

CENTRO DE INVESTIGACIÓN Y DOCENCIA ECONÓMICAS, A.C.



MÉTODOS INTENSIVOS PARA EL DESARROLLO DE PRONÓSTICOS DE SERIES DE
TIEMPO: AGREGADOS MACROECONÓMICOS EN MÉXICO (2018-2019)

TESINA

QUE PARA OBTENER EL TÍTULO DE

LICENCIADO EN ECONOMÍA

PRESENTA

ÁNGEL EDUARDO CASTILLO MARROQUÍN

DIRECTOR DE LA TESINA: DR. DANIEL VENTOSA-SANTAUlàRIA

CIUDAD DE MÉXICO

AGOSTO, 2019

Agradecimientos

A mis padres, quienes me dieron todo.

A mi asesor, el Dr. Daniel Ventosa-Santaulària y mis lectores, la Dra. Eva Arceo Gómez y el Dr. Alejandro López-Feldman, quienes me instruyeron extensamente en la econometría.

A mis amigos: Chava, AB, Santi, Oli y Plauchu (los badboys), por sobrevivir conmigo con el mínimo esfuerzo.

Índice General

Capítulo 1: Motivación y Revisión de Literatura	1
1.1 El procedimiento de <i>pooling</i> en el pronóstico de series de tiempo	2
1.2 La interpretación del problema de predicción y los métodos no lineales	3
Capítulo 2: Metodología	5
2.1 Simulaciones ‘dentro de muestra’ y portafolio de modelos	5
2.2 Criterios de desempeño de pronósticos	6
2.3 Construcción de pronósticos: procedimientos de <i>pooling</i> y selección	7
2.4 Estabilidad y pruebas de sesgo secuencial	10
2.5 Selección de método de construcción de pronóstico	11
2.6 Descripción de los modelos	11
2.6.1 ARIMA	11
2.6.2 Algoritmo Holt-Winters	13
2.6.3 Método de Suavizado Exponencial de Espacio de Estado	14
2.6.4 Modelo Theta	16
2.6.5 Modelo de Descomposición de Tendencia Estacional	16
2.6.6 TBATS: Suavizado Exponencial de Espacio de Estado con Transformación Box-Cox, Errores ARMA y Componentes de Tendencia y Estacionalidad	17
2.6.7 Redes Neuronales con <i>Algoritmo Extreme Learning Machine</i>	19
2.6.8 Red Neuronal Paralimentaria con una sola Capa Oculta y Entradas Rezagadas	20
Capítulo 3: Descripción de las Series	21
3.1 Series Mensuales	21
3.1 Series Trimestrales	23

Capítulo 4: Resultados	25
4.1 Desempeño de los Modelos en Estabilidad Secuencial y Discusión	25
Capítulo 5: Conclusiones	32
Bibliografía	33

Índice de Figuras

Tabla 1. Taxonomía de modelos de suavizado exponencial.....	15
Figura 1. Estructura de una Red Neuronal Paralela de una sola Capa Oculta.....	19
Gráfica 1. Serie histórica: IGAE, INEGI.....	21
Gráfica 2. Serie histórica: Inflación General, INEGI.....	22
Gráfica 3. Serie histórica: Inflación Subyacente, INEGI.....	22
Gráfica 4. Serie histórica: Tipo de Cambio, FRED.....	23
Gráfica 5. Serie histórica: PIB, INEGI.....	24
Gráfica 6. Serie histórica: Tasa de Desempleo, OECD.....	24
Gráfica 7. Histograma de Estabilidad Secuencial para el IGAE.....	26
Gráfica 8. Histograma de Estabilidad Secuencial para la Inflación General.....	27
Gráfica 9. Histograma de Estabilidad Secuencial para la Inflación Subyacente.....	27
Gráfica 10. Histograma de Estabilidad Secuencial para Tipo de Cambio.....	28
Gráfica 11. Histograma de Estabilidad Secuencial para PIB.....	29
Gráfica 12. Histograma de Estabilidad Secuencial para Tasa de Desempleo.....	30
Tabla 2. Error msMAPE dentro de muestra, promedio y para cada serie, de cada modelo ...	31

Abstract

La selección de modelos de series de tiempo por medio de criterios estadísticos supone algunas desventajas en la práctica. En este trabajo se propone un conjunto de métodos intensivos para estimar pronósticos de series univariadas, que involucra procedimientos de *pooling* a través de algoritmos de selección de diferentes modelos y métodos, y un criterio de estabilidad secuencial. Dicho criterio muestra el desempeño de los diferentes métodos -lineales y no lineales- a lo largo de la muestra, lo que en casos de inestabilidad justifica el *pooling*, además de que permite analizar el comportamiento de los procesos estocásticos.

Capítulo 1: Motivación y Revisión de Literatura

1.1 El procedimiento de *pooling* en el pronóstico de series de tiempo

Una parte importante de los trabajos existentes que tienen como estrategia empírica la aplicación intensiva (i.e. que utiliza de manera aditiva una serie de procesos) de métodos para la estimación de pronósticos de series de tiempo, suele tener como propósito evaluar el desempeño de los modelos per se. Una razón es que la metodología de los estudios con aplicaciones de análisis de series de tiempo suele centrarse en estudiar el comportamiento de una serie con base en una hipótesis intrínseca a ésta. En este sentido, se suele seleccionar un modelo, es decir que se adopta la interpretación matemática particular de un proceso estocástico, con base en el análisis de la serie en la hipótesis además de un criterio estadístico, regularmente el Criterio de Información de Akaike (AIC) o el Criterio de Información Bayesiano (BIC).

Acerca de la literatura existente sobre la aplicación intensiva de modelos de pronósticos para series de tiempo, existen dos trabajos principales, cuya metodología es idónea para el propósito de documentar los hallazgos y el conocimiento pertinente sobre este nicho de análisis. El primero es un artículo de Stock y Watson (1998). El segundo es un trabajo más reciente de Zou y Yang (2004). Stock y Watson (1998) se enfocaron en estudiar el desempeño de un rango extenso de métodos de pronósticos para 215 series macroeconómicas de Estados Unidos. Este tipo de diseño experimental trabaja con series acotadas para simular pronósticos en tiempo real de un portafolio extenso de pronósticos de series univariadas mensuales en horizontes de uno, seis y doce periodos, y así compararlos.

Zou y Yang (2004) defienden que la selección tradicional de modelos conlleva a un riesgo por inestabilidad, sobre todo debido a las imprecisiones relacionadas con la selección a través de pruebas estadísticas. Mencionan, además, que utilizar estimaciones de un solo modelo puede significar mayor variabilidad en los pronósticos, ya que “un modelo no necesariamente se desempeña mejor que otro en términos de riesgo de predicción”. Esta percepción conlleva a la

constante más notable en ambos estudios: el diseño de un método o algoritmo para combinar la información de más de un pronóstico.

En la literatura previa se habían utilizado los procedimientos de *pooling* a través de promedios simples de los modelos y se demostró que mejoraban el desempeño de los pronósticos. Una de las principales aportaciones de Stock y Watson (1998) fue el diseño de un proceso de *pooling* que empleaba el error cuadrático medio de los pronósticos de las simulaciones comparados con los datos reales. De esta forma se generaba un promedio ponderado mediante la proporción inversa de dichos errores. La conclusión más relevante de Stock y Watson (1998) es: a pesar de que las especificaciones más eficientes son las de modelos autorregresivos previamente sometidos a pruebas de raíz unitaria, estos pronósticos se pueden mejorar a través combinarlos con los estimados por otros métodos. Amisano y Geweke (2013), utilizaron un procedimiento de *pooling* con la media simple de tres modelos bayesianos para pronósticos de agregados macroeconómicos y midieron una mejora en la eficiencia promedio de 41.9%.

