarizados a lo largo del tiempo. Panel inferior izquierdo: función de autocorrelación (ACF) de los residuales hasta el lag 36. Panel inferior derecho: histograma de los residuales con curva normal superpuesta.



2.6 Diagnóstico de residuales

La Figura 5 presenta el diagnóstico completo de los residuales del modelo ARIMA$(12, 1, 0)$ seleccionado. El panel superior muestra los residuales estandarizados a lo largo del tiempo, oscilando aleatoriamente alrededor de cero sin patrones sistemáticos, tendencias o cambios evidentes en la varianza, consistente con homocedasticidad. El panel inferior izquierdo presenta la función ACF de los residuales hasta el lag 36, donde la mayoría de los coeficientes se encuentran dentro de las bandas de confianza al 95 %, con solo un rezago mostrando autocorrelación leve, lo cual será evaluado formalmente mediante la prueba de Ljung-Box. El panel inferior derecho muestra el histograma de los residuales con curva normal superpuesta, revelando una distribución aproximadamente simétrica y centrada en cero con ligeras desviaciones en las colas.

2.6.1 Prueba de autocorrelación: Ljung-Box

La prueba de Ljung-Box evalúa formalmente la hipótesis nula de que no existe autocorrelación en los residuales. La Tabla 4 presenta los resultados. Con un estadístico $Q^* = 22.45$ y 12 grados de libertad, se obtiene un valor p de 0.03277. Utilizando un nivel de significancia de $\alpha = 0.01$ para diagnóstico de residuales (Box & Jenkins, 1970), no se rechaza la hipótesis nula ($p = 0.03277 > 0.01$), confirmando que los residuales no presentan autocorrelación estadísticamente significativa y que el modelo ha capturado adecuadamente la estructura de dependencia temporal.

Tabla 4: Resultados de la Prueba de Ljung-Box.

Estadístico/Parámetro	Valor
Estadístico Q^*	22.45
Grados de libertad	12
Rezagos evaluados	24
Parámetros del modelo	12
Valor p	0.03277

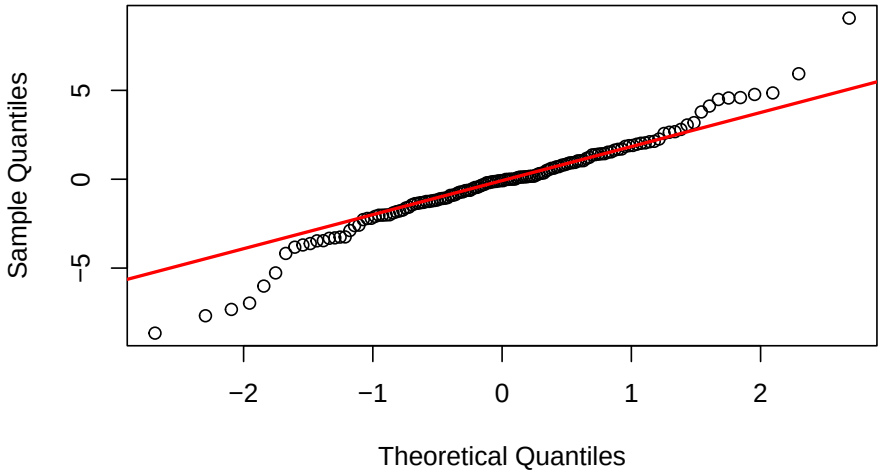
2.6.2 Verificación de homocedasticidad: Test de Breusch-Pagan

El test de Breusch-Pagan evalúa la hipótesis nula de varianza constante de los residuales a lo largo del tiempo. La Tabla 5 presenta los resultados. Con un estadístico $BP = 0.9155$ y valor $p = 0.3387$, no se rechaza la hipótesis nula de homocedasticidad al nivel de 0.05 ($p = 0.3387 > 0.05$), confirmando que la varianza de los residuales es estable, validando el supuesto de homocedasticidad necesario para la construcción de intervalos de confianza confiables.

Tabla 5: Resultados del Test de Breusch-Pagan.

