
Modelamiento del consumo de combustible en Colombia por medio de modelos lineales dinámicos

Modeling of fuel consumption in Colombia through dynamic models

Paula Hernández^a
mariahernandez@usantotomas.edu.co

Daniel Gómez^b
daniel.gomezr@usantotomas.edu.co

Alejandra García^c
gabrielagarcia@usantotomas.edu.co

Dagoberto Bermúdez^d
dagobertobermudez@usantotomas.edu.co

Resumen

El objetivo de este artículo es el modelamiento del consumo mensual de combustible regular en Colombia, tomando como covariables la cantidad de vehículos vendidos y matriculados mensualmente en el país, el promedio del valor de la tasa representativa del mercado (TRM) y el precio promedio mensual de la acción de Ecopetrol. Utilizando datos públicos encontrados en el DNP (Departamento Nacional de Planeación) entre el periodo de enero de 1996 y abril de 2018, el modelamiento de la variable objetivo en este estudio se realizó mediante el enfoque de los modelos lineales dinámicos. Sin embargo, en un principio se analizó el comportamiento de dicha variable desde un enfoque frecuentista utilizando modelos ARIMA, ARIMAX y SARIMAX.

Palabras clave: Consumo de combustible, Modelos lineales dinámicos, Vehículos matriculados, Tasa representativa del mercado (TRM), Modelos ARIMA.

Abstract

The objective of this article is the modeling of the monthly consumption of regular fuel in Colombia, taking as covariates the number of vehicles sold and registered monthly in the country, the average value of the representative market rate (TRM), and the monthly average price of Ecopetrol's stock. Using public data found in the DNP (National Planning Department) between January 1996 and April 2018, the modeling of the target variable in this study was carried out using the approach of dynamic linear models. However, initially, the behavior of this variable was analyzed from a frequentist approach using ARIMA, ARIMAX, and SARIMAX models.

Keywords: Fuel consumption, Dynamic linear models, Registered vehicles, Market representative rate (TRM), ARIMA models.

^aFacultad de Estadística Universidad Santo Tomás

^bFacultad de Estadística Universidad Santo Tomás

^cFacultad de Estadística Universidad Santo Tomás

^dFacultad de Estadística Universidad Santo Tomás

1. Introducción

Según la Vicepresidencia de Asuntos Económicos y Regulatorios de la ACP (2021), actualmente en el país, los combustibles líquidos representan el 96 % del consumo final de energía del sector transporte, y se espera que para el año 2050 continúen siendo los principales energéticos que soportan la movilidad del país.

Es importante resaltar que el consumo de combustible es un factor indispensable en la mayoría de los sectores económicos del país, como el sector industrial, en la producción de alimentos o en el transporte público e individual de los colombianos. El uso de combustible está directamente relacionado con los costos de operación en los que incurren las diferentes empresas en el país, pero también puede ser de vital importancia en el análisis del costo de vida de las personas.

Es por esta razón que es indispensable tener un panorama claro del comportamiento de este fenómeno a lo largo del tiempo, ya que puede ser de gran utilidad a la hora de tomar políticas públicas por parte del estado. Por ejemplo, según John Posada (2014), el análisis de costos es necesario cuando se considera un proyecto de carreteras, siendo los costos de operación vehicular (VOC, por sus siglas en inglés) un componente clave en este análisis. El consumo de combustible es de significativa importancia en estos VOC y es influenciado por diferentes aspectos como la geometría de la carretera, la superficie de rodadura, la velocidad y el estilo de conducción, entre otros.

Debido a la gran relevancia que tiene este fenómeno, se realizó el modelamiento del consumo de combustible desde un punto de vista estadístico, enfocando el análisis desde la metodología de los modelos lineales dinámicos, con el fin de llegar a la conclusión de la elección del modelo que mejor estime y explique el comportamiento de esta variable en el país.

2. Marco teórico

García et al. (2016) explora, mediante el modelo casi ideal de demanda, las elasticidades del precio de la demanda, el precio cruzado y el gasto en combustibles como la gasolina, el diésel y el GNV (Gas Natural Vehicular). Utiliza métodos como aproximaciones lineales y el índice SUR, que incorpora la dependencia espacial en los errores de los sistemas de regresión lineal. Con 120 datos de demanda y precios de combustibles, los hallazgos principales sugieren que la gasolina y el diésel son bienes inelásticos, mientras que el GNV muestra una elasticidad de la demanda más elástica.

