



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE POSGRADO EN INGENIERÍA

Pronósticos Dr. Jair Morales Camarena

Elaborado por: Sergio Ibarra Ramírez
CDMX, Mayo de 2023



“Modelos de Pronóstico aplicados a Demanda de Gas Natural”

Objetivo

Demostrar la aplicación de distintos modelos de pronóstico para el caso de la demanda de Gas Natural en el sector eléctrico mexicano.

Introducción

Se cuentan con datos mensuales desde Enero de 2005 hasta Septiembre de 2022 de la demanda de gas natural que se usa para producir energía eléctrica en México. Históricamente la Secretaría de Energía (SENER) había hecho hasta 2018 un “Informe Anual sobre gas natural”; en donde prospectaba consumos a 10-12 años con un error MAPE promedio de 20-25% ¹. A partir de 2019 no se tiene un pronóstico oficial de la demanda de dicho hidrocarburo en el país.

Marco Teórico

La teoría de pronósticos busca determinar el valor más probable de una variable dependiente (Y) en función de variables independientes (x1,x2, x3, etc.) . En este caso se abordarán dos casos principalmente ,

A) El modelo lineal generalizado (GLM, por sus siglas en inglés) es aquel en el que el valor de la variable a pronosticar Y se calcula como una combinación lineal de las variables predictoras x1, x2,...xn. Dicha combinación lineal puede ser “transformada” mediante una función denominada función link que permitirá usar el concepto de GLM para explicar relaciones lineales no simples (eg. Binomiales, Poisson). Se presenta la ecuación la ecuación general de los GML así como las funciones link que se usaría en ciertos casos particulares:

$$y(\eta) = \beta_0 + \eta \beta_1 * x_1 + \eta \beta_2 * x_2 + ... + \eta \beta_n * x_n$$

Donde:

- y: Valor a pronosticar de variable dependiente y
- β : Parámetros a ajustar para cada caso
- Xi: variables predictoras independientes entre si
- η n o gama: Función Link (para el caso de funciones no simples)

Distribución de Probabilidad	Función Link
Distribución Normal	Identity function
Distribución Binomial	Logit/Sigmoid function
Distribución Poisson	Log function (aka log-linear)

B) Los modelos de series de tiempo donde se pretende conocer el valor de la variable a pronosticar a tiempo t, como una combinación lineal de los valores de la misma variable y los errores a tiempos pasados (t-s) más un término de error o “ruido blanco” que da cuenta por la propia aleatoriedad del fenómeno. Un tipo de modelos muy usados en series de tiempo son los modelos llamados Auto-Regresivos y de Medias Móviles (ARIMA) que tienen la siguiente ecuación general:

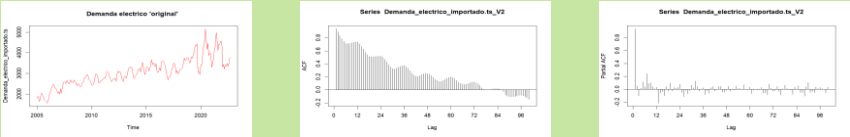
$$Y_t = \varphi_1 Y_{t-1} + \dots + \varphi_p Y_{t-p} + at + \theta_1 at-1 + \dots + \theta_q at-q$$

Donde

- yt: Valor a pronosticar de variable y a tiempo t
- θ, φ : Parámetros a ajustar para cada caso
- yt-s: Valores de la variable y a t-s
- $at - s$: Errores de la variable y a t-s

Aplicación

Se muestra la serie histórica de demanda de gas natural en el sector eléctrico en México y las funciones ACF y PACF de la serie sin diferenciar



Se aplico el GLM para la predicción de 12 meses de demanda de gas natural en el sector eléctrico mexicano como función de la Población en México, así como del PIB. (Utilizando datos de 201 meses como ‘datos de entrenamiento’)

Modelo	MAPE	Comentario(s)
GLM (Demanda de GN como función de población & PIB)	15.5%	Se usó la función Gaussian (identity) para GLM “lineal simple”
GLM (Demanda de GN como función únicamente de población)	15.0%	

