2. NOTA METODOLÓGICA PARA LA ELABORACIÓN DE LAS PROYECCIONES DE EGRESOS

Formato 7b) Proyecciones de Egresos – LDF

Metodología

Para realizar las proyecciones de egresos en el periodo 2020-2025, se recurrió a la metodología del modelo de series de tiempo propuesto por Box & Jenkins, adicionado de variables exógenas, mejor conocido como modelo ARMAX.

Uno de los objetivos de este modelo es simular los valores y pronósticos de la serie temporal analizada. A diferencia del tradicional modelo ARIMA (Autorregresivo Integrado de Medias Móviles) el cual corresponde a un modelo de ecuaciones univariadas sin variables exógenas y en el que "los datos hablan por sí mismos", el modelo ARMAX se trata de un modelo de ecuaciones multivariadas, donde la incorporación de variables exógenas permite que sean los términos de error de la ecuación los que asuman el proceso estocástico de la serie en cuestión; esto es:

1) Sea la ecuación objetivo:

$$Cap\ i000_t = \alpha + \sum \beta_k * X\ k_{t-s} + e_t$$

Donde:

Cap. $i000_t$
- 1000 si es Servicios Personales en el periodo t.
- 2000 si es Materiales y Suministros en el periodo t.
- 3000 si es Servicios Generales en el periodo t.
- 4000 si es Transferencias, Asignaciones, Subsidios y Otras Ayudas en el periodo t.
- 5000 si es Bienes Muebles, Inmuebles e Intangibles en el periodo t.
- 6000 si es Inversión Pública en el periodo t.
- 7000 si es Inversiones Financieras y otras Provisiones en el periodo t.
- 8000 si es Participaciones y Aportaciones en el periodo t.
- 9000 si es Deuda Pública en el periodo t.

- X_{kt-s} es un vector de variables exógenas incluidas en el modelo en los periodos t-s, donde s ≥ 0;

X_{kt-s}
- Índice Nacional de Precios al Consumidor en el periodo t-s.
- Producto Interno Bruto Nacional en el periodo t-s.
- Índice Trimestral de la Actividad Económica Estatal en el periodo t-s.
- Tasa de Interés Interbancaria de Equilibrio a 28 días en el periodo t-s.
- Índice de Inversión Fija Bruta en el periodo t-s.

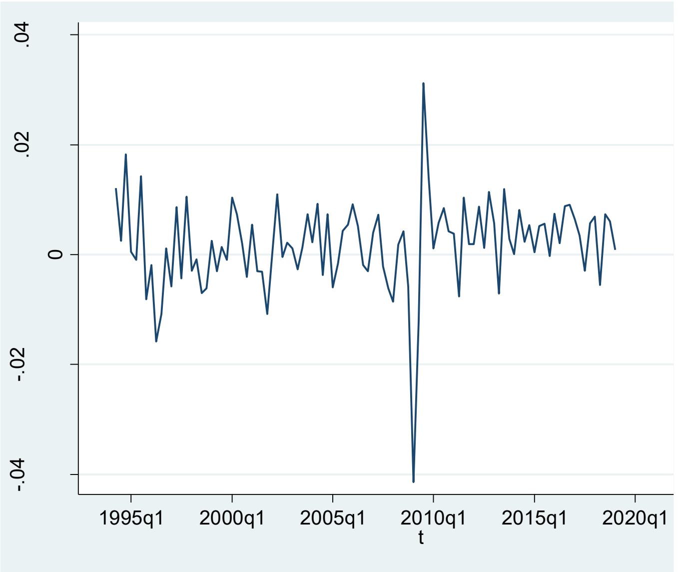
- β_k es un parámetro que pondera el efecto de la variable $X_{k.}$;
- α es una constante que explica el valor de la serie, independientemente de los efectos de $X_{kt-s.}$; y
- e_t es el término de error de la ecuación, al cual se le asignará el patrón ARMA del modelo.

2) Se obtienen los términos de error como:

$$e_t = Cap\ i000_t - \alpha - \sum_{k=1} \beta_k * X_{kt-s}$$

Se corrobora que los términos de error arrojados por la ecuación objetivo sea una serie estacionaria, condición necesaria para llevar a cabo el modelo. Este procedimiento se realiza graficando la serie (Gráfico 1), observando el patrón que sigue en el tiempo.