Por otro lado, Fuentes et al. (2020) destaca el uso de modelos predictivos para estimar el consumo de combustible en vehículos de transporte por carretera, considerando variables como las características del vehículo, la carga transportada y la ruta. Originalmente basados en datos empíricos, estos modelos se han ido perfeccionando con datos experimentales, enfocándose en la relación entre el consumo de combustible y las fuerzas que se oponen al movimiento. La metodología se centra en la interacción de factores determinantes de costos, como estudios, construcción, operación y mantenimiento. Modelos empíricos, mecanicistas, HDM-III, COPER y VOC son comúnmente utilizados para predecir el consumo de combustible en vehículos.

Modelos lineales dinámicos

Una de las ramas más importantes en la estadística es la que concierne a los modelos; estos son bastante útiles en la descripción y predicción de los fenómenos en las áreas de estudio ya mencionadas, desde que se tienen nociones del modelamiento de variables por medio de la estadística los esfuerzos de los teóricos en buscar nuevas alternativas y proponer nuevos métodos han sido incesantes, yendo desde los modelos tradicionales de regresión lineal propuestos por primera vez por Legendre en 1805 y posteriormente profundizados por Gauss hasta la implementación de los modelos dinámicos los cuales fueron desarrollados a principios de la década de 1960 en ingeniería para monitorear sistemas dinámicos.

Se puede decir que los modelos lineales dinámicos son una extensión de los modelos de regresión tradicionales implementando un enfoque bayesiano.

Recordando la estructura de un modelo de regresión lineal tradicional se tiene que.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon_i, \text{ Donde } \varepsilon_i \sim N(0, \sigma_i^2)$$

Desde el enfoque dinámico se parte desde la misma estructura, pero se supone que los parámetros siguen una distribución de probabilidad, teniendo en cuenta esto se tiene que.

$$Y_t = \beta_{0t} + \beta_{1t} X_{1t} + \cdots + \beta_{kt} X_{kt} + \varepsilon_t, \text{ Donde } \varepsilon_t \sim N(0, \sigma_t^2)$$

Donde $\beta_{it} = \beta_{i,t-1} + W_{it}$, donde $W_{it} \sim N(c_{it}, v_{it})$, en donde c_{it} y v_{it} son los hiperparámetros de la distribución de probabilidad de W_{it} , para σ_t^2 se utiliza una distribución inversa gamma, o en el caso de trabajar con ϕ_t (la precisión de la distribución normal) se utiliza una distribución gamma de parámetros (α, β) .

Debido a que estos modelos sugieren normalidad en la variable respuesta es bastante útil para modelar variables continuas las cuales su dominio se encuentran de $(-\infty, \infty)$.

Sin embargo, en muchas ocasiones las variables respuesta del estudio puede que no sean cuantitativas o además se muevan en otro dominio diferente a $(-\infty, \infty)$ y es en este punto donde resulta de bastante utilidad utilizar los modelos lineales generalizados.

A diferencia de los modelos lineales, los modelos lineales generalizados son modelos con variables respuestas categóricas, que siguen una distribución diferente a la distribución Normal, por ende se obtiene que la relación entre la respuesta y las variables no siempre son lineales.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + \varepsilon_i$$

Siendo Y_i la variable respuesta la cual se modela con una función lineal de variables predictoras y un término de error.

Para el objetivo de este trabajo el modelo generalizado Poisson es de bastante utilidad ya que la variable de estudio es de carácter discreta.

El modelo de regresión Poisson o también conocido como el modelo log-lineal, debido que al transformar una relación no lineal en una relación lineal se utiliza el logaritmo de la regresión Poisson denominada función de enlace, es utilizado para el modelamiento de datos de conteo.

$$\ln(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$$

$$\text{Donde } Y \sim \text{Poisson}(\lambda = e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p})$$

Donde:

- Y : variable respuesta
- β_p : coeficientes numéricos
- X : variables explicativas

Donde, β_p se estiman utilizando métodos como la estimación de máxima verosimilitud.

Una de las características más importantes del modelo es la equidispersión, es decir, su media y su varianza son iguales.