Estadístico/Parámetro	Valor
Estadístico BP	0.9155
Grados de libertad	1
Valor p	0.3387

Figura 6: Q-Q Plot de los residuales del modelo ARIMA(12,1,0).



2.6.3 Prueba de normalidad: Shapiro-Wilk y Q-Q Plot

La Tabla 6 presenta los resultados de la prueba de Shapiro-Wilk. Con un estadístico $W = 0.9641$ y valor p de 0.001085, se rechaza la hipótesis nula de normalidad al nivel de 0.05.

Tabla 6: Resultados de la Prueba de Shapiro-Wilk.

Estadístico/Parámetro	Valor
Estadístico W	0.9641
Valor p	0.001085

La Figura 6 presenta el gráfico Q-Q que compara los cuantiles empíricos de los residuales con los cuantiles teóricos de una distribución normal estándar. Los residuales se ajustan razonablemente bien a la distribución normal en su parte central, con la mayoría de los puntos alineados sobre la línea de referencia. Las desviaciones sistemáticas en ambas colas indican presencia de colas más pesadas que las de una distribución normal (leptocurtosis), lo cual explica el rechazo de normalidad en la prueba de Shapiro-Wilk, aunque esta desviación se concentra principalmente en los valores extremos.

El análisis exhaustivo permite concluir que el modelo ARIMA$(12, 1, 0)$ cumple satisfactoriamente con los supuestos fundamentales: (i) ausencia de autocorrelación residual (Ljung-Box: $p = 0.033 > 0.01$), (ii) varianza estable en el tiempo (Breusch-Pagan: $p = 0.339 > 0.05$), y (iii) desviación leve de normalidad concentrada en las colas que no compromete la validez del modelo dado el tamaño muestral disponible. Por lo tanto, el modelo es adecuado para realizar pronósticos válidos y confiables del IPI del sector petrolero colombiano.

3 Resultados y discusión

Esta sección presenta los resultados de la modelación y pronóstico del IPI del sector petrolero colombiano. Se describe el modelo ARIMA$(12, 1, 0)$ seleccionado con sus coeficientes estimados y especificación completa, seguido de los pronósticos a 12 meses con intervalos de confianza al 95 %. Adicionalmente, se presentan los resultados del modelo bayesiano BSTS como aproximación complementaria.

3.1 Modelo ARIMA seleccionado

El modelo ARIMA$(12, 1, 0)$ seleccionado mediante el proceso descrito en la Sección 2.5 presenta tres coeficientes autorregresivos significativos (rezagos 1, 11 y 12) aplicados a la serie diferenciada. La Tabla 7 presenta los coeficientes estimados.

3.1.1 Coeficientes estimados

La Tabla 7 presenta los coeficientes estimados del modelo junto con sus errores estándar. Los coeficientes AR(2) a AR(10) fueron fijados en cero durante el proceso de simplificación, ya que resultaron estadísticamente no significativos.

Tabla 7: Coeficientes estimados del modelo ARIMA$(12, 1, 0)$.

Parámetro	Estimación	Error Estándar
AR(1)	-0.2108	0.0510
AR(11)	-0.1714	0.0528
AR(12)	0.5971	0.0681

El coeficiente AR$(1) = -0.2108$ indica correlación negativa con el periodo inmediatamente anterior. El coeficiente AR$(11) = -0.1714$ captura dependencias de mediano plazo. El coeficiente AR$(12) = 0.5971$ es el de mayor magnitud absoluta y refleja un fuerte componente estacional anual, donde los valores actuales están positivamente correlacionados con los del mismo mes del año anterior. El modelo presenta una varianza residual $\sigma^2 = 6.628$ (desviación estándar $\sigma \approx 2.57$ puntos del IPI del sector analizado), representando aproximadamente el 2.6 % del nivel promedio del índice. El Error Porcentual Absoluto Medio (MAPE) sobre el conjunto de entrenamiento

es de 1.89 %, es decir, las predicciones del modelo difieren en promedio menos de 2 puntos porcentuales de los valores observados, indicando un ajuste preciso a los datos históricos.