Gráfico 1): Términos de Error de la Ecuación Objetivo Estimada para el PIB



3) Se procede a corroborar con la prueba Dickey-Fuller Aumentada (Prueba de Raíz Unitaria):

Un proceso es estacionario si los ponderadores de rezago tienen un valor menor a uno:
Partiendo de la ecuación de un proceso Autorregresivo de primer orden $\boldsymbol{e_t = \mu e_{t-1} + u_t}$. y restando de ambos lados e_{t-1}, tenemos:

$$\Delta e_t = (\mu-1)e_{t-1} + ut$$

Si $\boldsymbol{|\mu-1|<1}$ entonces e_t es un proceso estacionario en el tiempo; es decir, con media y varianza constante.
Las hipótesis son las siguientes:

H_0: $|\mu| = 1$ $\longrightarrow$ No se rechaza la hipótesis nula. (No es estacionario)
H_1: $|\mu| < 1$ $\longrightarrow$ Se rechaza la hipótesis nula. (Es estacionario)

4) Se obtienen las gráficas de Autocorrelación Parcial (PACF) y Autocorrelación (ACF) para determinar los términos Autorregresivos (AR) y de Medias Móviles (MA) respectivamente, mismas que se presentan en el gráfico 2 y se obtienen mediante las siguientes funciones:

$$PACF = corr\,(e_t,\, e_{t-k} \mid e_{t-1}\, e_{t-2} \ldots e_{t-k+1}) \qquad ACF = corr(e_t, e_{t-k})$$

$$\frac{Cov(e_t-\hat{e}_t, e_{t-k}-\hat{e}_{t-k})}{\sqrt{V(et-\hat{e}t)}\quad\sqrt{V(e_{t-k}-\hat{e}_{t-k})}} \qquad \frac{Cov\,(e_t,\, e_{t-k})}{\sqrt{Var(e_t)} * \sqrt{Var(e_{t-k})}}$$

Gráfico 2a): Función de Autocorrelación

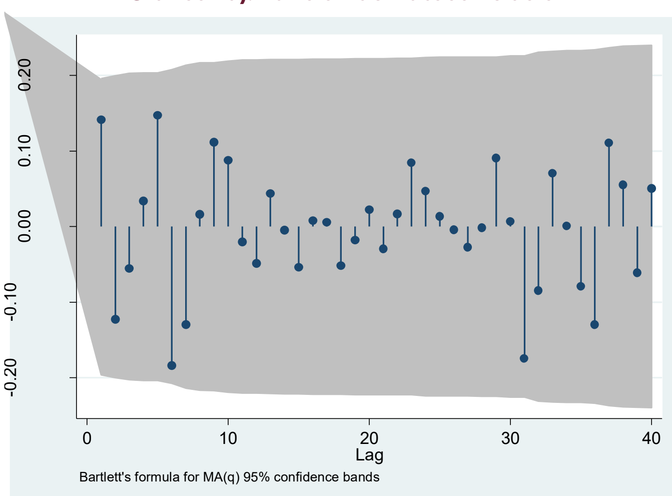
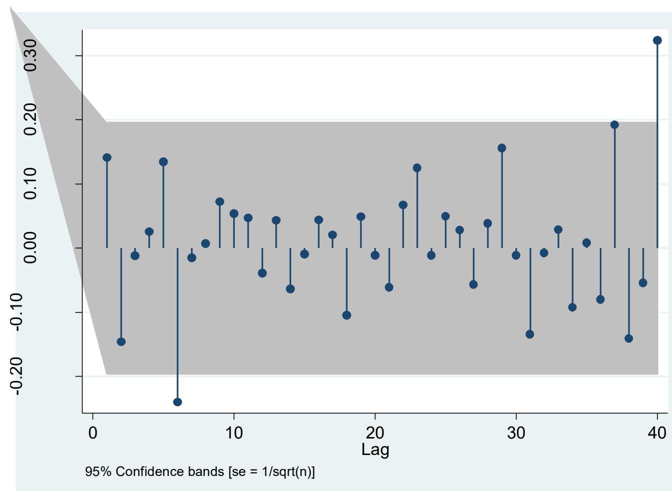


Gráfico 2b): Función de Autocorrelación Parcial



De las gráficas anteriores, se deduce que la especificación ARMA de los términos de error solo seguiría un término autorregresivo en el sexto rezago, mientras que ninguna de las otras observaciones sobresale de la banda de 95.0 por ciento de confianza para confirmar su significancia estadística.

5) Se aplican los criterios de Akaike (AIC) y de Schwartz (SC), donde el valor mínimo determina el número de rezagos óptimos en los términos Autorregresivos a incluir, de acuerdo a las siguientes funciones:

$$\text{Criterio de Akaike: } AIC = \frac{Ln(SSE)}{T} + \frac{2k}{T} \qquad \text{Criterio de Schwartz: } SC = \frac{Ln(SSE)}{T} + \frac{k^{*}Ln(T)}{T}$$

Donde:

SSE: Suma de los términos de error de la ecuación elevados al cuadrado.