Extendiendo esta analogía a la metodología dinámica se tiene lo siguiente.

$$\begin{aligned} Y_t &\sim \text{Poisson}(\lambda) \\ \lambda &= e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p} \\ \beta_{p,t} &= \beta_{p,t-1} + W_{p,t} \\ B_{p,t} &\sim N(B_{p,t-1}, W_{p,t}) \\ W_{p,t} &\sim \text{Gamma}(\alpha_0, \beta_0) \end{aligned}$$

3. Estimación de los modelos

El objetivo del trabajo es modelar la cantidad de barriles de gasolina de 42 galones consumidos en Colombia. Para ello, se utilizaron las siguientes variables auxiliares:

Variables	Descripción
Y	Cantidad de barriles de gasolina de 42 galones en Colombia
X_1	Cantidad de vehículos vendidos y matriculados en Colombia
X_2	Tasa representativa del mercado TRM
X_3	Precio de la acción de Ecopetrol

Tabla 1: Variables de estudio

A continuación, se presentan gráficas de la serie histórica de las variables de estudio.

Se realizó el análisis de la serie de tiempo del consumo de combustible en Colombia, observando una disminución en el consumo desde inicios del año 2000 hasta el año 2010, seguido de un aumento posterior. Además, se analizó la serie de tiempo de la venta y matrícula de vehículos en Colombia, la cual mostró un crecimiento desde el año 2000 con mayor volatilidad después del año 2010. También

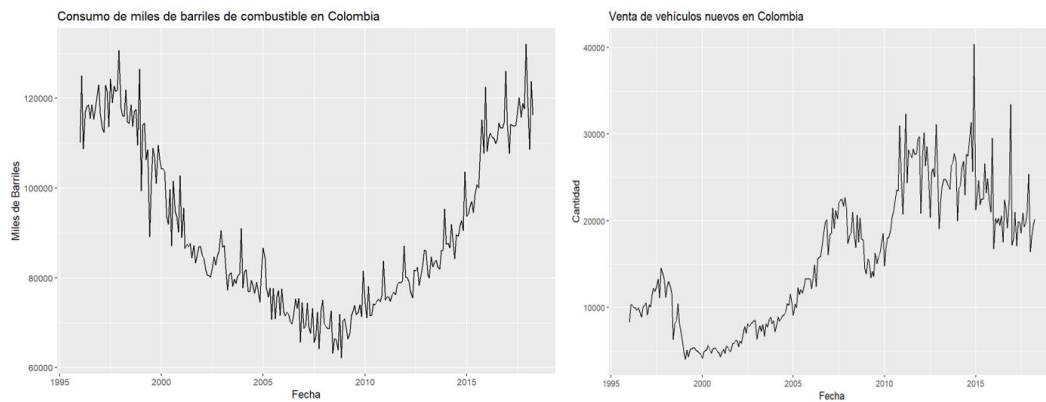


Figura 1: Consumo de combustible y venta y matrícula de vehículos. Fuente: ANDEMOS. Elaboración: Propia.

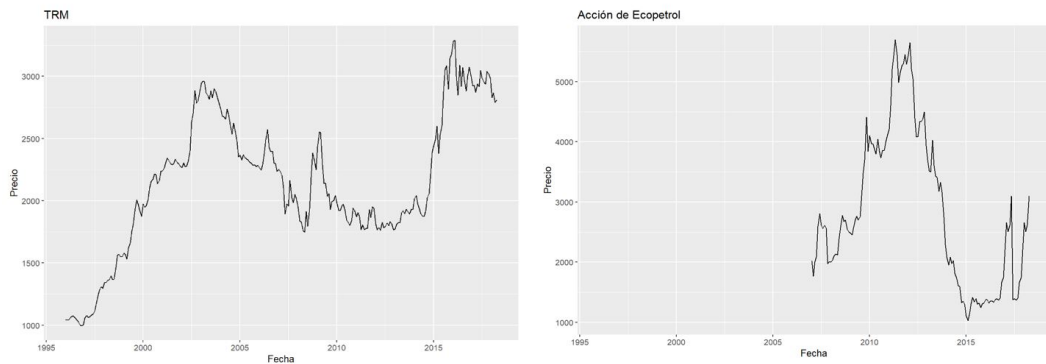


Figura 2: Serie histórica de la TRM y la acción de Ecopetrol. Fuente: Investing.com. Elaboración: Propia.

se estudiaron las series de tiempo del promedio de precios de la tasa representativa del mercado (TRM) y de la acción de Ecopetrol, observando una tendencia al alza en la TRM y una tendencia a la baja en la acción de Ecopetrol en los últimos años.