3.1.2 Ecuación del modelo y pronósticos

El modelo ARIMA$(12, 1, 0)$ en su forma diferenciada se expresa como:

$$\nabla Y_t = -0.2108\nabla Y_{t-1} - 0.1714\nabla Y_{t-11} + 0.5971\nabla Y_{t-12} + \varepsilon_t.$$

donde $\nabla Y_t = Y_t - Y_{t-1}$ representa la primera diferencia, Y_t es el índice del sector en el mes t del sector analizado, y $\varepsilon_t \sim N(0, \sigma^2)$ con $\sigma^2 = 6.628$. Transformando a niveles originales mediante sustitución y combinando términos semejantes tenemos

$$Y_t = 0.7892Y_{t-1} + 0.2108Y_{t-2} - 0.1714Y_{t-11} + 0.7685Y_{t-12} - 0.5971Y_{t-13} + \varepsilon_t.$$

El coeficiente de Y_{t-1} (0.7892) indica que aproximadamente el 79 % del nivel de producción del mes anterior persiste en el mes actual. El coeficiente de Y_{t-12} (0.7685) muestra que aproximadamente el 77 % del nivel de producción del mismo mes del año anterior se transmite al periodo actual.

La Tabla 8 y la Figura 7 presentan los pronósticos del sector para los próximos 6 meses con intervalos de confianza al 95 %. La tabla muestra los valores puntuales mes a mes junto con sus errores estándar, mientras que la figura ilustra la trayectoria pronosticada con su banda de incertidumbre.

Los pronósticos revelan tres características importantes. Primero, el error estándar aumenta progresivamente desde 2.57 puntos en el primer mes hasta 5.38 puntos en el mes 6, reflejando el incremento natural de incertidumbre con el horizonte. Segundo, los pronósticos puntuales oscilan entre 83.1 (mes 3) y 87.4 (meses 1 y 2), con un promedio de 85.6 puntos, aproximadamente 14 % por debajo del año base 2018 (100 puntos) y 12 % por debajo del promedio histórico del periodo analizado (97.68 puntos). Tercero, la amplitud de los intervalos de confianza se expande desde ± 5.0 puntos en el primer mes hasta ± 10.6 puntos en el mes 6, implicando mayor precisión en los pronósticos de corto plazo.

3.2 Modelo BSTS

El modelo BSTS constituye un enfoque alternativo de pronóstico que permite incorporar incertidumbre estructural mediante métodos bayesianos. A diferencia del enfoque clásico de Box-Jenkins utilizado anteriormente, el BSTS emplea Cadenas de Markov Monte Carlo para la estimación de parámetros, generando distribuciones posteriores que capturan la incertidumbre del pronóstico de manera explícita.

Para el ajuste del modelo se utilizó la serie truncada de 18 observaciones mensuales correspondientes al periodo enero 2024 - junio 2025. La especificación del modelo incluyó únicamente un componente de tendencia semilocal lineal, prescindiendo de componentes estacionales dado el tamaño limitado de la muestra.

El modelo generó pronósticos para un horizonte de 6 meses, presentando un MAPE de 1.06 %, es decir, las predicciones del modelo difieren en promedio aproximadamente 1 punto porcentual de los valores observados. Este nivel de precisión resulta notable y sugiere que el modelo captura adecuadamente la dinámica de corto plazo de la serie. Los valores pronosticados se presentan en la Tabla 9, evidenciando un patrón oscilante con tendencia descendente moderada, consistente con el comportamiento observado en el periodo de análisis.

La Figura 8 ilustra la serie en el periodo considerado junto con el pronóstico BSTS y su intervalo de confianza al 95 %. Se observa que los intervalos de confianza se amplían gradualmente conforme aumenta el horizonte de pronóstico, reflejando el incremento natural de la incertidumbre. Los pronósticos sugieren una continuación de la tendencia observada en el periodo reciente, con valores que oscilan entre 83.88 y 86.47 puntos del índice.