A partir de los años 80 ha sido una tendencia utilizar procedimientos de *pooling*, a pesar de las críticas. Clement y Hendry (1998) a través de Zou y Yang (2004), mencionan que el método de combinar modelos econométricos basados en el mismo conjunto de información tiene poca justificación estadística. Sin embargo, si los resultados empíricos sugieren que, efectivamente, el *pooling* mejora los pronósticos, aportar un esquema alternativo para la ponderación (*weighting*) de los pronósticos, puede significar un avance en lo que se conoce sobre métodos intensivos para la realización de pronósticos de series univariadas, como defienden Zou y Yang (2004).

1.2 La interpretación del problema de predicción y los métodos no lineales

Bontempi et al. (2013) analizan el problema de producir pronósticos para horizontes largos, principalmente la acumulación de errores. En general, la manera en la que se formalizan los modelos no lineales, enfocándose en aquellos que utilizan algún método de aprendizaje automático (*machine learning*), suele proveer una justificación matemática de por qué estos pueden capturar mejor los procesos estocásticos en el comportamiento de las series que aquellos

no lineales: Los autores se basan en teoría de sistemas dinámicos para afirmar que el comportamiento “aparentemente aleatorio puede ser generado por sistemas determinísticos (...) interactuando de manera no lineal.”

La lógica utilizada en el diseño experimental de Stock y Watson (1998) y otros trabajos anteriores, específicamente sobre las simulaciones dentro de muestra, son precisamente el marco analítico de los modelos que definen los pronósticos como un problema de *supervised learning*. Bontempi et al. (2013) describen que una de las limitaciones en la interpretación desde *machine learning* del problema de predicción para horizontes largos es la sensibilidad a los errores de estimación en la estrategia recursiva. Sin embargo, para algunos modelos, como el de Redes Neuronales Recurrentes, los pronósticos son bastante eficientes.

En gran parte, la utilidad de los modelos no-lineales en los pronósticos de series de tiempo tiene que ver con problema del trade-off entre varianza y sesgo. Más allá del problema de predicción, esta propiedad es la que justifica los procesos de *pooling* y el uso de métodos de *machine learning* a la par de modelos lineales.

Capítulo 2: Metodología

2.1 Simulaciones ‘dentro de muestra’ y portafolio de modelos

En el contexto de los conjuntos de datos en este estudio, se define Y_t como la realización más reciente o la última observación de una serie histórica. Similarmente, Y_{t+1} es el valor consecuente de la serie que se plantea pronosticar, que a su vez está basado en la serie histórica mencionada $\{Y_i\}_{i=1}^t = Y_1, Y_2, \dots, Y_t$, donde $t \geq 1$, es el tamaño de la muestra y de igual forma denota el periodo de la última actualización de dicha serie.

El método propuesto para realizar las simulaciones en el marco analítico de este estudio es con simulaciones ‘dentro de muestra’, como se explica a continuación. Se parte de una serie de longitud t y se recorta un horizonte determinado H , de aquí en adelante ‘horizonte de *training*’. Dicho horizonte comprenderá regularmente 1, 6 o 12 realizaciones de la muestra; la razón es que éstas serán las longitudes estándar para el horizonte de predicción fuera de muestra, es decir, los pronósticos reales.

De una lista de modelos predeterminados, se selecciona el modelo j y se le introducen los datos del rango $\{Y_i\}_{i=1}^{t-H}$. Las estimaciones $\{\hat{Y}_{j,i}\}_{i=t-H+1}^t$ corresponden a los pronósticos del modelo j para el rango de realizaciones $\{Y_i\}_{i=t-H+1}^t$, que son las observaciones que se cortaron para realizar el *training* i.e. la evaluación del desempeño y la selección de los modelos para el pronóstico. De los modelos ‘primitivos’ se computa una especificación; estos son descritos a detalle en la sección 2.6 en este capítulo y se resumen en la siguiente lista:

- ARIMA¹
- Algoritmo Holt-Winters²
- Método de Suavizado Exponencial de Espacio de Estado³

¹ Se utilizó la función `forecast.auto.arima` en R.

² Librería `HoltWinters` en R.

³ Librería `ETS` en R.

- Modelo Theta⁴
- Modelo de Descomposición de Tendencia Estacional⁵
- Extreme Learning Machines (ELM) – Red Neuronal Artificial⁶
- Modelo TBATS⁷
- Red Neuronal Paralela – redes con capa oculta y entradas rezagadas⁸

2.2 Criterios de desempeño de pronósticos

Para un horizonte de *training* dentro de muestra ($H = 1, H = 6, H = 12$), se define un criterio de desempeño de pronóstico como el Error Porcentual Absoluto Medio (*MAPE*, en la literatura) del modelo j

$$MAPE_j = \frac{1}{H} 100 \sum_{i=1}^H \left| \frac{Y_{t+i} - \hat{Y}_{j,t+i}}{Y_{t+i}} \right|$$

que, en general, mide la precisión en términos porcentuales. Por esta razón es útil para medir el desempeño de diferentes técnicas para una misma serie. Hanke y Wichern (2014), mencionan que la medición MAPE es útil cuando el error con relación al tamaño de la serie es relevante para el pronóstico.

De igual forma, se puede utilizar el Error Porcentual Absoluto Medio Simétrico (*sMAPE*) que proponen Makridakis y Hibon (2000) en sus conclusiones sobre la Competencia M-3, en la cual se compilan diferentes mediciones para evaluar la precisión de diferentes técnicas de pronósticos.

$$sMAPE_j = \frac{1}{H} 100 \sum_{i=1}^H \frac{|Y_{t+i} - \hat{Y}_{j,t+i}|}{(|Y_{t+i}| + |\hat{Y}_{j,t+i}|)/2}$$

⁴ Librería *thetam* en R.

⁵ Se utilizó la función *hybrid* con configuración *stlm* en R.

⁶ Librería *elm* en R.

⁷ Librería *tbats* en R.

⁸ Librería *nnetar* en R.

Chen y Yang (2004) en un artículo dedicado a evaluar las medidas de desempeño de pronósticos, encuentran que el MAPE y el sMAPE tienen un desempeño muy deficiente cuando el valor real del pronóstico ronda por el cero, por lo que implementan una transformación en el denominador del sMAPE para corregir esta falla y lo llaman msMAPE (sMAPE modificado).

$$msMAPE_j = \frac{1}{H} 100 \sum_{i=1}^H \frac{|Y_{t+i} - \hat{Y}_{j,t+i}|}{\frac{|Y_{t+i}| + |\hat{Y}_{j,t+i}|}{2} + S_i}$$

en donde S_i se define como

$$S_i = \frac{1}{i-1} \sum_{k=1}^{i-1} |Y_k - \bar{Y}_{i-1}|$$

y, a su vez, \bar{Y}_{i-1} se obtiene de

$$\bar{Y}_{i-1} = \frac{1}{i-1} \sum_{k=1}^{i-1} Y_k$$

Chen y Yang (2004) encuentran que esta modificación mejora la eficiencia de la medida sMAPE y MAPE. Las técnicas descritas antes, que usan como base el MAPE, son escogidos como criterios de evaluación para diferentes modelos en una misma serie; aquí se interpretan como funciones de pérdida para los pronósticos, tomando en cuenta las correcciones para sus usos contingentes, principalmente el msMAPE en el caso en que los valores de los pronósticos son cercanos a cero.

2.3 Construcción de pronósticos: procedimientos de *pooling* y selección

- *Método 1: Modelo monolítico*

El método de pronóstico 1 selecciona un solo modelo utilizando el criterio msMAPE más bajo, es decir elige el modelo que despliegue menor error en los pronósticos dentro de muestra, de

acuerdo con el procedimiento descrito en 2.1 y 2.2. El pronóstico δ_j representa al modelo j y se selecciona:

$$\delta_j = \min_{\hat{Y}_{j,t+i}} \{msMAPE_j\}$$

- *Método 2: Combinación mediante algoritmo AFTER*

Este método explora el algoritmo AFTER (*Aggregated Forecast Through Exponential Reweighting*) para crear combinaciones de modelos propuesto por Yang (2001), el cual consiste en un esquema de ponderación basado en la varianza de los pronósticos.