Se llevaron a cabo pruebas de estacionariedad Dickey-Fuller para la serie de consumo de combustible y su serie diferenciada. Los resultados indicaron que la serie original no es estacionaria (p -valor > 0.05), mientras que la serie diferenciada sí lo es (p -valor = 0.01). Esto es relevante porque, a diferencia de los enfoques frecuentistas que requieren estacionariedad para ajustar modelos, los modelos dinámicos no tienen esta restricción, lo que los hace particularmente útiles para el análisis de series de tiempo que no cumplen con el supuesto de estacionariedad.

3.1. Estimación del mejor modelo SARIMAX

Se realizaron pruebas con varios modelos de la familia ARIMA y se calculó el AIC para determinar qué modelo se ajusta mejor a la serie del consumo de combustible en el país, obteniéndose los siguientes resultados. Se probaron diferentes órdenes de modelos ARIMA (p,q) que se presentan en la siguiente tabla.

p	q	AIC
0	2	5072.51
1	1	5072.80
3	3	5073.25
2	3	5073.73
2	0	5073.93
2	1	5073.99
0	3	5074.37
1	2	5074.42
3	0	5074.89
2	2	5075.30
3	1	5075.41
1	3	5076.37
3	2	5077.05
0	1	5078.02
1	0	5096.52

Tabla 2: AIC para los modelos ARIMA estimados

Se observó que los modelos con menor AIC son el ARIMA(0,1,2) y el ARIMA(1,1,1), por lo que se tomaron como referencia para la estimación del modelo definitivo. Posteriormente, se introdujo la variable exógena de venta de vehículos y un componente estacional a estos modelos para observar si mejoraban en términos del AIC. Los resultados se presentan en la siguiente tabla.

Modelo	AIC
ARIMA(0,1,2)	5072.514
ARIMA(1,1,1)	5072.803
ARIMAX(0,1,2)	4813.495
ARIMAX(1,1,1)	4814.27
SARIMAX(0,1,2)(1,0,0)	4792.183
SARIMAX(1,1,1)(1,0,0)	4793.218

Tabla 3: Comparación del AIC de los modelos estimados

Se puede observar que el modelo con menor AIC es el SARIMAX(0, 1, 2)(1, 0, 0)₁₂, por lo cual se eligió para realizar la comparación con los modelos dinámicos. El modelo estimado fue el siguiente:

$$\begin{aligned}
Y_t &= -20.846 - 0.272X_{t-11} + 0.376X_{t-12} + \tau_t \\
\tau_t &\sim SARIMA_{12}(0, 1, 2)(1, 0, 0) \\
\tau_t &= (1 - \Phi_1 B^{12})(1 - B)Y_t = (1 + \theta_1 B + \theta_2 B^2)a_t \\
\tau_t &= (1 - 0.342B^{12})(1 - B)Y_t = (1 - 0.811B + 0.196B^2)a_t \\
a_t &\sim RB(0, \hat{\sigma} = 5729.323)
\end{aligned}$$

3.2. Estimación del modelo dinámico polinomial de primer orden

Se realizó la estimación de tres modelos polinomiales de primer orden para el consumo de combustible en Colombia, considerando constante la varianza de la distribución de probabilidad de Y_t como la varianza de la distribución a priori θ_t .

Modelo 1:

$$\begin{aligned}
Y_t &= \theta_t + \gamma_t, \quad \gamma_t \sim N(0, V = 10) \\
\theta_t &= \theta_{t-1} + W_t, \quad W_t \sim N(0, W = 10) \\
\frac{W}{V} &= 1
\end{aligned}$$

Modelo 2:

$$\begin{aligned}
Y_t &= \theta_t + \gamma_t, \quad \gamma_t \sim N(0, V = 1000) \\
\theta_t &= \theta_{t-1} + W_t, \quad W_t \sim N(0, W = 10) \\
\frac{W}{V} &= 0.01
\end{aligned}$$

Modelo 3:

$$\begin{aligned}
Y_t &= \theta_t + \gamma_t, \quad \gamma_t \sim N(0, V = 10) \\
\theta_t &= \theta_{t-1} + W_t, \quad W_t \sim N(0, W = 1000) \\
\frac{W}{V} &= 100
\end{aligned}$$

Se estimaron tres modelos polinomiales de primer orden, el primero asumiendo la razón entre las varianzas $\frac{W}{V} = 1$, el segundo con $\frac{W}{V} = 0.01$ y el tercero con $\frac{W}{V} = 100$.

