



## Empleando modelos jerárquicos para encontrar el mejor modelo para pronosticar los galones de gasolina corriente demandados en Bogotá (Colombia)

ALONSO CIFUENTES, JULIO CÉSAR

Universidad Icesi, Cali (Colombia)

Correo electrónico: jcalonso@icesi.edu

DÍAZ, JAVIER GUSTAVO

Universidad Icesi, Cali (Colombia)

Correo electrónico: jgdiaz@icesi.edu

ESTRADA, DANIELA

Universidad Icesi (Cali) y Alianza Coba (Bogotá), Colombia.

Correo electrónico: destrada@icesi.edu

FIGUEROA, CÉSAR ALFONSO

Oficina de Inteligencia Tributaria, Bogotá (Colombia)

Correo electrónico: cfigueroa@shd.gov.co

TAMURA, GABRIEL

Universidad Icesi, Cali (Colombia)

Correo electrónico: gtamura@icesi.edu

### RESUMEN

El documento tiene como objetivo encontrar el mejor modelo jerárquico que permita proyectar la demanda total de gasolina corriente y por tanto el recaudo por sobretasa a la gasolina en Bogotá, Colombia, impuesto importante para el financiamiento de la malla vial y sistemas de transporte masivos. Para lograr este objetivo, se emplean datos de los galones reportados por los 6 mayoristas de gasolina corriente de la ciudad bajo dos aproximaciones univariadas (ARIMA y el método de suavizamiento exponencial (ETS por sus siglas en inglés)), cinco métodos y diferentes algoritmos de minimización. Se encuentra que la mejor combinación de estos parámetros para pronosticar los galones de gasolina corriente demandados es el modelo ETS bajo un pronóstico univariado simple.

**Palabras clave:** Colombia, gasolina, modelos jerárquicos, series de tiempo, pronósticos.

**Clasificación JEL:** C22; C53; H2.

**MSC2010:** 62P20; 00A99.

# Use of hierarchical models to find the best model to forecast the gallons of regular gasoline demanded in Bogotá (Colombia)

## ABSTRACT

The objective of this analysis is to find the best hierarchical model to forecast the total demand for regular gasoline in Bogotá, Colombia and, therefore, the collection of gasoline surcharges, which is an important tax used to finance road networks and massive transportation systems. We used data reported by 6 wholesalers of regular gasoline in the city, and used two univariate approaches (ARIMA and exponential smoothing (ETS)), five methods and different minimization algorithms to forecast gallons of regular gasoline. Results show that the best combination of these parameters is an ETS model under a simple univariate forecast.

**Keywords:** Colombia, gasoline, hierarchical models, time series, forecasts.

**JEL classification:** C22; C53; H2.

**MSC2010:** 62P20; 00A99.



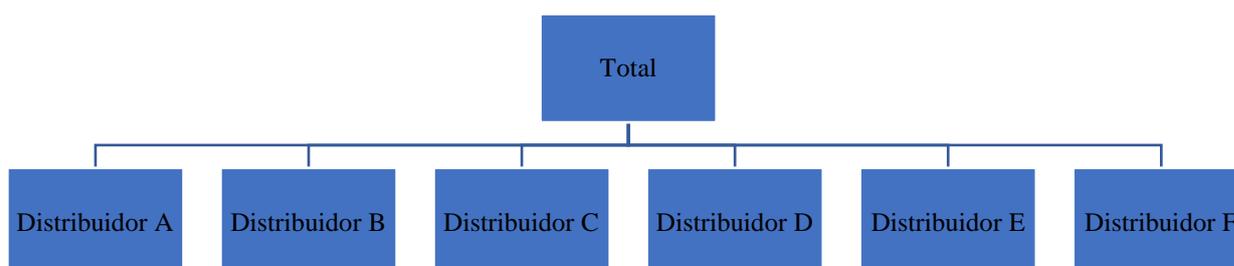
## 1. Introducción.