La propuesta es reducir el error del pronóstico de una serie mediante la transformación de varios métodos a una combinación lineal convexa, atendiendo el problema de inestabilidad que describen Zou y Yang (2004) sobre la selección entre varios modelos; en este estudio, el método ‘monolítico’. Esta combinación se obtiene por medio del AFTER, que consiste en

$$W_{j,h} = \frac{\prod_{i=1}^{h-1} \hat{v}_{j,i}^{1/2} \exp\{-\frac{1}{2} \sum_{i=1}^{h-1} [(Y_{t+i} - \hat{Y}_{j,t+i})^2 / \hat{v}_{j,i}]\}}{\sum_{j'=1}^J \prod_{i=1}^{h-1} \hat{v}_{j',i}^{1/2} \exp\{-\frac{1}{2} \sum_{i=1}^{h-1} [(Y_{t+i} - \hat{Y}_{j',t+i})^2 / \hat{v}_{j',i}]\}}$$

como lo definen Zou y Yang (2004), donde $W_{j,h}$ es el peso del método j que corresponde en la combinación lineal convexa que conforma al pronóstico en la realización h del horizonte de *training* de longitud H ; es decir que $\sum_{j=1}^J W_{j,h} = 1$, para un número $J \geq 2$ de métodos de pronóstico y, a su vez,

$$\hat{Y}_{t+h} = \sum_{j=1}^J W_{j,h} \hat{Y}_{j,t+h}$$

es el pronóstico combinado a través de las ponderaciones de métodos de pronóstico para la realización $h \in [1, H]$, $H \geq 2$ del horizonte de training.

- *Método 3: Combinación convexa y proporción inversa del MAPE*⁹

El método 3 utiliza el mismo principio que el algoritmo AFTER en el sentido de la proporcionalidad de cada método penalizada por la varianza de predicción de éste i.e. el error del pronóstico con respecto a las realizaciones en la muestra. La diferencia es que el método 3 no pondera de manera exponencial dicha penalización, sino que utiliza uno de los criterios de desempeño basados en el MAPE (Ver 2.2). A manera de ejemplo, en el caso donde se utiliza el msMAPE, los pesos se calcularían de la siguiente manera:

$$W_{j,h} = \frac{(msMAPE_j)^{-1}}{\sum_{j'}^J (msMAPE_{j'})^{-1}}$$

Donde j' se refiere al resto de las estimaciones del msMAPE arrojadas por los otros modelos, de modo que el pronóstico con el método 3 se calcularía, similarmente:

$$\hat{Y}_{t+h} = \sum_{j=1}^J W_{j,h} \hat{Y}_{j,t+h}$$

⁹Este procedimiento tiene como propósito servir de *benchmark* para el método anterior.

2.4 Estabilidad y pruebas de sesgo secuencial

La premisa del trabajo de Zou y Yang (2004) es que la selección de modelos puede conllevar a inestabilidad cuando hay incertidumbre sobre cuál es el modelo más eficiente, lo que justifica el uso del procedimiento de *pooling*. En este sentido, los autores proponen una medida para la estabilidad de los criterios de selección de modelos, específicamente: una medida de consistencia secuencial que señala el porcentaje de veces que se selecciona una especificación definida de un modelo, utilizando el mismo proceso de selección, cuando cambia el tamaño de la muestra.

Si bien, los autores utilizan este coeficiente para medir la estabilidad de un método de selección de modelo e.g. AIC, la propuesta es 1) utilizarla para este fin en la selección de la especificación del modelo ARIMA que se usará después en los métodos de construcción de pronósticos y 2) extrapolar esta técnica a una prueba alternativa para justificar la estabilidad de las combinaciones, en más detalle:

- 1) Se propone acortar la muestra $\{Y_i\}_{i=1}^{t-H}$ para una determinada serie y especificar qué criterio estadístico (AIC, AICc o BIC) selecciona la especificación del modelo ARIMA con mayor eficiencia en los pronósticos dentro de muestra, para cada nivel de longitud de muestra. De esta forma, el criterio que arroje menores errores en las simulaciones se utilizará para ajustar la especificación ARIMA para las evaluaciones posteriores entre modelos.
- 2) La prueba alternativa que se propone consiste en observar la consistencia de las ponderaciones de los métodos de construcción de pronósticos 2 y 3, para cada nivel de longitud de muestra. La persistencia de algunos modelos puede caracterizar la estabilidad de estos y así condicionar qué modelos son relevantes para introducir en el algoritmo AFTER y en el método 3.

Estas dos pruebas se realizan forzosamente en este orden, dado que la selección de la especificación ARIMA es previa a la evaluación de todos los métodos de pronóstico i.e. cualquiera de las formas descritas en 2.3. Para ambas pruebas, se utiliza

$$\rho_j = \frac{\sum_{i=1}^{t-H} D_i}{t-H}$$

que determina el porcentaje de selección del modelo/criterio para cada nivel de tamaño de muestra, donde D_i es una variable dicotómica, de modo que

$$D_i = \begin{cases} 1, & j \text{ se selecciona} \\ 0, & j \text{ no se selecciona} \end{cases}$$

2.5 Selección de método de construcción de pronóstico

Tomando en cuenta el *tradeoff* entre error y estabilidad, se plantean dos casos. En el primero se escoge el ‘método monolítico’, mientras que en el segundo se utiliza el algoritmo AFTER para hacer una combinación de modelos.

Caso 1: Al comparar la función de pérdida de todos los modelos (MAPE, sMAPE o msMAPE), un método sobresale, por lo que éste arroja mejores pronósticos, y no se percibe consistencia secuencial de otros modelos para diferentes niveles de muestra, es decir, que en términos de medición de error no hay otros modelos que estimen pronósticos con errores relativamente bajos si se acorta la muestra.

Caso 2: Dos o más modelos arrojan pronósticos muy similares y, por lo tanto, muestran errores similares, u otros modelos generan pronósticos relativamente menos eficientes, pero que muestran estabilidad secuencial.

2.6 Descripción de los modelos

2.6.1 ARIMA

El Modelo Autorregresivo Integrado de Media Móvil (ARIMA), fue desarrollado originalmente por Box y Jenkins en 1976. Box et al. (2016), acerca de las series no estacionarias,

explican que “a pesar de que el nivel global de las fluctuaciones pueda ser diferente en diferentes tiempos, el comportamiento general de la serie, cuando se permiten diferencias dentro de cada nivel, puede ser similar a través del tiempo”. Esta explicación es útil para caracterizar el modelo ARIMA.

El operador rezago B es el operador lineal que transforma una realización Y_t en la anterior Y_{t-1} , de modo que:

$$B^n Y_t = Y_{t-n}$$

Se puede observar que al obtener la primera diferencia de una observación Y_t , ésta se puede representar en términos del operador rezago:

$$Y_t - Y_{t-1} = Y_t - B Y_t = (1 - B) Y_t = \Delta Y_t$$

donde ΔY_t es la notación para la primera diferencia. De forma general,

$$\Delta^d Y_t = (1 - B)^d Y_t$$

de modo que d se refiere al orden de diferencias; es el parámetro al que se refieren Box et al. (2016), cuando dicen “diferencias dentro de cada nivel”. La metodología Box-Jenkins para estimar modelos ARIMA requiere que la serie sea estacionaria. Por esta razón es que en adelante, pero en el mismo sentido, d se referirá al orden de diferencias que se necesitan para transformar una serie en estacionaria.