K: Número de variables, incluyendo el intercepto.

T: Tamaño de la muestra.

6) Una vez determinados los componentes AR y MA del modelo, se procede a estimar los parámetros y a pronosticar los términos de error en los periodos *t+1, t+2, ..., t+s*.

Un ejemplo de pronóstico para un modelo ARMA (1,1); es decir, con un rezago en la variable dependiente (el término de error) y con una media móvil, es el siguiente:

$$e_{t+1} = \mu + \alpha_1 * e_t + \alpha_2 * u_t + u_{t+1}$$

donde u_{t+1} es el componente no determinista o de error en la ecuación, cuyo valor esperado es cero y se asume ser ruido blanco[1].

Por lo tanto:

$$e_{t+1} = \mu + \alpha_1 * e_t + \alpha_2 * u_t$$

$$e_{t+2} = \mu + \alpha_1 * e_{t+1} + \alpha_2 * u_{t+1} + u_{t+2}$$

$$e_{t+2} = \mu + \alpha_1 * (\mu + \alpha_1 * e_t + \alpha_2 * u_t) + \alpha_2 * u_{t+1} + u_{t+2}$$

$$e_{t+2} = \mu(1+\alpha_1) + \alpha_1^2 e_t + \alpha_1\alpha_2 * u_t$$

7) Ya que se obtienen los pronósticos de los términos de error en los *t+s* periodos adelante, estos se incluyen a la regresión objetivo, como a continuación se especifica:

$$Cap.\ i000_{t+1} = \alpha + \sum_{k=1}^{K} \beta_k * X_{kt+p-s} + E_{t+1}$$

$$Cap.\ i000_{t+2} = \alpha + \sum_{k=1}^{K} \beta_k * X_{kt+p-s} + E_{t+2}$$

$$Cap.\ i000_{t+p} = \alpha + \sum_{k=1}^{K} \beta_k * X_{kt+p-s} + E_{t+p}$$

[1] El término ruido blanco o “White noise” como se conoce en la literatura, hace referencia a que los términos de error se distribuyen con normalidad y no están correlacionados en el tiempo.

Donde E_{t+p} son los términos de error estimados con la especificación ARMA.

Es fundamental señalar que, para realizar los pronósticos de los *t+s* periodos adelante de los capítulos de gasto, se necesita información disponible de las variables exógenas a emplear en *t+s* periodos adelante.

Gráficamente, la proyección de una serie temporal se observa de la siguiente manera:[2] del lado izquierdo se ofrece una muestra del pronóstico del Índice Trimestral de la Actividad Económica del Estado de Puebla y del lado derecho la proyección del Capítulo 3000 No Etiquetado.

Gráfico 3a): Simulación de los valores del ITAEE De 1993 a 2025

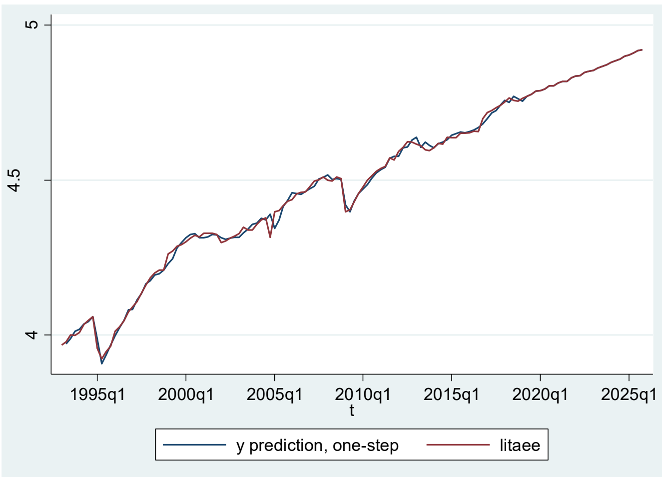
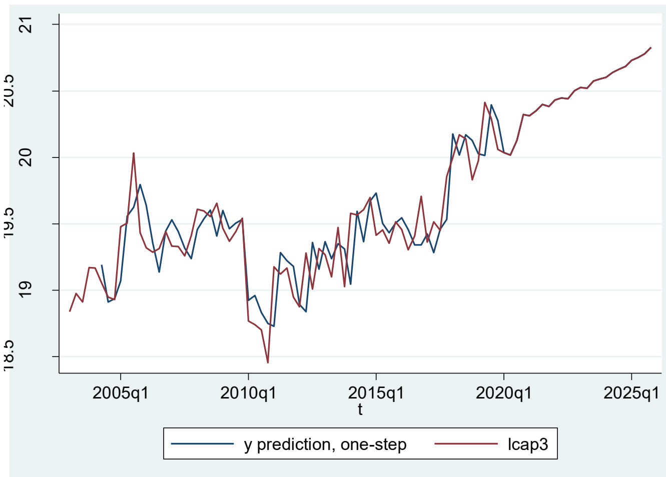