Se realizó la comparación de estos modelos a través del MAPE, obteniéndose los siguientes resultados:

La estimación de los modelos polinomiales de primer orden para el consumo de combustible en Colombia reveló que el modelo 3, con una razón entre las varianzas de $\frac{W}{V} = 100$, fue el que presentó un mejor ajuste según el criterio del MAPE. Esto sugiere que un mayor peso en la varianza del proceso de evolución θ_t , en comparación con la varianza del error de observación γ_t , proporciona una mejor representación de la dinámica del consumo de combustible en el contexto estudiado. Estos resultados subrayan la importancia de la selección adecuada de los parámetros en modelos dinámicos para lograr predicciones precisas y útiles en el análisis de series temporales económicas.

Tabla 4: Medidas de comparación

Modelo	MLD 1 (W/V = 1)	MLD 2 (W/V = 0.01)	MLD 3 (W/V = 100)
MAPE	3.417	3.999	3.402

3.3. Elección del mejor modelo de regresión lineal dinámico

Bajo la estructura de los modelos de regresión lineal dinámicos:

$$Y_t = \beta_{0t} + \beta_{1t}X_{1t} + \dots + \beta_{kt}X_{kt} + \varepsilon_t, \quad \text{donde } \varepsilon_t \sim N(0, \phi_t) \tag{1}$$

Donde:

$$\begin{aligned} \beta_{it} &= \beta_{i,t-1} + W_{it} \\ W_{it} &\sim N(c_{it} = 0, v_{it} = 0.1) \\ \phi_t &\sim \text{Gamma}(\alpha = 0.1, \beta = 0.1) \end{aligned}$$

Teniendo en cuenta la tabla 1 sobre las variables de estudio, se propusieron varios modelos de regresión lineal dinámico bajo el supuesto de normalidad en la variable respuesta y se compararon en términos del DIC, donde se obtuvieron los siguientes resultados.

Tabla 5: Resultados de la estimación

Modelo	Definición	DIC	MAPE
1	$Y_t = \beta_{0,t} + \beta_{1,t}X_{1,t}$	12381.67	1.83
2	$Y_t = \beta_{0,t} + \beta_{1,t}X_{2,t}$	15077.26	1.02
3	$Y_t = \beta_{0,t} + \beta_{1,t}X_{1,t} + \beta_{2,t}X_{2,t}$	4808.73	0.92
4	$Y_t = \beta_{0,t} + \beta_{1,t}X_{3,t}$	7960.18	2.02
5	$Y_t = \beta_{0,t} + \beta_{1,t}X_{1,t} + \beta_{2,t}X_{3,t}$	9354.15	1.97
6	$Y_t = \beta_{0,t} + \beta_{1,t}X_{2,t} + \beta_{2,t}X_{3,t}$	13645.13	1.76

La comparación de los modelos de regresión lineal dinámicos para el consumo de combustible en Colombia reveló que el modelo 3, que incluye las variables $X_{1,t}$ y $X_{2,t}$, presentó el mejor ajuste según el criterio del DIC con un valor de 4808.73 y el menor MAPE de 0.92%. Esto indica que la combinación de estas dos variables explicativas proporciona una representación más precisa de la dinámica del consumo de combustible en comparación con los otros modelos evaluados. Estos resultados subrayan la importancia de seleccionar cuidadosamente las variables y la estructura del modelo para lograr predicciones precisas en el análisis de series temporales en el contexto económico.