Dicho lo anterior, el proceso ARIMA de orden (p, d, q) está definido por:

$$Y_t = \mu + \sum_{i=1}^p \phi_i \Delta^d Y_{t-p} + \sum_{i=1}^q \theta_i \Delta^d e_{t-q}$$

donde p es el orden de rezagos, el número de observaciones anteriores incluidas en el modelo; q es el orden de media móvil, el número de valores rezagados del término de error definido

como e_t ; $\phi_1 \dots \phi_p$ son los parámetros del proceso autorregresivo (AR); y $\theta_1 \dots \theta_q$ los parámetros del proceso de medias móviles, del $ARIMA(p, d, q)$.

Los valores p y q se determinan mediante la función de autocorrelación (ACF) y la función de autocorrelación parcial (PACF) de la serie. La ACF genera los coeficientes de correlación entre una Y_t y sus observaciones previas. La PACF, similarmente, estima la correlación de Y_t con cada rezago, que no se explica por los rezagos anteriores, generalmente, la realización de la PACF para el rezago k muestra la correlación entre Y_t y Y_{t-k} que no se explica por $Y_{t-1} \dots Y_{t-(k-1)}$.

2.6.2 Algoritmo Holt-Winters

Esta extensión del modelo de Suavizado Exponencial para series univariadas surgió a partir de las aplicaciones prácticas de Holt (1957) y después fue mejorada por Winters (1960). Para desarrollarlo es necesario explicar el modelo de Suavizado Exponencial Simple, el cual describe el siguiente proceso:

$$Y_t = \mu_t + e_t$$

donde μ_t es la media y e_t es el error independiente con media cero y desviación estándar σ . La estimación de μ_t se hace a través de la media móvil ponderada de la realización anterior de la serie:

$$\mu_t = \alpha Y_t + (1 - \alpha)\mu_{t-1}$$

El parámetro α es el coeficiente de suavizado, la tasa a la que la memoria de las realizaciones anteriores se ‘extingue’ exponencialmente. El método Holt-Winters también es conocido como Suavizado Exponencial Triple, debido a que se integran los componentes de nivel (intercepto), tendencia (pendiente) y estacionalidad de la serie al suavizado:

$$l_t = \alpha Y_t + (1 - \alpha)(l_{t-1} + b_{t-1})$$

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1}$$

$$s_t = \gamma(Y_t - l_t) + (1 - \gamma)s_{t-p}$$

Los términos l_t , b_t y s_t se refieren al nivel, tendencia y estacionalidad, respectivamente. p es la longitud del intervalo estacional. β y γ son los coeficientes de suavizado de para la tendencia y el componente estacional. Cabe mencionar que estos coeficientes se estiman minimizando los errores cuadrados de predicción. Dicho lo anterior, y para cada realización h del horizonte $H \geq 1$ de predicción, el algoritmo Holt-Winters está definido por:

$$\hat{Y}_{t+h} = l_t + hb_t + s_{t+h-p}$$

2.6.3 Método de Suavizado Exponencial de Espacio de Estado

El modelo de Suavizado Exponencial con enfoque en la metodología de Espacio con Estado describe un marco de análisis para seleccionar un modelo de entre la taxonomía de modelos de suavizado exponencial de Makridakis et al. (1998), a través de Hyndman et al. (2002), sobre la cual se elabora en la Tabla 1.

Modelos de Suavizado Exponencial			
Componente de Tendencia	Componente Estacional		
	N (ninguno)	A (aditivo)	M (multiplicativo)
N (ninguno)	NN	NA	NM
A (aditivo)	AN	AA	AM
M (multiplicativo)	MN	MA	MM
D (atenuado)	DN	DA	DM

Tabla 1. Taxonomía de modelos de suavizado exponencial.¹⁰

Las configuraciones corresponden a la manera en la que interactúan los componentes, estacional y de tendencia, con las ecuaciones de nivel, tendencia y estacionalidad. Por ejemplo, algoritmo Holt-Winters que fue descrito antes (ver 2.6.2), corresponde a la combinación AA. Dicho algoritmo con estacionalidad multiplicativa correspondería a la combinación AM y las siguientes ecuaciones:

$$l_t = \alpha(l_t/s_{t-p}) + (1 - \alpha)(l_{t-1} + b_{t-1})$$

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1}$$

$$s_t = \gamma(Y_t/s_{t-p}) + (1 - \gamma)s_{t-p}$$

Como explican Hyndman et al. (2008), se selecciona el modelo de espacio de estado que minimice el criterio de Akaike:

$$AIC = \mathcal{L}^*(\hat{\theta}, \hat{Y}_0) + 2q$$

donde \mathcal{L}^* es la función de verosimilitud y q indica el número de parámetros en θ contando los grados de libertad en Y_0 .

¹⁰ Esta tabla está basada en la de Hyndman et al. (2002).

2.6.4 Modelo Theta

El Modelo Theta fue presentado por Assimakopoulos et al. (2000) como un método de pronósticos para series univariadas, en el que un coeficiente se aplica a las segundas diferencias de la serie (ver 2.6.1) para modificar la curvatura local de ésta. Como explica, de manera simplificada, Hyndman (2001): a partir de una serie $\{Y_1, \dots, Y_t\}$ se construye otra $\{Y_{1,\theta}, \dots, Y_{t,\theta}\}$, de tal forma que:

$$\Delta^2 Y_n = \Delta^2 Y_{n,\theta}$$

para $n = 3, \dots, t$. La solución a la ecuación anterior es:

$$Y_{n,\theta} = a_\theta + b_\theta(t-1) + \theta Y_n$$

para $n = 1, \dots, t$. Además, a_θ y b_θ son constantes. De modo que se buscan los valores de $Y_{1,\theta}$ y $\Delta Y_{1,\theta}$ que minimicen la suma de las diferencias al cuadrado $\sum_{n=1}^t [Y_n - Y_{n,\theta}]^2$.

De acuerdo con la configuración de Assimakopoulos et al. (2000) para el M3, los pronósticos del método theta se estiman a través de un promedio ponderado de $Y_{n,\theta}$ para $\theta = \{0,2\}$, de modo que:

$$\hat{Y}_{t+h} = \frac{1}{2} [\hat{Y}_{t+h,0} + \hat{Y}_{t+h,2}]$$

2.6.5 Modelo de Descomposición de Tendencia Estacional

La metodología de Descomposición de Tendencia Estacional usando Loess fue desarrollada por Cleveland et al. (1990). Este algoritmo descompone una serie en sus componentes de Estacionalidad, Tendencia y Residuo.

El algoritmo STL (*Seasonal Trend decomposition by Loess*), utiliza un proceso de Suavizado; la Regresión Local (Loess), cuyo procedimiento se explica en el contexto de las series de tiempo, a continuación:

Se tiene un conjunto de datos (una serie) Y . Para cada observación Y_n , $n = \{1, \dots, t\}$ se escoge un subconjunto de la serie, de longitud $q \leq t$, con las q observaciones más cercanas a Y_n en el tiempo. Después, se realiza un ajuste de mínimos cuadrados ponderados en este subconjunto; dicho modelo pondera la cercanía entre Y_n y cada valor del subconjunto i.e. los puntos más cercanos tendrán mayor influencia en el ajuste. De esta forma, se obtiene la estimación \hat{Y}_n . Al iterarse este procedimiento para cada Y_n de la serie, resulta un ajuste suavizado de los datos.

Finalmente, para aislar el efecto de datos aberrantes, se vuelve a ponderar, ahora considerando la distancia entre Y y el conjunto de estimaciones \hat{Y} . El resultado es la regresión local (Loess), la cual generará una función $\hat{g}(n)$.

Dicho lo anterior, se puede explicar el algoritmo STL: La serie se dividirá en ciclos de sub-series, correspondientes a los intervalos de estacionalidad, a los cuales se les aplicará Loess y generarán el componente estacional de la serie. Al extraer dicho componente de la serie ‘bruta’, lo que se obtiene es el componente de tendencia y el residuo. De esta forma, se pueden aplicar métodos de pronósticos no-estacionales y después agregar el componente estacional.