Gráfico 3b): Simulación de los valores del Capítulo 3000 No Etiquetado De 2003 a 2025



Consideraciones

1) Dado que la combinación lineal de la variable dependiente y las variables exógenas deben arrojar una serie estacionaria de los términos de error, cuando esto no sucede es necesario recurrir a diferenciar las series y relacionarlas. Esto lleva el nombre de orden de integración de las series.

   Así, si la serie del $Cap.\ i000_t$ es no estacionaria, se procede a tomar su primera diferencia para lograr el comportamiento estable de la variable en el tiempo; esto es[3]:

$$\Delta\ Cap.\ i000_t = Cap.\ i000_t - Cap.\ i000_{t-1}$$

   Y su segunda diferencia se denota como

[2] Las proyecciones de egresos se realizaron con el software estadístico STATA 14.

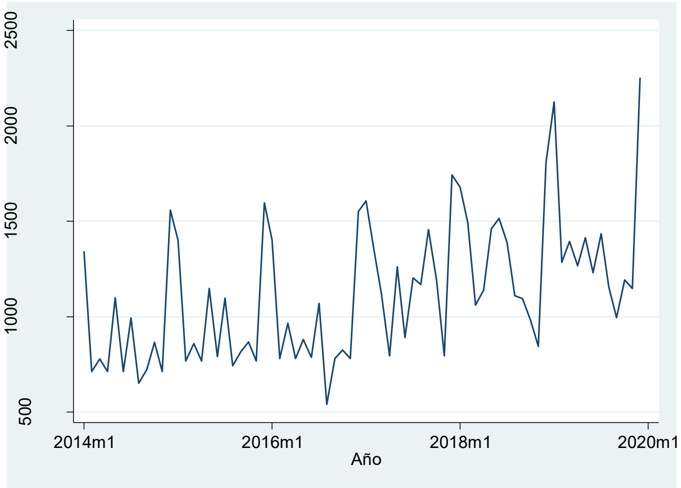
[3] Para lograr la estacionariedad en las series requeridas, no hubo que diferenciar más que una vez cada variable.

$$\Delta_2\ Cap.\ i000_t = \Delta Cap.\ i000_t - \Delta Cap.\ i000_{t-1}$$

2) Para llevar a cabo el modelo realizado, se dio un tratamiento previo a algunas series, con el fin de remover la estacionalidad de estas.

La estacionalidad se refiere a los comportamientos similares que ocurren en la serie con el paso del tiempo; es decir, fluctuaciones que se repiten en cada mismo periodo para cada año. Gráficamente, una serie de tiempo con componente estacional se observa de la siguiente manera:

Gráfico 4): Capítulo 1000 No Etiquetado con efecto estacional.
De 2014 a 2019
(Millones de pesos)



Es ampliamente recomendable trabajar con cifras desestacionalizadas. De lo contrario, es posible que los resultados de los coeficientes asociados a cada ecuación puedan resultar ser no significativos cuando en realidad si lo son, además de mejorar el análisis de la variable.

Para obtener las series ajustadas por estacionalidad se aplicó el método *Census X-12*, mejor conocido como *X-12 ARIMA*, desarrollado por la Oficina del Censo de los Estados Unidos.

Dicho método toma en consideración la frecuencia de cada día en un mes (lunes, martes, miércoles, etc.), ya que estos pueden influenciar de manera significativa las variables en cuestión.

La ponderación de los días por mes se denota de la siguiente manera:

$$M_t = \frac{N_t}{N^*_t} \; exp \sum_{t=1}^{6} \beta_i (D_{it} - D_{7t})$$

Donde:

M_t es un factor de ajuste por la frecuencia de los días de la semana
i = 1 corresponde a lunes, i = 2 corresponde a martes, …
D_{it} es el número de días j en el mes t
N_t es el número de días del mes t
N^*_t es el promedio, sobre cuatro años, de la longitud del mes (30, 31 ó 28.25)
β_i son los parámetros de las ponderaciones diarias.