3.4. Elección del mejor modelo de regresión dinámico generalizado Poisson

Bajo la estructura de los modelos de regresión dinámicos Poisson:

$$\begin{aligned}
Y_t &\sim \text{Poisson}(\lambda_t) \\
\lambda_t &= e^{\beta_{0t} + \beta_{1t}X_{1t} + \beta_{2t}X_{2t} + \dots + \beta_{kt}X_{kt}} \\
\beta_{kt} &= \beta_{k,t-1} + W_{kt} \\
\beta_{k,t-1} &\sim N(0, 0.1) \\
W_{kt} &\sim \text{Gamma}(\alpha_0 = 0.1, \beta_0 = 0.1)
\end{aligned}$$

Teniendo en cuenta la tabla 1 sobre las variables de estudio, se propusieron varios modelos de regresión dinámico Poisson y se compararon en términos del DIC y el MAPE, donde se obtuvieron los siguientes resultados.

Tabla 6: Resultados de la estimación

Modelo	Definición	DIC	MAPE
1	$\lambda_t = e^{\beta_{0t} + \beta_{1t}X_{1t}}$	385.46	2.55
2	$\lambda_t = e^{\beta_{0t} + \beta_{1t}X_{2t}}$	11388.96	5.60
3	$\lambda_t = e^{\beta_{0t} + \beta_{1t}X_{1t} + \beta_{2t}X_{2t}}$	10876.22	3.83
4	$\lambda_t = e^{\beta_{0t} + \beta_{1t}X_{3t}}$	11860.23	6.27
5	$\lambda_t = e^{\beta_{0t} + \beta_{1t}X_{1t} + \beta_{2t}X_{3t}}$	9357.56	4.19
6	$\lambda_t = e^{\beta_{0t} + \beta_{1t}X_{2t} + \beta_{2t}X_{3t}}$	8365.46	5.76

La comparación de los modelos de regresión dinámico generalizado Poisson para el consumo de combustible en Colombia demostró que el modelo 1, que solo incluye la variable X_{1t} , presentó el mejor ajuste según los criterios del DIC y el MAPE, con valores de 385.46 y 2.55 % respectivamente. Este resultado sugiere que la variable X_{1t} es un predictor significativo y suficiente para modelar la dinámica del consumo de combustible en el contexto estudiado. La selección de un modelo apropiado es crucial para obtener predicciones precisas y para la toma de decisiones basada en datos en el ámbito de la economía y la gestión de recursos.

3.5. Elección del mejor modelo jerárquico

Bajo la estructura del modelo de regresión jerárquico:

$$\begin{aligned}
Y_t &= \beta_t Y_{t-1} + \epsilon_t, \quad \text{donde } \epsilon_t \sim N(0, \sigma^2) \\
\text{donde } \beta_t &= \alpha_{0t} + \alpha_{1t}X_t + w_t, \quad \text{con } w_t \sim N(0, \sigma_w^2) \\
\text{con } \alpha_{0t} &= \alpha_{0,t-1} + v_{1t}, \quad \text{con } v_{1t} \sim N(0, \sigma_{v1}^2) \\
\text{con } \alpha_{1t} &= \alpha_{1,t-1} + v_{2t}, \quad \text{con } v_{2t} \sim N(0, \sigma_{v2}^2)
\end{aligned}$$

Luego de la evaluación de los tres modelos jerárquicos propuestos para el análisis de la serie temporal, se determinó que el modelo 1, que incorpora la variable X_{1t} , presentó el mejor ajuste de acuerdo con los criterios del DIC y el MAPE, con valores de 62681.9 y 0.12 % respectivamente. Este resultado

Tabla 7: Resultados de la estimación

Modelo	Definición	DIC	MAPE
1	$Y_t = \beta_t Y_{t-1} + \epsilon_t$, donde $\beta_t = \alpha_{0t} + \alpha_{1t} X_{1t}$	62681.9	0.12
2	$Y_t = \beta_t Y_{t-1} + \epsilon_t$, donde $\beta_t = \alpha_{0t} + \alpha_{1t} X_{2t}$	337284	0.20
3	$Y_t = \beta_t Y_{t-1} + \epsilon_t$, donde $\beta_t = \alpha_{0t} + \alpha_{1t} X_{3t}$	431368	1.24

destaca la importancia de la variable X_{1t} en la modelación del fenómeno estudiado y subraya la utilidad del enfoque jerárquico en la captura de la dinámica subyacente de la serie temporal.

3.6. Comparación de los modelos

En la sección anterior se realizó la elección del mejor modelo para cada grupo. A continuación, se compararán estos modelos con el fin de elegir el mejor modelo.