Construir un pronóstico es un ejercicio relativamente común en los negocios y en el sector público. En el sector público, la construcción de un buen pronóstico permite una mejor planificación de los recursos y una programación financiera eficiente. En Colombia, uno de los tres principales tributos que financian a las municipalidades es la sobretasa a la gasolina. Por ejemplo, se estima que la sobretasa a la gasolina tiene una participación entre el 8% y el 12% dentro de los ingresos tributarios de los municipios (Yepes, Martínez, & Aguilar, 2014). Los ingresos provenientes de este impuesto son utilizados para financiar proyectos estratégicos en mantenimiento y construcción de infraestructura vial y la financiación de proyectos de transporte masivo.

Este tributo tiene como base de liquidación el producto entre el volumen de gasolina vendido en una municipalidad (medido en galones) y un precio de referencia establecido por el Gobierno Nacional por medio de una resolución. La tasa de este impuesto está establecida por una Ley de carácter nacional que no permite a las municipalidades modificarla. De esta manera, para proyectar los ingresos por sobretasa a la gasolina en una municipalidad en Colombia, se requiere únicamente proyectar el volumen de gasolina que será vendido.

Por otro lado, unas de las peculiaridades de este tributo en el caso colombiano, es que desde la Ley 488 de 1998 (24 de diciembre) se estableció que los responsables de la recaudación son los distribuidores mayoristas (antes eran los minoristas). Esta definición del responsable de la recaudación implica que la fiscalización del tributo se hace relativamente fácil. Por ejemplo, en la ciudad de Bogotá, se contaba en enero de 2017 con seis grandes mayoristas de gasolina corriente y cinco de gasolina extra; estos responsables representaban más del 90% de la recaudación. Así, el volumen total de galones de gasolina corriente o extra corresponde a la suma de los galones distribuidos por cada mayorista. Siguiendo a Hyndman, Ahmed, Athanasopoulos y Shang (2011) esta estructura corresponde a una estructura jerárquica simple como la que se representa en la Figura 1.

**Figura 1. Estructura Jerárquica de la demanda de gasolina.**



Fuente: Elaboración propia.

Este documento tiene como objetivo explotar la naturaleza jerárquica de la demanda total de gasolina en la ciudad de Bogotá para encontrar el mejor modelo jerárquico que permita proyectar la demanda total de gasolina corriente expresada en galones y, por tanto, la recaudación por sobretasa a la gasolina.

Hasta donde llega el conocimiento de los autores no existe ningún trabajo publicado previamente que intente caracterizar estadísticamente o generar pronósticos de la demanda de galones de gasolina para Bogotá o nivel nacional. Solo se encuentran dos trabajos, Alonso y Solano (2006) y Alonso, Solano y Duque (2007), que modelan el comportamiento de la recaudación total del impuesto de la sobretasa a la gasolina a partir de proyecciones de las ventas en galones de gasolina corriente y extra para la ciudad de Cali. Estos autores no presentan un

análisis desagregado a nivel de distribuidor. A nivel internacional tampoco se encuentran referencias de un ejercicio similar al propuesto aquí para la demanda de gasolina.

Este documento está compuesto por cinco secciones adicionales. La siguiente sección discute diferentes opciones para encontrar el mejor modelo para pronosticar la demanda total de gasolina corriente teniendo en cuenta la estructura jerárquica. La tercera sección describe los datos empleados y nuestra aproximación. La cuarta sección muestra los resultados. El documento concluye con unos comentarios finales.

## 2. Pronósticos de modelos jerárquicos.

Siguiendo a Hyndman, Ahmed, Athanasopoulos y Shang (2011), la naturaleza jerárquica de los datos se puede expresar de la siguiente manera:

$$\begin{bmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{C,t} \\ y_{D,t} \\ y_{E,t} \\ y_{F,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{A,t} \\ y_{B,t} \\ y_{C,t} \\ y_{D,t} \\ y_{E,t} \\ y_{F,t} \end{bmatrix}, \quad (1)$$

donde  $y_t$  representa la serie total en el periodo  $t$  y  $y_{i,t}$  representa el número de galones que vende el distribuidor  $i$  en el periodo  $t$ , con  $i \in \{A, B, C, D, E, F\}$ . Si se define el vector  $\mathbf{y}_{k,t}$  como todas las observaciones del nivel inferior del diagrama 1 en el periodo  $t$ , (1) puede expresarse como:

$$\mathbf{y}_t = \mathbf{S}\mathbf{y}_{k,t} \quad (2)$$

donde  $\mathbf{S}$  representa la matriz de suma dada por:

$$\mathbf{S} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

En nuestro caso estamos interesados en generar pronósticos  $h$  pasos adelante (*h-step-ahead*) del Total (primer elemento de  $\mathbf{y}_t$ ) dado que ya hemos observado  $T$  periodos de la serie ( $\hat{y}_{T+h|T}$ ).