2.6.6 TBATS: Suavizado Exponencial de Espacio de Estado con Transformación Box-Cox, Errores ARMA y Componentes de Tendencia y Estacionalidad

Como menciona Delivera et al. (2010), después del empleo del marco de análisis de espacio con estado en modelos estacionales, Taylor (2003) realizó una extensión de las especificaciones del algoritmo Holt-Winters en su variante lineal (ver 2.6.2), para añadir un segundo componente estacional. A su vez, Delivera et al. (2010), desarrollaron el algoritmo TBATS, el cuál es una extensión con componentes estacionales trigonométricos, del algoritmo BATS. El TBATS incorpora una transformación de Box-Cox y errores ARMA, un parámetro T para considerar otros posibles patrones estacionales y representaciones de componentes

estacionales basadas en series de Fourier, siguiendo la metodología y notación de Delivera et al. (2010):

$$Y_t^{(\omega)} = \begin{cases} \frac{Y_t^{(\omega)} - 1}{\omega} & \omega \neq 0 \\ \log Y_t & \omega = 0 \end{cases}$$

$$Y_t^{(\omega)} = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^T s^{(i)}_{t-m_i} + d_t$$

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha d_t$$

$$b_t = (1 - \phi)b + \phi b_{t-1} + \beta d_t$$

$$s_t^{(i)} = s^{(i)}_{t-m_i} + \gamma_i d_t$$

$$d_t = \sum_{i=1}^p \phi d_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t$$

$$s_t^{(i)} = \sum_{j=1}^{k_i} s_{j,t}^{*(i)}$$

$$s_{j,t}^{*(i)} = s_{j,t-1}^{*(i)} \cos \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \sin \lambda_j^{(i)} + \gamma_1^{(i)} d_t$$

$$s_{j,t}^{*(i)} = -s_{j,t-1}^{*(i)} \sin \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \cos \lambda_j^{(i)} + \gamma_2^{(i)} d_t$$

$$\lambda_j^{(i)} = \frac{2\pi j}{m_i}$$

$$\varepsilon_t \sim NID(0, \sigma^2)$$

El parámetro que define el periodo estacional, antes p , es ahora definido por m_i , $i = \{1, \dots, t\}$ dado que ahora p se refiere al orden de los rezagos del ARMA(p, q) (ver 2.6.1), $s_t^{(i)}$ es el componente estacional de orden i , ε_t es ruido blanco. Sobre los componentes estacionales trigonométricos, $s_{j,t}^{*(i)}$ describe el nivel en el componente estacional y $s_{j,t}^{*(i)}$ el gradiente de éste anterior y k_i se refiere al número de ciclos harmónicos que requiere el componente estacional.

2.6.7 Redes Neuronales con *Algoritmo Extreme Learning Machine*

ELM es un Algoritmo diseñado por Huang et al. (2004) para las redes neuronales paralimentarias de una sola capa oculta. Éstas son redes neuronales artificiales cuyas conexiones se alimentan la información en una sola dirección. Como se explica en la Figura 1, los nodos de entrada ‘entregan’ la información ponderada a cada nodo de una caja negra: la capa oculta de nodos. Cada nodo en esta capa oculta produce un resultado a través de una función de activación. Esta función es la que define la información en los nodos de la capa de salida.

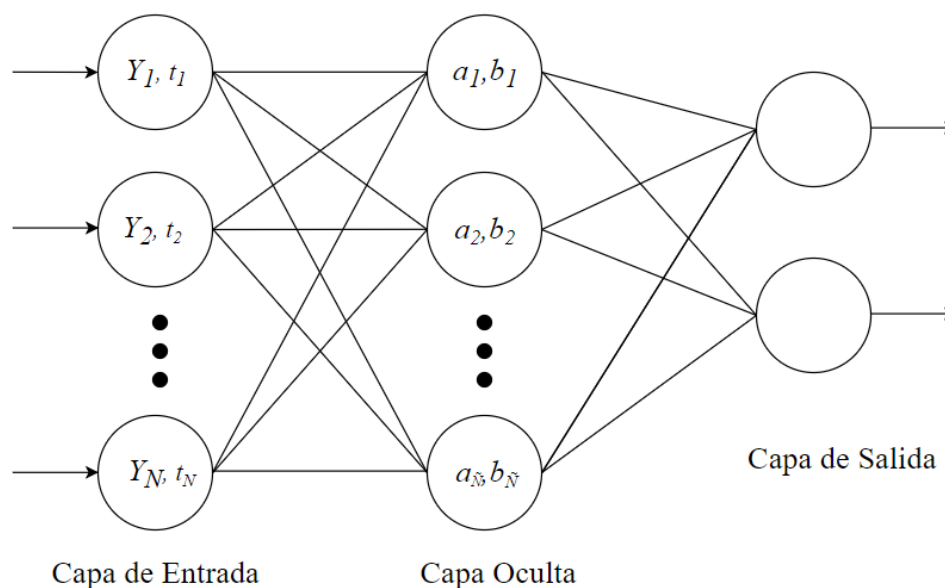


Figura 1. Estructura de una Red Neuronal Paralimentaria de una sola Capa Oculta.¹¹

El algoritmo ELM, como lo desarrolla Huang et al. (2004) a través de Ding et al. (2013) y siguiendo la notación de la Figura 1, consiste en:

- 1) Se introduce el conjunto de training $Y_i, t_i \in R^n \times R^n$ ($i = 1, \dots, N$) y la función de activación f y el nodo oculto \tilde{N} , lo que da como resultado las ponderaciones β .

¹¹ Figura de elaboración propia.

- 2) Se asignan de forma aleatoria los parámetros de aprendizaje de los nodos ocultos:
 $a_i, b_i, i = 1, \dots, \tilde{N}$
- 3) La capa oculta genera la capa de salida H .
- 4) Calcular la ponderación de resultado β .

El resultado se puede definir como:

$$o_j = \sum_{i=1}^{\tilde{N}} \beta_i f_i(x_j; a_i, b_i), \quad j = 1, \dots, N$$

la cual es un sistema lineal, de modo que β , se pueden determinar encontrando la solución de mínimos cuadrados.

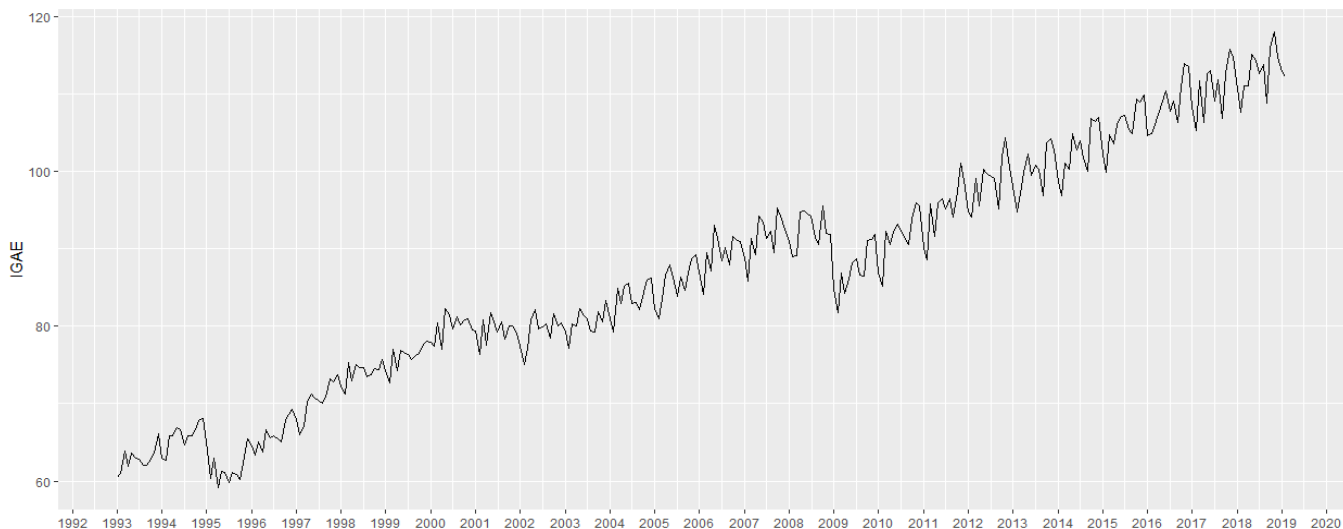
2.6.8 Red Neuronal Paralimentaria con una sola Capa Oculta y Entradas Rezagadas

Este método consiste en el ajuste de una red neuronal como la de la Figura 1 en la metodología del punto anterior (ver 2.6.7), pero con los valores rezagados Y_i, t_i de la serie como información para los nodos de entrada. La red arroja pronósticos en horizonte $H = 1$; para $H > 1$ la red computa los pronósticos de manera recursiva i.e. utilizando los pronósticos anteriores como información de entrada.