Los modelos se compararon a través del MAPE, y los resultados se presentan a continuación.

Tabla 8: Modelos definitivos

Familia del modelo	Definición	MAPE
SARIMAX	SARIMAX(0, 1, 2)(1, 0, 0) ₁₂	4.246 %
Modelo Polinomial Primer Orden (M.P.P)	$Y_t = \theta_t + \gamma_t$, con $\frac{W}{V} = 100$	3.402 %
Modelo Lineal Dinámico (M.L.D) Normal	$Y_t = \beta_{0,t} + \beta_{1,t} X_{1,t} + \beta_{2,t} X_{2,t}$	0.92 %
M.L.D Poisson	$\lambda_t = e^{\beta_{0,t} + \beta_{1,t} X_{1,t}}$	2.55 %
Modelo jerárquico	$Y_t = \beta_t Y_{t-1} + \epsilon_t$, donde $\beta_t = \alpha_{0t} + \alpha_{1t} X_{1t}$	0.12 %

Se pudo observar que el modelo que presentó un menor MAPE fue el modelo de regresión lineal dinámico jerárquico con un 0.12%. Además, al realizar la gráfica de los parámetros β del modelo, se pudo observar que estos fueron significativos. Por lo tanto, el mejor modelo estimado corresponde al modelo lineal jerárquico utilizando la variable X_1 (vehículos vendidos y matriculados en Colombia) como variable explicativa durante el periodo de tiempo de estudio.

3.7. Modelo estimado

El modelo estimado tiene la siguiente forma:

$$Y_t = \beta_t Y_{t-1} + \epsilon_t, \quad \text{donde } \epsilon_t \sim N(0, \sigma^2) \tag{2}$$

$$\beta_t = \alpha_{0t} + \alpha_{1t} X_t + w_t, \quad \text{con } w_t \sim N(0, \sigma_w^2) \tag{3}$$

$$\alpha_{0t} = \alpha_{0t-1} + v_{1t}, \quad \text{con } v_{1t} \sim N(0, \sigma_{v1t}^2) \tag{4}$$

$$\alpha_{1t} = \alpha_{1t-1} + v_{2t}, \quad \text{con } v_{2t} \sim N(0, \sigma_{v_{2t}}^2) \quad (5)$$

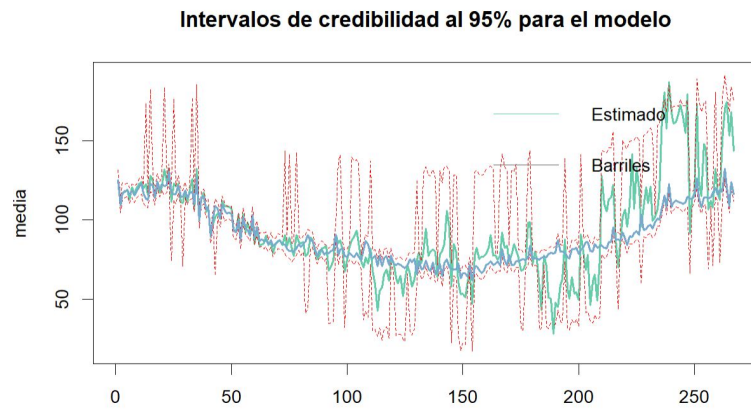


Figura 3: Intervalos de confianza para el modelo estimado y los datos reales. Fuente: DNP. Elaboración: Propia.

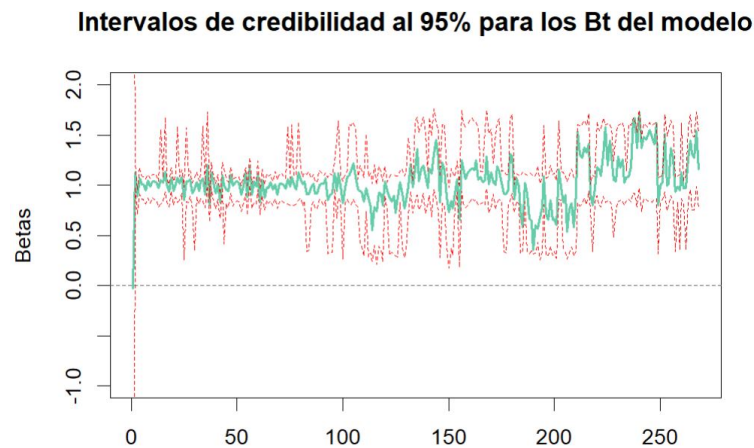


Figura 4: Intervalos de confianza para los β_0 estimados. Fuente: DNP. Elaboración: Propia.