$\hat{y}_{T+h|T}$  puede ser calculado de diferentes maneras. Por ejemplo, se puede pronosticar directamente por medio de un modelo univariado para el nivel total. Esta opción se denominará pronóstico inicial. Otra opción es emplear métodos univariados para la base (*bottom* en inglés) y combinarlos para obtener la parte de arriba (*Top* en inglés). A esta aproximación se le denomina Bottom to Top (BTT). Es decir,

$$\tilde{y}_{T+h|T} = \mathbf{S}\hat{y}_{k,T+h|T} \quad (2)$$

Una ventaja de esta aproximación es que se emplea toda la información de la base. Pero esto tiene un costo: tener que proyectar series desagregadas que pueden ser más volátiles (ver Alonso y Rivera (2017) para una mayor discusión).

Otra aproximación es la propuesta por Hyndman et al. (2011) denominada combinación óptima de pronósticos (*Optimal forecast combination* en inglés, de aquí en adelante OFC). Esta aproximación implica combinar los pronósticos para las series individuales de la base y no sumarlos. La intuición detrás de esto radica en que al generarse de manera individual cada uno de los pronósticos de la base, no necesariamente estos sumarán de manera consistente el pronóstico de arriba. Así, combinarlos podría generar una mejor aproximación que sumarlos.

Hyndman et al. (2011) proponen entonces emplear una aproximación análoga a la regresión múltiple. Es decir,

$$\hat{\mathbf{y}}_{T+h|T} = \mathbf{S}\boldsymbol{\beta}_{K,T+h|T} + \boldsymbol{\varepsilon}_{K,T+h|T}, \quad (3)$$

donde  $\boldsymbol{\beta}_{K,T+h|T}$  corresponde al vector desconocido de medias para los valores pronosticados de la base también denominados los pesos y  $\boldsymbol{\varepsilon}_{K,T+h|T}$  es un vector con vector de medias cero y matriz de varianzas y covarianzas  $\boldsymbol{\Sigma}_{K,T+h|T}$ . En otras palabras,  $\boldsymbol{\varepsilon}_{K,T+h|T}$  es el error de la regresión.

Hyndman et al. (2011) muestran que el mejor estimador lineal insesgado para  $\boldsymbol{\beta}_{K,T+h|T}$ , si  $\boldsymbol{\Sigma}_{K,T+h|T}$  es esférico, es

$$\hat{\boldsymbol{\beta}}_{K,T+h|T} = (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T\hat{\mathbf{y}}_{T+h|T}. \quad (4)$$

Es decir, el estimador de Mínimos Cuadrados Ordinarios (OLS por su sigla en inglés). Por lo tanto, el pronóstico OFC empleando OLS (OFC-OLS) para el nivel total corresponderá a:

$$\hat{\mathbf{y}}_{T+h|T}^{OLS} = \mathbf{S}(\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T\hat{\mathbf{y}}_{T+h|T}. \quad (4)$$

Hyndman, Lee y Wang (2016) sugieren una aproximación diferente para calcular  $\boldsymbol{\beta}_{K,T+h|T}$  que reconozca la posibilidad de que  $\boldsymbol{\Sigma}_{K,T+h|T}$  no sea esférica y pueda presentar algún tipo de heteroscedasticidad. Los autores sugieren emplear mínimos cuadrados ponderados (WLS por su sigla en inglés) donde los pesos están dados por las varianzas estimadas de cada elemento de  $\boldsymbol{\varepsilon}_{K,T+h|T}$ . Es decir, el pronóstico OFC empleando WLS (OFC-WLS) estaría dado por

$$\hat{\mathbf{y}}_{T+h|T}^{WLS} = \mathbf{S}(\mathbf{S}^T\boldsymbol{\Lambda}_{T+h|T}\mathbf{S})^{-1}\mathbf{S}^T\boldsymbol{\Lambda}_{T+h|T}\hat{\mathbf{y}}_{T+h|T}, \quad (5)$$

donde  $\boldsymbol{\Lambda}_{T+h|T}$  es una matriz diagonal cuyos elementos corresponden a la inversa de cada una de las varianzas estimadas del respectivo término de  $\boldsymbol{\varepsilon}_{K,T+h|T}$ .