Capítulo 3: Descripción de las Series

3.1 Series Mensuales

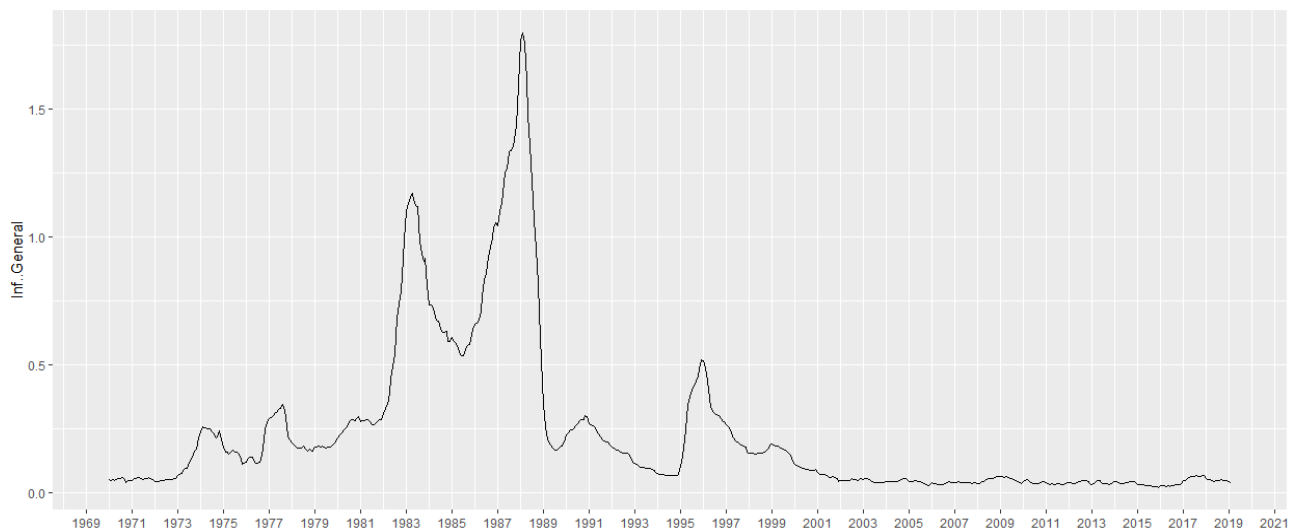
- **Indicador Global de la Actividad Económica (IGAE):** es un índice que representa la evolución del sector real de la economía. Se calcula a través de índices de volumen físico de la producción real, se utilizaron los datos de 1993-01-01 a 2018-12-01. (INEGI, 2018)



Gráfica 1. Serie histórica: IGAE, INEGI.¹²

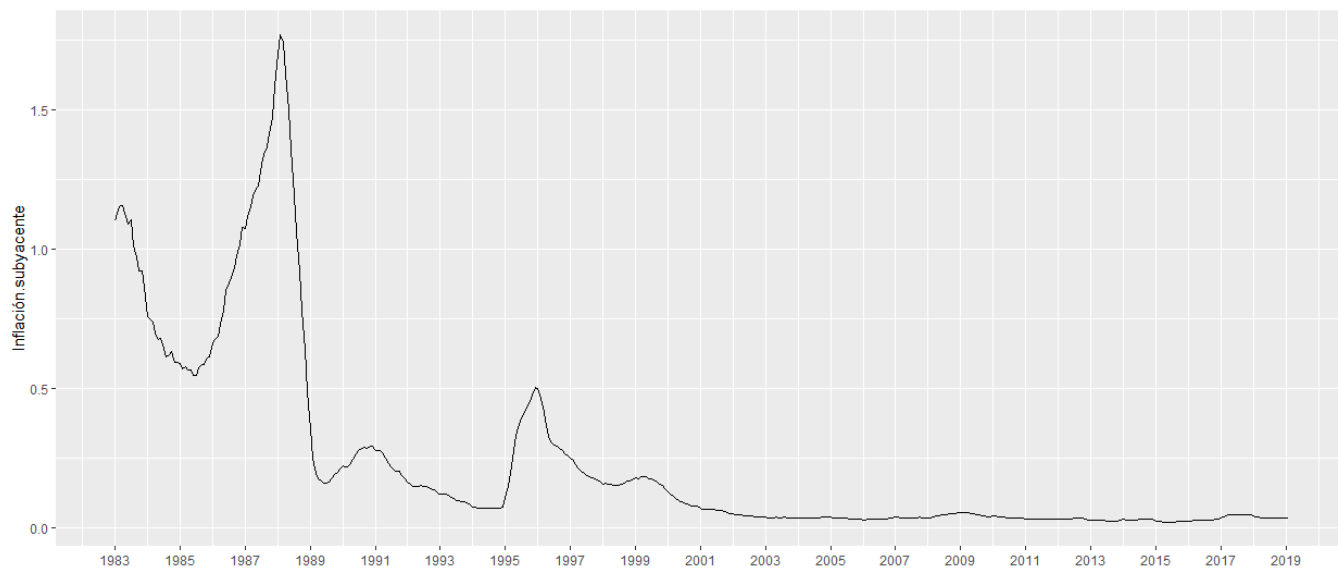
- **Inflación General:** representa el aumento porcentual generalizado de los precios. Se utilizó la inflación mensual interanual, la cual se calcula con la variación porcentual del Índice de Precios al Consumidor, del mismo mes, de un año a otro. Se utilizaron datos de 1970-01-01 a 2018-12-01. (INEGI, 2018)

¹² Gráfico de elaboración propia, realizado con software *R*, librería *ggplot2*.



Gráfica 2. Serie histórica: Inflación General, INEGI.¹³

- Inflación subyacente: se trata de la inflación calculada con el IPC, sin considerar los precios de los energéticos y alimentos antes de su elaboración, debido a que estos dos son más volátiles. Se utilizaron los datos de 1983-01-01 hasta 2018-12-01. (INEGI, 2018)



Gráfica 3. Serie histórica: Inflación Subyacente, INEGI.¹⁴

¹³ Gráfico de elaboración propia, realizado con software R, librería *ggplot2*.

¹⁴ Gráfico de elaboración propia, realizado con software R, librería *ggplot2*.