De igual manera se llevó a cabo un pronóstico de los mismos 12 meses usando los modelos de series de tiempo de ARIMA (Utilizando datos de 201 meses como ‘datos de entrenamiento’) con base en los resultados de las gráficas de ACF & PACF

	AR(p)	MA(q)	ARMA (p,q)
ACF	Decaimiento geométrico	Valor significativo en ciertos valores de lag p y posterior corte después de lag p	Decaimiento geométrico
PACF	Valor significativo en ciertos valores de lag p y posterior corte después de lag p	Decaimiento geométrico	Decaimiento geométrico

Modelo	MAPE	Comentario(s)
ARIMA (1,1,1)	18.5%	La serie debe ser diferenciada una vez con el objetivo de hacerla estacionaria
SARIMA (1,1,1) (1,1,0)[12]	16.5%	Se considera una estacionalidad de que se repite cada 12 meses

Se llevaron a cabo transformaciones de la variable dependiente (log, raíz cuadrada) y/o Suavizamiento de los datos históricos de la serie (Simple Moving Average (SMA), etc) o incluso se “delimitó” el conjunto de datos de entrenamiento evitando considerar los datos atípicos de los años 2020 y 2021 provocados por la pandemia de Covid 19 con el objetivo de lograr mejores valores de predicción y evitar que los valores atípicos afecten tan significativamente a los modelos.

Modelo	MAPE	Comentario(s)
ARIMA (1,1,1) de LOG	17.5%	Se transforma la demanda a log(Demanda)
ARIMA (1,1,1) de Serie previamente ajustada con (SMA)	17.8%	Se ajusta un modelo para el caso de la serie previamente suavizada
SARIMA (1,1,1) (1,1,0)[12] (Con datos de train recortados a 2019)	12.4%	Se ajusta un modelo para el caso de la serie donde los datos de entrenamiento se recortan a solo 2015-2019 (recortando valores atípicos provocados por covid-19)

Conclusiones

Los modelos lineales Generalizados (GLM) ofrecen una buena alternativa de pronóstico de una variable como la demanda de gas en función de otros datos tales como el precio de importación del hidrocarburo, el PIB, etc, sin embargo, el pronóstico final resulta no ser del todo preciso y está sujeto a la estimación de las variables predictoras que son en si otros pronósticos.

Los modelos de series de tiempos están basados en la relación de la variable con sus propios valores en tiempos pasados, dicha relación se cuantifica mediante las funciones de autocorrelación y de autocorrelación parcial (ACF & PACFF, por sus siglas en inglés). Estos modelos suelen ser muy sensibles a valores atípicos de la serie, lo que suele resultar en pronósticos sesgados.

Algunas opciones para tratar datos con *outliers* tan marcados como los provocados por el covid 19 son:

- A. Intentar suavizar la serie original (SMA, WMA, etc)
- B. Intentar sustituir los *outliers* por valores como promedios
- C. Aislar los *outliers* para el entrenamiento del modelo

El mejor resultado (menor error MAPE) se dio con el modelo SARIMA (1,1,1) (1,1,0)[12] (Con datos de entrnemaiento recortados a 2019), es decir, que “no se tomaron en cuenta los resultados atípicos derivados de la pandemia de Covid-19”. Dicho modelo permitió obtener forecast de 12 meses con un MAPE de 12% que resulta mejor al reportado históricamente por la SENER hasta el 2018.

Bibliografía

- SENER | Sistema de Información Energética | Demanda Interna de Gas Natural por Estado Sector Eléctrico (energia.gob.mx)
- S. Ali, “Reading the ACF and PACF Plots — The Missing Manual / Cheatsheet”. linkedin.com. <https://www.linkedin.com/pulse/reading-acf-pacf-plots-missing-manual-cheatsheet-saqib-ali/>
- “Arauto”, “How to choose the parameters for the model”. arauto.readthedocs.io. https://arauto.readthedocs.io/en/latest/how_to_choose_terms.html