Por otro lado, Wickramasuriya, Athanasopoulos y Hyndman (2015) sugieren emplear un estimador de mínimos cuadrados generalizados (GLS por su sigla en inglés) para  $\boldsymbol{\beta}_{K,T+h|T}$ . Para simplificar el problema sugieren emplear una reconciliación empleando la traza mínima (MinT) para calcular la inversa generalizada de  $\boldsymbol{\Sigma}_{K,T+h|T}$ . Es decir, el pronóstico OFC empleando GLS por medio del método MinT (OFC-MinT) estaría dado por

$$\hat{\mathbf{y}}_{T+h|T}^{MinT} = \mathbf{S}(\mathbf{S}^T\boldsymbol{\Sigma}_{K,T+h|T}^G\mathbf{S})^{-1}\mathbf{S}^T\boldsymbol{\Sigma}_{K,T+h|T}^G\hat{\mathbf{y}}_{T+h|T}, \quad (6)$$

donde  $\boldsymbol{\Sigma}_{K,T+h|T}^G$  es la inversa generalizada de  $\boldsymbol{\Sigma}_{K,T+h|T}$  calculada empleando la aproximación MinT.

Así tendremos seis opciones de pronóstico para el total ( $y_t$  que corresponde al primer elemento de  $\mathbf{y}_t$ ):

- i) el pronóstico univariado del Total (pronóstico inicial  $-\hat{y}_{T+h|T}$ ),
- ii) el de abajo hacia arriba (BTT) que corresponde a la suma simple de los pronósticos individuales de cada una de las series de la base (primer elemento de  $\tilde{y}_{T+h|T}$ ),
- iii) la combinación óptima de los pronósticos empleando OLS (OFC-OLS) individuales de cada una de las series de la base (primer elemento de  $\hat{y}_{T+h|T}^{OLS}$ ),
- iv) la combinación óptima de los pronósticos empleando WLS (OFC-WLS) individuales de cada una de las series de la base (primer elemento de  $\hat{y}_{T+h|T}^{WLS}$ ),
- v) la combinación óptima de los pronósticos empleando MinT (OFC-MinT) individuales de cada una de las series de la base (primer elemento de  $\hat{y}_{T+h|T}^{MinT}$ ). Para el método OFC-MinT se emplearán dos métodos para calcular la matriz de varianzas y covarianzas de  $\epsilon_{K,T+h|T}$ : una es la matriz de varianzas y covarianzas muestral (OFC-MinT-sam) y
- vi) otra es emplear un estimador de reducción hacia la diagonal (en inglés *shrinkage estimator with shrinkage towards the diagonal*) (OFC-MinT-shr).

### 3. Caso de estudio.

Para realizar nuestro ejercicio se cuenta con una base de datos que contiene los galones vendidos mensualmente por los seis distribuidores de gasolina corriente de la ciudad de Bogotá. Por razones de confidencialidad, no es posible mostrar gráficas o estadísticas de las series de los distribuidores de manera individual. La base de datos contiene datos desde enero de 2006 hasta febrero de 2017. La información es provista por la Oficina de Inteligencia Tributaria de la Secretaría Distrital de Hacienda de Bogotá. Se emplea una muestra desde enero de 2006 hasta febrero de 2016 para la estimación de los modelos y se guardará la muestra de marzo de 2016 a febrero de 2017 para evaluar el desempeño de los pronósticos.

Nuestro ejercicio implica los siguientes pasos:

1. Emplear la muestra enero-2006 a enero-2016 para estimar modelos univariados para la serie total ( $y_t$ ) y las series de la base ( $y_{i,t}$  con  $i \in \{A, B, C, D, E, F\}$ ).
2. Realizar un pronóstico para el siguiente mes del total ( $y_{T+1}$ ) empleando el pronóstico univariado (pronóstico inicial), un pronóstico BTT, uno OFC-OLS, uno OFC-WLS y uno OFC-MinT.
3. Agregar un dato más a la muestra.
4. Repetir 2 a 3 hasta agotar la muestra.