Los resultados del modelo BSTS revelan tres características importantes. Primero, el error estándar aumenta progresivamente desde 3.81 puntos en el primer mes hasta 7.13 puntos en el mes 6, reflejando el incremento natural de incertidumbre con el horizonte. Segundo, los pronósticos puntuales oscilan entre 85.8 (mes 5) y 88.1 (mes 2), con un promedio de 86.9 puntos, aproximadamente 13 % por debajo del año base 2018 (100 puntos) y 11 % por debajo del promedio histórico del periodo analizado (97.68 puntos). Tercero, la amplitud de los intervalos de confianza se expande desde $\pm$7.3 puntos en el primer mes hasta $\pm$14.1 puntos en el mes 6, implicando mayor precisión en los pronósticos de corto plazo.

4 Conclusiones

Este trabajo desarrolló modelos de pronóstico para el Índice de Producción Industrial del sector de extracción de petróleo crudo y gas natural en Colombia mediante dos enfoques complementarios: ARIMA y BSTS. El análisis sobre datos mensuales del DANE (enero 2014 - junio 2025) identificó el modelo ARIMA$(12, 1, 0)$ como especificación óptima (AIC $= 666.08$, MAPE $= 1.89\,\%$), superando 35 modelos alternativos y satisfaciendo los supuestos de ausencia de autocorrelación residual, homocedasticidad y distribución aproximadamente normal.

Los pronósticos a 6 meses del modelo ARIMA indican que el IPI se mantendrá en niveles entre 83.1 y 87.4 puntos (promedio: 85.6), aproximadamente 14 % por debajo del año base 2018 y 12 % por debajo del promedio histórico. El modelo BSTS, aplicado sobre el periodo reciente (enero 2024 - junio 2025), generó pronósticos con MAPE de 1.06 %, proyectando niveles entre 85.8 y 88.1 puntos (promedio: 86.9), consistentes con los resultados del ARIMA. Ambos modelos confirman la persistencia de una contracción estructural sin señales de recuperación en el corto plazo.

La comparación metodológica revela ventajas complementarias: el ARIMA captura patrones de largo plazo sobre 138 observaciones con estructura autorregresiva,

Tabla 8: Pronósticos del IPI del sector de extracción de petróleo y gas mediante modelo ARIMA a 6 meses.

Horizonte	Pronóstico	Error estándar	Límite inferior	Límite superior
1 mes	87.4	2.57	82.3	92.4
2 meses	87.4	3.28	81.0	93.8
3 meses	83.1	3.92	75.4	90.7
4 meses	86.7	4.46	77.9	95.4
5 meses	83.9	4.94	74.2	93.6
6 meses	85.3	5.38	74.7	95.8

Figura 7: Pronóstico del IPI del sector de extracción de petróleo y gas a 6 meses con intervalos de confianza al 95 %.