Este método también toma en consideración el ajuste que se le debe dar a los efectos de la semana santa, ya que esta festividad no se da en el mismo mes para cada año. Dicho factor de ajuste se obtiene de la siguiente manera:

$$E_{ij} = \hat{a} + bX_{ij}\,(w)$$

Donde:

E_{ij} es el factor de ajuste por la semana santa
a y b son parámetros estimados por la regresión
w es el número de días previos a la semana santa
w_{ij} es el número de días de los w días antes de la semana santa que caen en el periodo j
$X_{ij}(w) = w_{ij} / w - Z_j(w)$
$Z_j(w)$ es el promedio w_{ij} / w

Una vez determinados los efectos de calendario y de semana santa, se procede a estimar la serie mediante un modelo ARIMA que ajuste de manera correcta a la serie.

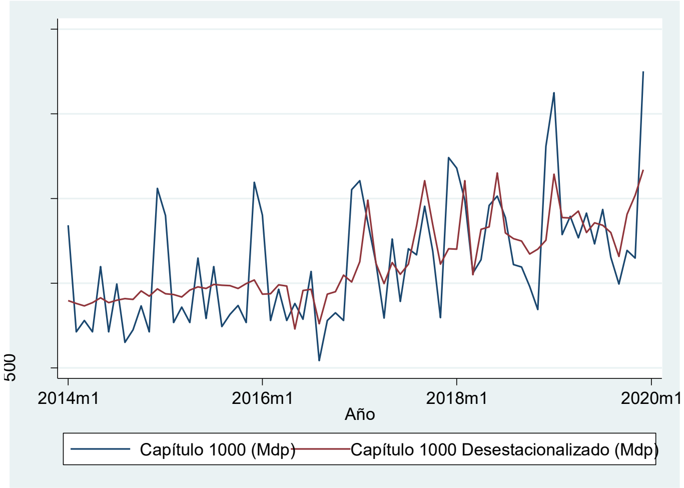
Dado que una serie de tiempo se descompone en sus factores cíclico-tendencia, de estacionalidad y de su componente irregular, para obtener una serie ajustada en su nivel se descontará el factor estacional.

Serie Original = Tendencia-ciclo + Estacional + Irregular

Serie desestacionalizada = Serie Original – Estacional

De este modo, se obtiene la serie desestacionalizada, que gráficamente se observa de la siguiente manera en un ejemplo para el Capítulo 1000 No Etiquetado:

Gráfico 5): Capítulo 1000 No Etiquetado con efecto estacional y desestacionalizado de 2014 a 2019
(Millones de pesos)



Base de datos

Para realizar las proyecciones, se dispuso de información histórica mensual del año 2003 al 2019 de los capítulos de gasto en su momento contable aprobado, distinguiendo entre gasto etiquetado y no etiquetado. Dicha información fue transformada para obtener una base de datos con frecuencia trimestral, arrojando de esta manera 68 observaciones por cada capítulo; es decir, una base de datos con un total de 1224 observaciones.

Por su parte, también se contó con una base de datos trimestral del año 1993 al segundo trimestre de 2019 para las variables referentes al Índice Nacional de Precios al Consumidor, el Producto Interno Bruto, el Índice Trimestral de la Actividad Económica Estatal y el Índice de Inversión Fija Bruta, información obtenida del Banco de Información Económica del Instituto Nacional de Estadística y Geografía (INEGI). Además, se obtuvieron observaciones de 1995 al segundo trimestre del presente de la Tasa de Interés Interbancaria de Equilibrio a 28 días, obtenidas del Sistema de Información Económica del Banco de México. Con ello, se obtuvo un total de una base de datos con 522 observaciones.

Bibliografía

- Ortiz de Dios, C. (2012). Modelos Econométricos y de Redes Neuronales para predecir la Oferta Maderera en México: ARIMA vs NAR y ARMAX vs NARX (Tesis de Maestría). Universidad Autónoma Metropolitana, Cd. de México.
- Philip H. Franses. (1991). Primary Demand for Beer in the Netherlands: An Application of ARMAX Model Specification. 09-07-2019, de SAGE Publications, Inc. Sitio web: https://www.jstor.org/stable/3172813
- Stata Press. (2013). Stata Times-Series Reference Manual Release 13. Texas: Stata Press.

- INEGI. (2013). Metodología del ajuste estacional. 12/07/2019, de Instituto Nacional de Estadística y Geografía Sitio web: http://internet.contenidos.inegi.org.mx/contenidos/Productos/prod_serv/contenidos/espanol/bvinegi/productos/metodologias/ajus_estacional/Metodajustestacional.pdf