El ejercicio de pronósticos univariados para cada serie se repite empleando modelos autoregresivos integrados de media móvil (ARIMA por su sigla en inglés) y modelos de suavización exponencial (ETS por su sigla en inglés).

Por otro lado, para los casos de los OFC se requiere un proceso de minimización. Para encontrar numéricamente el mínimo se adoptan los siguientes algoritmos: Descomposición LU (del término en inglés Lower-Upper), el método de gradiente conjugado (CG por su sigla en inglés) y la descomposición de Cholesky (CHOL). Es decir, tendremos las siguientes opciones: OFC-OLS-LU, OFC-OLS-CG, OFC-OLS-CHOL, OFC-WLS-LU, OFC-WLS-CG, OFC-WLS-CHOL, OFC-MinT-sam-LU, OFC-MinT-sam-CG, OFC-MinT-sam-CHOL, OFC-MinT-shr-LU, OFC-MinT-shr-OLS y OFC-MinT-shr-CHOL.

Finalmente, para cada una de las posibles combinaciones de aproximaciones univariadas (ARIMA y ETS), métodos (pronóstico inicial, BTT, OFC-OLS, OFC-WLS, OFC-MinT-sam y OFC-MinT-shr) y los respectivos algoritmos de minimización, se tendrán 12 pronósticos para el siguiente mes (para el periodo marzo de 2016 a febrero de 2017) que son comparados con los valores realmente observados.

Para la comparación de los pronósticos emplearemos las métricas tradicionales que se muestran en la Tabla 1.

**Tabla 1. Métricas para evaluar la bondad de ajuste de las aproximaciones.**

Métrica	Fórmula
Mean absolute error (MAE)	$\frac{1}{T} \sum_{t=1}^T  Y_t - \hat{Y}_t $
Mean absolute percent error (MAPE)	$\left( \frac{1}{T} \sum_{t=1}^T \left  \frac{Y_t - \hat{Y}_t}{Y_t} \right  \right) * 100\%$
Root mean square error (RMSE)	$\sqrt{\frac{1}{T} \sum_{t=1}^T (Y_t - \hat{Y}_t)^2}$
Mean percentage error (MPE)	$\left( \frac{1}{T} \sum_{t=1}^T \frac{Y_t - \hat{Y}_t}{Y_t} \right) * 100\%$

Fuente: Elaboración propia.

#### 4. Resultados.

En la Tabla 2 y en la Tabla 3 se presentan los resultados de cada una de las métricas para las dos aproximaciones univariadas consideradas.

**Tabla 2. Resultados de las métricas tradicionales para aproximación univariada ARIMA.**

Método	RMSE	MAE	MPE	MAPE
Inicial	1045641.000	761697.400	0.066	2.924
BTT	1014550.700	805987.900	1.361	3.047
OFC-WLS-LU	972639.800	724529.000	0.677	2.765
OFC-OLS-LU	1002539.900	779520.300	1.197	2.960
OFC-MinT-shr-lu	918829.100	682283.900	-0.142	2.620
OFC-MinT-sam-lu	897899.600	708053.700	-0.381	2.716
OFC-WLS-CG	972639.800	724529.000	0.677	2.765
OFC-OLS-CG	1002539.900	779520.300	1.197	2.960
OFC-MinT-shr-CG	918829.100	682283.900	-0.142	2.620
OFC-MinT-sam-CG	897899.600	708053.700	-0.381	2.716
OFC-WLS-CHOL	972639.800	724529.000	0.677	2.765
OFC-OLS-CHOL	1002539.900	779520.300	1.197	2.960
OFC-MinT-shr-CHOL	918829.100	682283.900	-0.142	2.620
OFC-MinT-sam-CHOL	897899.600	708053.700	-0.381	2.716

Fuente: Elaboración propia.

Bajo la aproximación univariada ARIMA se obtiene que los métodos OFC-MinT-sam-LU, OFC-MinT-sam-CG, y OFC-MinT-sam-CHOL son los mejores si se emplean los criterios RMSE, MAE o MAPE. En cambio, si se utiliza el criterio MPE el mejor método es el inicial, es decir, pronosticar la gasolina corriente total empleando un modelo ARIMA.