- Tipo de Cambio: Es el precio del dólar estadounidense (USD) en pesos mexicanos (MXN) en el mercado cambiario. Se calculó con los promedios de los cierres diarios. Se utilizaron los datos de 1994-01-01 hasta 2018-12-01. (FRED, 2018)

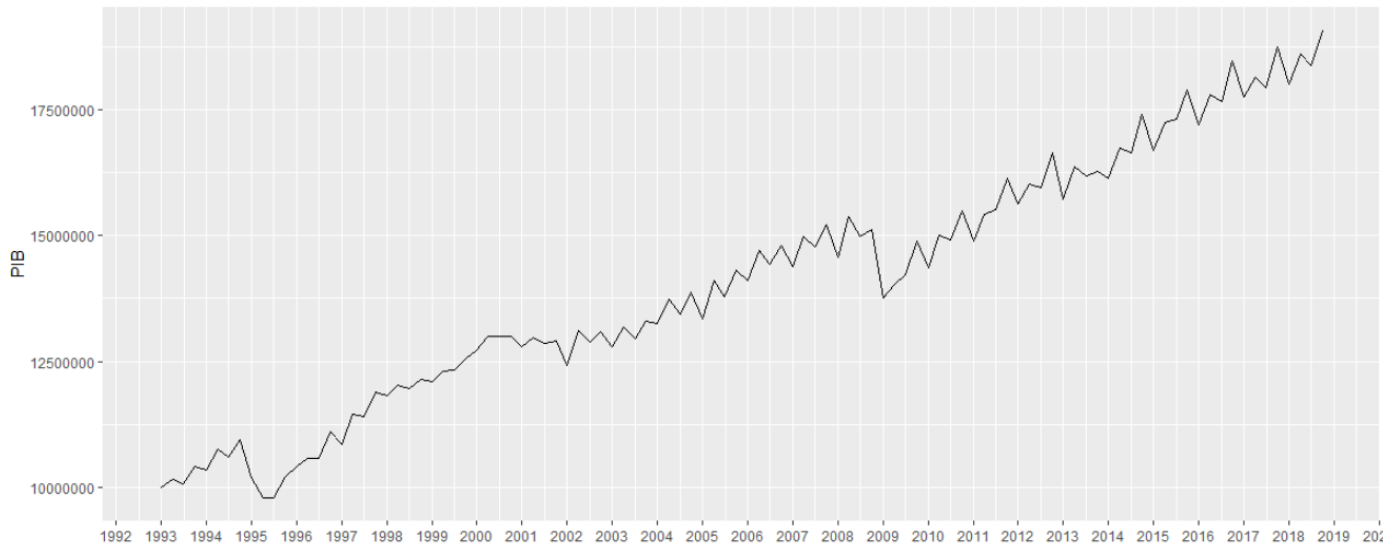


Gráfica 4. Serie histórica: Tipo de Cambio, FRED.¹⁵

3.1 Series Trimestrales

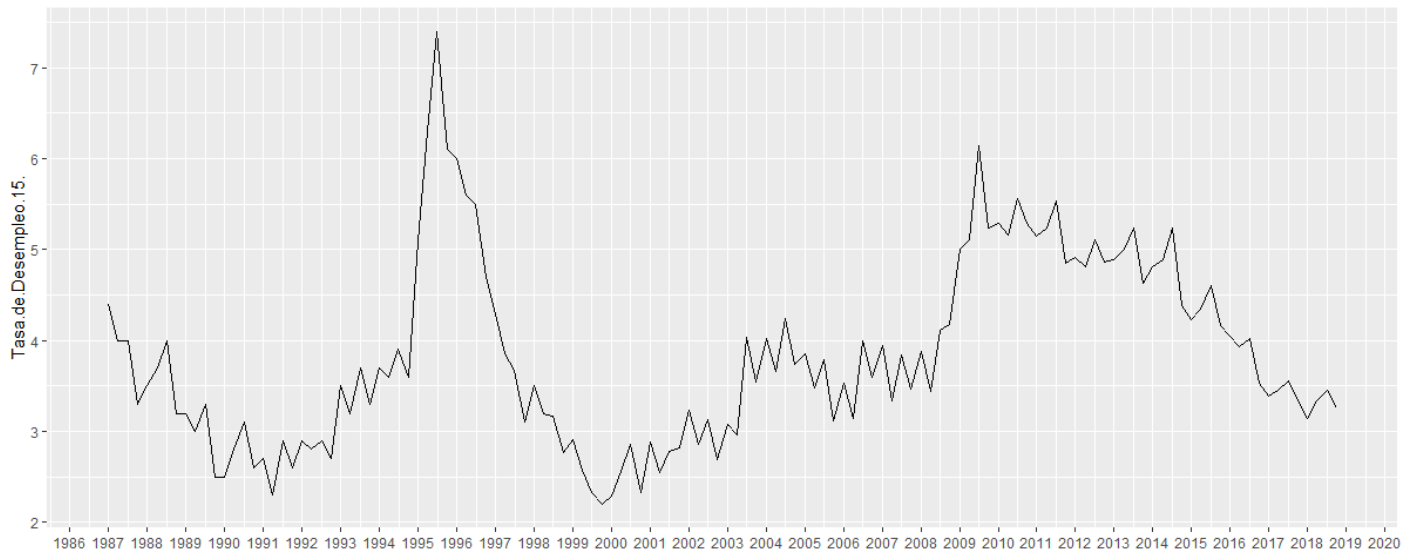
- Producto Interno Bruto (PIB): indica el valor monetario de la producción nacional total de bienes y servicios en cierto periodo. Se utilizaron los datos de 1993-01-01 hasta 2018-10-01. (INEGI, 2018)

¹⁵ Gráfico de elaboración propia, realizado con software R, librería *ggplot2*.



Gráfica 5. Serie histórica: PIB, INEGI.¹⁶

- Tasa de Desempleo: representa la proporción de la población de 15 años o mayor, que carece de empleo. Se utilizaron los datos de 1987-01-01 hasta 2018-10-01. (OECD, 2018)



Gráfica 6. Serie histórica: Tasa de Desempleo, OECD.¹⁷

¹⁶ Gráfico de elaboración propia, realizado con software R, librería *ggplot2*.

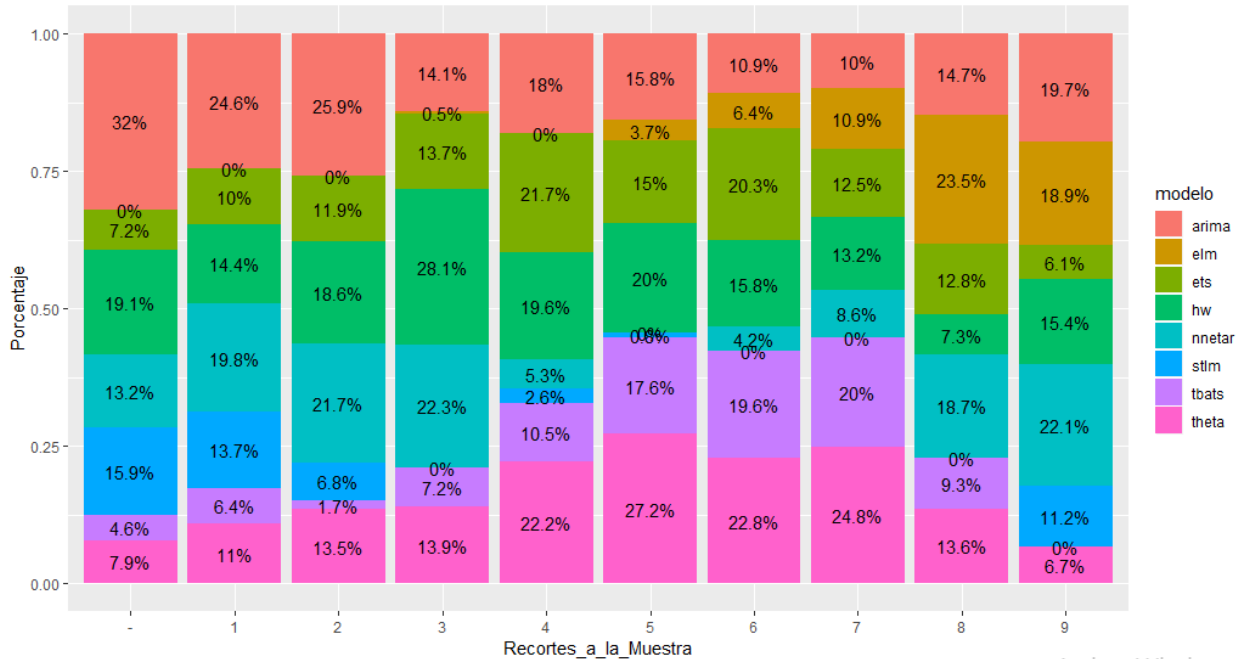
¹⁷ Gráfico de elaboración propia, realizado con software R, librería *ggplot2*.