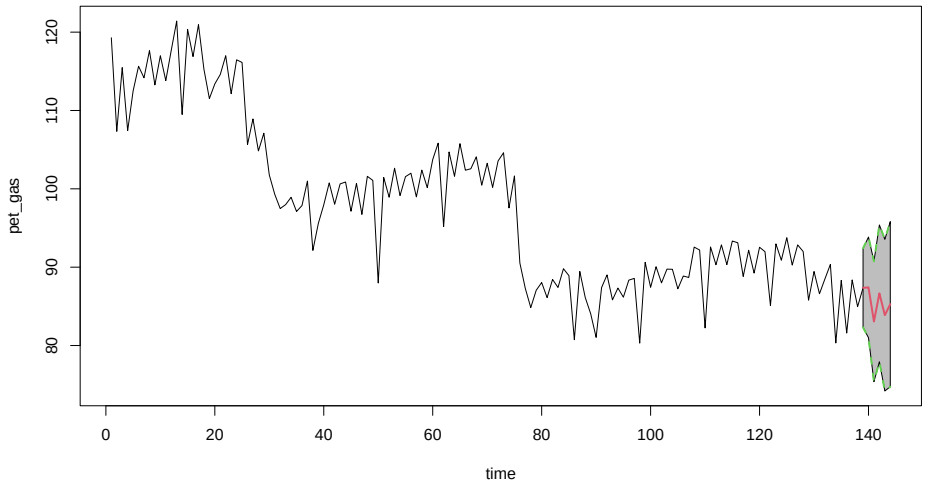
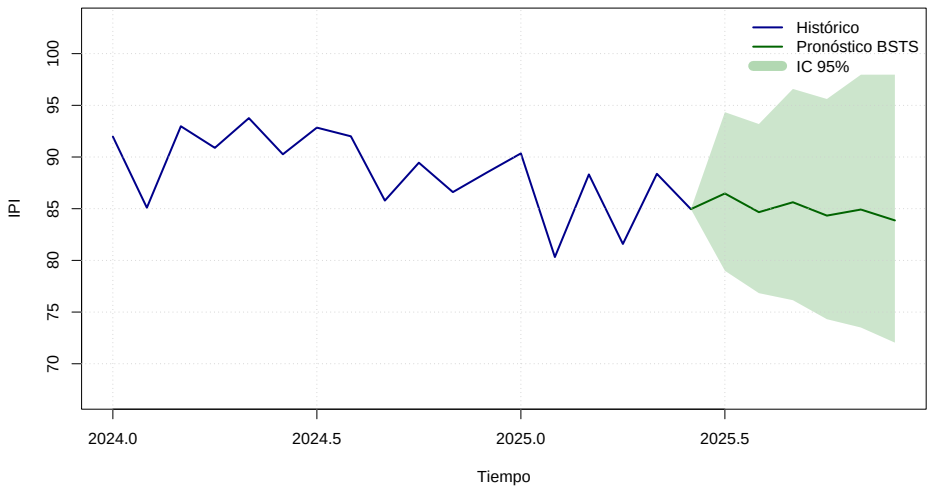


Figura 8: Serie del IPI del año 2024 a junio de 2025 y pronóstico BSTS a 6 meses con intervalo de confianza al 95 %.



mientras que el BSTS incorpora incertidumbre estructural mediante inferencia bayesiana, mostrando mayor precisión (MAPE inferior) sobre la dinámica reciente del sector. Estos hallazgos tienen implicaciones importantes para la formulación de política económica, planificación presupuestal y toma de decisiones de inversión en el sector extractivo colombiano.

Extensiones futuras podrían incorporar modelos multivariados considerando otras variables explicativas adicionales, que capturen rupturas estructurales asociadas a choques de precios, cambios regulatorios o innovaciones tecnológicas.

Referencias

Akpinar, M., & Yumusak, N. (2021). Forecasting Natural Gas Production and Consumption in United States—Evidence from SARIMA and SARIMAX Models. *Energies*, *14*(19), 6021. https://doi.org/10.3390/en14196021

Tabla 9: Pronósticos del IPI del sector mediante modelo BSTS a 6 meses.

Horizonte	Pronóstico	Error estándar	Límite inferior	Límite superior
1 mes	86.6	3.81	79.8	94.4
2 meses	88.1	4.08	80.4	96.6
3 meses	86.2	5.24	76.2	96.5
4 meses	87.4	5.49	76.7	99.0
5 meses	85.8	6.62	73.2	100.9
6 meses	87.1	7.13	73.8	102.0

Asociación Colombiana del Petróleo y Gas. (2023, febrero). *Tendencias de Inversión en Exploración y Producción (E&P) de Petróleo y Gas en Colombia 2022 y Perspectivas 2023* (Informe Económico). Asociación Colombiana del Petróleo y Gas.

Box, G. E. P., & Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day.

Chacón Cruz, T. P., & Riaño Amaya, C. A. (2020). *Análisis del Sector Petrolero en Colombia, Carga Tributaria y Comparación con Perú, México y Ecuador* [Trabajo de Grado]. Universidad Jorge Tadeo Lozano [Especialización en Administración y Auditoría Tributaria].