Bajo la aproximación univariada ETS se obtiene que el mejor método es el inicial si se emplean los criterios RMSE, MAE o MAPE. Por el contrario, si se utiliza el criterio MPE hay dos métodos que deberían utilizarse: OFC-MinT-shr-LU y OFC-MinT-shr-CHOL.

Por último, si se comparan las dos aproximaciones (ARIMA y ETS) se obtiene que el mejor método bajo la métrica RMSE, MAE y MAPE es el método inicial aplicado a un modelo univariado ETS. Mientras que por la métrica MPE también el modelo a escoger sería el ETS, pero empleando el método OFC-MinT-shr-LU o OFC-MinT-shr-CHOL.

**Tabla 3. Resultados de las métricas tradicionales para aproximación univariada ETS.**

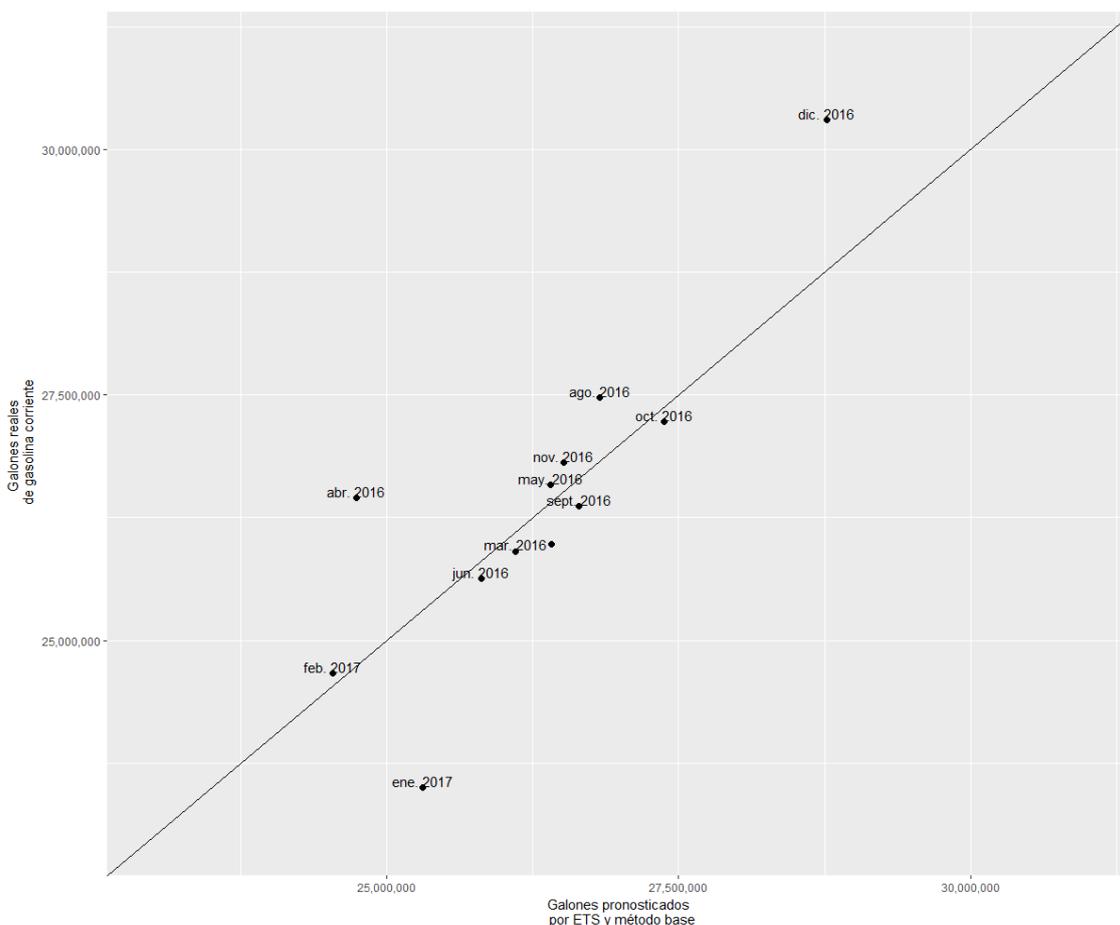
<b>Método</b>	<b>RMSE</b>	<b>MAE</b>	<b>MPE</b>	<b>MAPE</b>
Inicial	829110.300	591241.600	0.189	2.250
BTT	885758.100	626596.300	0.325	2.377
OFC-WLS-LU	860611.100	605753.500	0.117	2.312
OFC-OLS-LU	882139.900	623542.400	0.273	2.377
OFC-MinT-shr-LU	851038.800	601176.100	0.021	2.297
OFC-MinT-sam-LU	864678.400	605858.000	0.114	2.311
OFC-WLS-CG	860611.100	605753.500	0.117	2.312
OFC-OLS-CG	882139.900	623542.400	0.273	2.377
OFC-MinT-sam-CG	864678.400	605858.000	0.114	2.311
OFC-WLS-CHOL	860611.100	605753.500	0.117	2.312
OFC-OLS-CHOL	882139.900	623542.400	0.273	2.377
OFC-MinT-shr-CHOL	851038.800	601176.100	0.021	2.297
OFC-MinT-sam-CHOL	864678.400	605858.000	0.114	2.311