El análisis de los modelos propuestos para el consumo de combustible en Colombia reveló que el modelo de regresión lineal dinámico jerárquico, utilizando la variable de vehículos vendidos y matriculados como variable explicativa, presentó el mejor ajuste con un MAPE de 0.12%. Las gráficas del modelo estimado y los parámetros β mostraron que estos son significativos, lo que refuerza la validez del modelo seleccionado. Este resultado subraya la importancia de la selección adecuada de variables y estructuras de modelos en el análisis de series temporales para lograr

predicciones precisas y útiles en la toma de decisiones relacionadas con la gestión de recursos y la planificación económica.

4. Conclusiones

En el análisis del consumo de combustible en Colombia, se aplicaron diferentes metodologías para la selección del modelo más adecuado. Utilizando la metodología frecuentista, se encontró que el modelo SARIMAX(0,1,2)(1,0,0)₁₂, que incorpora la variable de compra y matrícula de vehículos, proporciona un buen ajuste al comportamiento de los datos. Además, en la estimación de modelos dinámicos polinomiales de primer orden, se observó que el mejor ajuste se logra cuando la relación W/V es igual a 100. Por otro lado, en los modelos dinámicos normales, el modelo $Y_t = \beta_{0,t} + \beta_{1,t}X_{1,t} + \beta_{2,t}X_{2,t}$ demostró tener el mejor ajuste.

En cuanto a los modelos dinámicos Poisson, el mejor ajuste se obtuvo con el modelo $\lambda_t = e^{\beta_{0,t} + \beta_{1,t}X_{1,t}}$. Por otra parte, en la estimación de los modelos jerárquicos, se encontró que el modelo $Y_t = \beta_t Y_{t-1} + \epsilon_t$, donde $\beta_t = \alpha_{0t} + \alpha_{1t}X_{1t}$, presentó el mejor ajuste. Este hallazgo resalta la importancia de la variable de venta de vehículos en la estimación del consumo de combustible en el país.

Después de comparar diferentes modelos y metodologías, se concluyó que el modelo lineal dinámico jerárquico, utilizando la variable de venta de vehículos como variable explicativa, ofreció el mejor ajuste en términos del MAPE y la significancia de los parámetros. Esto subraya la relevancia de seleccionar cuidadosamente el modelo y las variables explicativas para obtener estimaciones precisas y significativas en el análisis del consumo de combustible en Colombia.

Recibido: noviembre 2022

Aceptado: mayo 2023

Referencias

- J. Durbin and S. J. Koopman. Time series analysis by state space methods. *Oxford Statistical Science Series*, 24, 2001.
- H. J. Fuentes, Y. Muñoz, and L. E. Contreras. Un análisis del mercado de gasolina en Colombia para el periodo 2000-2018. *Revista Espacios*, 41(15), 2020.
- J. J. García, D. Pérez, M. Orrego, J. M. Castaño, et al. Un modelo casi ideal de demanda de combustibles para la industria de transporte. 2016.
- A. C. Harvey. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, 1990.
- V. Montiel, A. Machado, C. Araujo, and Ochoa. Determinación del consumo de combustible en un vehículo que utiliza gasolina o gnc mediante la simulación de su comportamiento dinámico. *Revista Técnica de la Facultad de Ingeniería Universidad del Zulia*, 25(3):158–165, 2002.

- G. Petris, S. Petrone, and P. Campagnoli. Dynamic linear models with r. *New York: Springer*, 2009.
- J. Posada, I. Sarmiento, and A. Correa. Consumo de combustible en camiones según peso del vehículo y otras variables. *Universidad Nacional de Colombia: Bogotá, Colombia*, 2014.
- N. A. Vera Meneses. Metodología para el modelaje estático de una supergrid latinoamericana. 2017.
- M. West and J. Harrison. *Bayesian Forecasting and Dynamic Models*. Springer, 2nd edition, 1997.