DANE. (2020a, septiembre). *Ficha Metodológica IPI - Índice de Producción Industrial* (inf. téc.) (Código: DSO-IPI-FME-001, Versión 4). Departamento Administrativo Nacional de Estadística. Colombia.

DANE. (2020b, septiembre). *Metodología General Índice de Producción Industrial - IPI* (inf. téc.) (Código: DSO-IPI-MET-001, Versión 4). Departamento Administrativo Nacional de Estadística. Colombia.

DANE. (2022). *Clasificación Industrial Internacional Uniforme de todas las actividades económicas. Revisión 4 adaptada para Colombia* (inf. téc.) (Actualización 2022). Departamento Administrativo Nacional de Estadística. Bogotá, Colombia. https://www.dane.gov.co/files/sen/nomenclatura/ciiu/CIIU_Rev_4_AC2022.pdf

DANE. (2025, agosto). *Boletín Técnico: Índice de Producción Industrial (IPI) - Junio 2025* (inf. téc.). Departamento Administrativo Nacional de Estadística. Bogotá D.C., Colombia.

Ejaz, M., & Iqbal, J. (2021). Estimation and Forecasting of Industrial Production Index. *Lahore Journal of Economics*, *26*(1), 1-30.

Franses, P. H., & Ooms, M. (1997). Forecasting industrial production using structural time series models. *European Journal of Operational Research*, *110*(3), 499-516. https://doi.org/10.1016/S0377-2217(98)00024-3

Golinelli, R., & Parigi, G. (2009). Forecasting monthly industrial production in real-time: from single equations to factor-based models. *Empirical Economics*, *37*, 303-336. https://doi.org/10.1007/s00181-008-0234-z

Hamza, H. A., & Mohammad, M. J. (2024). A Comparison Of The Bayesian Structural Time Series Technique With The Autoregressive Integrated Moving Average Model For Forecasting. *Migration Letters*, *21*(S4), 528-538.

Lu, Q., Li, Y., Chai, J., & Wang, S. (2020). Crude oil price analysis and forecasting: A perspective of “new triangle”. *Energy Economics*, *87*, 104721. https://doi.org/10.1016/j.eneco.2020.104721

Marzo Carpio, M. (2015, febrero). *El desplome 2014-2015 de los precios del crudo: Causas y previsiones a corto plazo* (Informe estratégico N.º 001-2015) (Documento FUNSEAM 001-2015). Fundación para la Sostenibilidad Energética y Ambiental (FUNSEAM). Barcelona, España. https://funseam.com/wp-content/uploads/2015/02/k2_attachments_Informe_Funseam_Desplome_2014-2015_de_los_precios_del_crudo-Febrero2015.pdf

Nguyen, T. H., Nguyen, N. M. T., Le, T. P. T., Nguyen, T. H. P., & Hoang, T. H. (2025). Application of ARIMA and LSTM Models for Crude Oil Price Forecasting. *International Journal of Scientific Engineering and Science*, *9*(4), 31-34.

Scott, S. L., & Varian, H. R. (2014). Predicting the present with Bayesian structural time series. *International Journal of Mathematical Modelling and Numerical Optimisation*, *5*(1/2), 4-23. https://doi.org/10.1504/IJMMNO.2014.059942

Suleiman, M., Muhammad, I., Adamu, A. Z., Zakari, Y., Iliyasu, R., Muhammad, A., Adamu, I., & Abdu, M. (2025). Modelling Nigeria Crude Oil Prices using ARIMA Time Series Models. *Journal of Science and Technology Research*, *7*(1).

Torres Hernández, A. S., Hamon Parra, E. G., & Morales Palacios, J. A. (2020). *Importancia de la producción y exportación de petróleo y gas en la industria colombiana en el marco de la Alianza del Pacífico* [Trabajo de grado]. Universitaria Agustiniana.

Zhao, C.-l., & Wang, B. (2014). Forecasting crude oil price with an autoregressive integrated moving average (ARIMA) model. En *Fuzzy Information & Engineering and Operations Research & Management* (pp. 275-286). Springer. https://doi.org/10.1007/978-3-642-38667-1_28