Nota: No se presentan los resultados del método OFC-MinT-shr-CG porque no converge.

Fuente: Elaboración propia.

En síntesis, si se desea pronosticar la demanda total de gasolina corriente (galones) y, por tanto, la recaudación por sobretasa a la gasolina, el mejor modelo a utilizar sería el ETS bajo el método inicial. En la Figura 2 se puede observar el pronóstico bajo este modelo y método versus el valor real de galones demandados durante el periodo marzo 2016 a febrero de 2017.

**Figura 2. Valor pronosticado vs valor real de galones de gasolina corriente empleando el mejor modelo y método.**



Fuente: Elaboración propia.

## 5. Comentarios finales.

Los resultados obtenidos implican que, entre los métodos comparados y de acuerdo a los criterios elegidos, la mejor forma de pronosticar la demanda total de galones en Bogotá corresponde a pronosticar la serie total por medio de un modelo ETS. En este caso el modelo y la aproximación más sencilla brinda un mejor ajuste a los datos. Es decir, no se gana poder en la predicción de explotar la naturaleza jerárquica de los datos, resultado que a priori no es evidente.

Por otro lado, independientemente de los resultados, es importante reconocer la gran importancia para las municipalidades de contar con pronósticos de sus ingresos tributarios que le dan sentido a este tipo de ejercicios. En especial en el caso colombiano, pronosticar el número de galones de gasolina es sumamente importante porque es un insumo para saber cuánto se va a recaudar por el impuesto de sobretasa a la gasolina. Lo anterior porque este impuesto es el cuarto más importante en los ingresos tributarios de los entes territoriales y es el que les permite financiar la construcción de obras de infraestructura relacionadas con los sistemas de transporte masivo y cubrir el mantenimiento, reparación y ampliación de la malla vial. Luego, mejores pronósticos permiten planear y estructurar presupuestos de las vigencias futuras con un menor margen de error, además de que puede ser una herramienta poderosa para identificar posibles responsables que son inexactos potenciales, es decir, que no están pagando el valor del impuesto correspondiente al valor esperado de galones demandados.

Respecto al primer beneficio que tiene haber encontrado el mejor modelo de acuerdo con los criterios utilizados, se puede utilizar para saber si el municipio o distrito va a tener suficientes recursos o a qué plazo los puede tener para financiar las obras de infraestructura viales o de transporte masivo. Lo anterior se aplica en Colombia en la construcción de las obras de infraestructura y sistemas de transporte masivo, ya que se ha establecido un esquema de cofinanciación entre la Nación y los entes territoriales que consiste en que el 70% de la financiación está a cargo de la nación y el resto a cargo del ente territorial donde se va a ejecutar la obra.

Por ejemplo, en el caso específico de Bogotá, la Administración Distrital se ha comprometido a utilizar el respectivo porcentaje del recaudo de la sobretasa para cubrir la construcción de la primera línea del metro de Bogotá (PLMB). En este sentido, el 30% de la aportación del Distrito a la PLMB, que podría alcanzar los \$4.14 billones de pesos, será cubierto con: cerca de un 19% proveniente del cupo de endeudamiento, el 64% que corresponde al valor presente neto desde el año 2018 y hasta por 20 años del 50% de la recaudación de la sobretasa a la gasolina motor que por ley se puede destinar a los sistemas de transporte masivo y un 17% del pago de dividendos y otras fuentes.

Luego, dados los compromisos que ha adquirido la Administración Distrital para cumplir con su responsabilidad en la financiación de dicha obra, es vital el permanentemente monitoreo de la recaudación de la sobretasa a la gasolina motor real respecto a lo esperado y la identificación de responsables que estén evadiendo el pago parcial del impuesto (es decir, potenciales inexactos) puesto que eso significa un detrimento patrimonial que puede afectar el cumplimiento de las responsabilidades adquiridas por la Administración y por ende de la ejecución de la obra.

Por último, el siguiente paso se enfocará en aplicar la misma metodología a gasolina extra y encontrar el modelo más adecuado para pronosticar la demanda de este tipo de combustible. Asimismo, este estudio se ampliará para incluir el cálculo de los intervalos de pronóstico, los cuales son igual de importantes que los pronósticos puntuales tanto para la gasolina corriente como para la gasolina extra, y poder, por ejemplo, generar rangos de mínimo y máximo ingreso potencial. Asimismo, se puede estudiar cómo reacciona la demanda de gasolina (corriente o extra) ante cambios en la tasa de crecimiento del PIB, introducción de carros eléctricos e híbridos al país, cambios en el precio del gas vehicular, entre otros, y por ende se puede analizar cómo dichos cambios podrían afectar la recaudación esperada para financiar el proyecto PLMB.

## **Agradecimientos**

Esta investigación fue realizada por el Centro de Excelencia y Apropiación en Big Data y Data Analytics (CAOBA), con el apoyo del Ministerio de Tecnologías de la Información y Telecomunicaciones de la República de Colombia (MinTIC) a través del Departamento Administrativo Colombiano de Ciencia, Tecnología e Innovación (COLCIENCIAS) dentro del contrato N°. FP44842-anex46-2015. Queremos hacer un reconocimiento especial a todos los miembros de la oficina de telecomunicaciones del Departamento Nacional de Planeación (DNP) y al equipo CAOBA-Icesi, tanto profesores como estudiantes.

## Referencias

- Alonso, J.C., & Rivera, A.F. (2017). Pronosticando la inflación mensual en Colombia un paso hacia delante: una aproximación «de abajo hacia arriba». *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 23, 98-118. Recuperado de: <https://www.upo.es/revistas/index.php/RevMetCuant/article/view/2688>
- Alonso, J.C. & Solano, J.A. (2006). Sistema de transporte masivo en la ciudad de Cali: ¿Podrá el municipio financiar el proyecto? *Estudios Gerenciales*, 22(98), 111-129.
- Alonso, J.C., Solano, J.A., & Duque, X. (2007). El futuro de la sobretasa a la gasolina y la financiación del Sistema Integrado de Transporte Masivo. *Boletín Polis: Observatorio de Políticas Públicas*, 2(2), 9-12.
- Hyndman, R.J., Lee, A.J., & Wang, E. (2016). Fast computation of reconciled forecasts for hierarchical and grouped time series. *Computational Statistics and Data Analysis*, 97, 16-32. <http://doi.org/10.1016/j.csda.2015.11.007>
- Hyndman, R. J., Ahmed, R.A., Athanasopoulos, G., & Shang, H.L. (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis*, 55(9), 2579-2589. <http://doi.org/10.1016/j.csda.2011.03.006>
- Wickramasuriya, S.L., Athanasopoulos, G., & Hyndman, R.J. (2015). *Forecasting hierarchical and grouped time series through trace minimization* (Nº. 15). Recuperado de: <https://business.monash.edu/econometrics-and-business-statistics/research/publications/ebs/wp15-15.pdf>
- Yepes, T., Martínez, S., & Aguilar, J. (2014). *Financiación de la infraestructura en los municipios del sistema de ciudades de Colombia*. Documento para la Misión de Ciudades. Bogotá: Fedesarrollo. Recuperado de: <https://repository.fedesarrollo.org.co/handle/11445/1117>