Capítulo 4: Resultados

4.1 Desempeño de los Modelos en Estabilidad Secuencial y Discusión

Cada una de las barras en los Histogramas de Estabilidad Secuencial muestra las ponderaciones de los métodos utilizados, estimadas por el algoritmo AFTER para el proceso de *pooling* (Ver 2.3), donde cada iteración i es un subconjunto de la serie con i recortes desde el dato más reciente. Gracias a los histogramas de estabilidad secuencial, podemos observar cómo se alterna la persistencia de los diferentes modelos a través del tiempo, i.e. para subconjuntos de la muestra. En todos los histogramas se observa un intervalo de persistencia del modelo ARIMA (ver 2.6.1). Cuando la estabilidad de los modelos no es constante, este intervalo se alterna con otro intervalo en el que prepondera otro modelo, en la mayoría de los casos, no lineal. En la Gráfica 7, por ejemplo, se puede observar que, hasta el tercer recorte de la muestra, el proceso ARIMA es el modelo con mejor precisión en términos de la función de pérdida msMAPE, hasta que el modelo Theta lo supera en el intervalo de 4 a 7 recortes, de modo que éste alcanza una eficiencia de más de 200% mejor que el ARIMA.

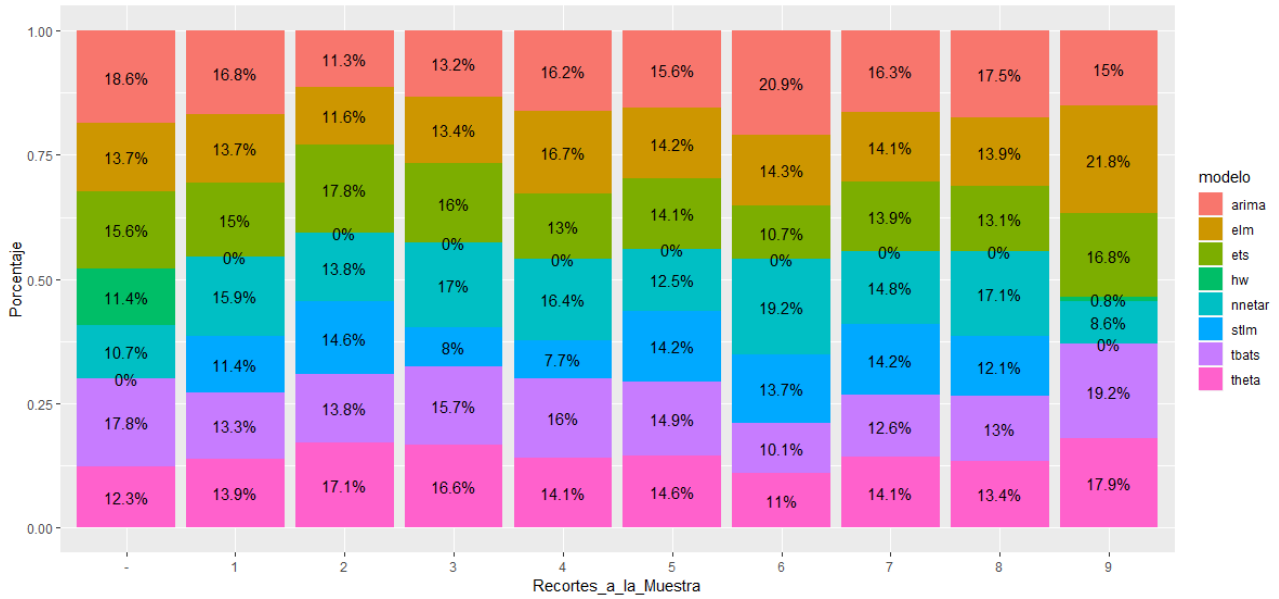
En esta misma Gráfica 7 se presenta un aumento exponencial de la proporción escogida por AFTER, de la red neuronal con algoritmo ELM, a partir del recorte 6. Es decir, la tendencia sugiere que a partir de ese corte en el tiempo, esta metodología dejó de capturar de forma exponencial el proceso estocástico de la serie. La mayor implicación de esto es que la tendencia de la ponderación AFTER a través del tiempo puede ser una herramienta útil para descartar modelos en pronósticos de series univariadas de horizonte $H = 1$



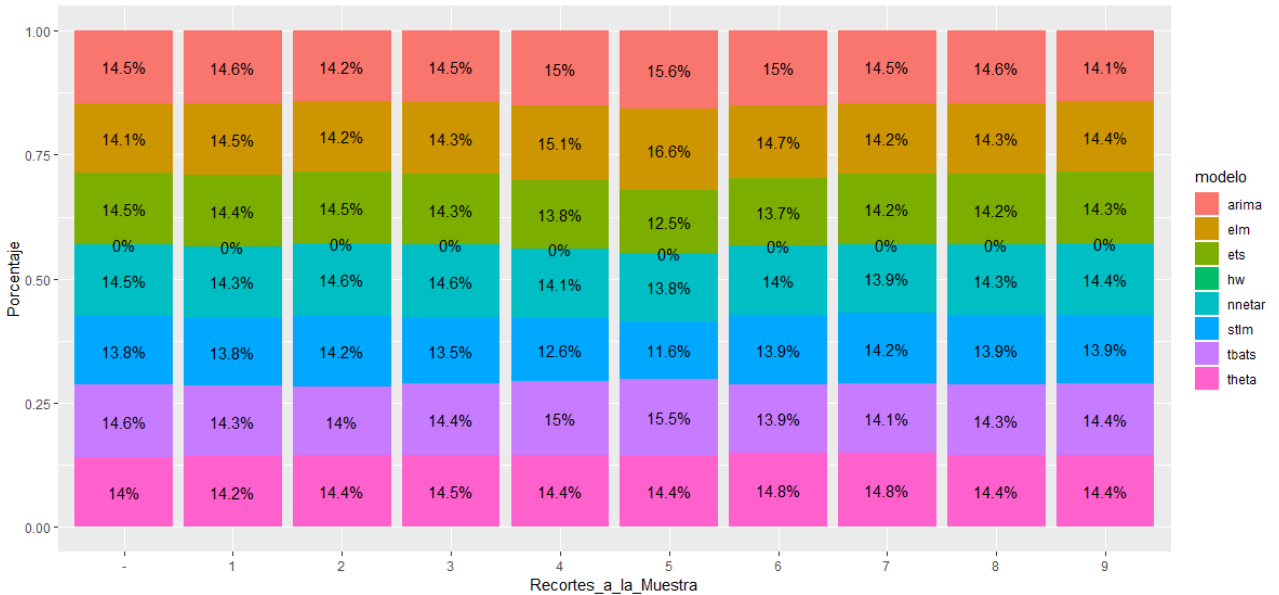
Gráfica 7. Histograma de Estabilidad Secuencial para el IGAE: Porcentajes de ponderación de los modelos utilizados, calculados con el algoritmo AFTER, para 9 recortes a la muestra.¹⁸

Un segundo caso, contrario al primero, se observa en la Gráfica 8 y Gráfica 9; los modelos presentan una estabilidad secuencial constante para estas series, sobre todo en la de inflación subyacente. Lo más lógico, al no observar una tendencia en la estabilidad secuencial, es descartar los modelos que, en promedio, tengan el peor desempeño y después computar el algoritmo AFTER. Cabe destacar que estos dos histogramas sugieren, de forma contingente, que las series más volátiles podrían tener menor estabilidad secuencial en sus modelos, como sucede con la serie de inflación general en comparación a la subyacente.

¹⁸ Gráfico de elaboración propia, realizado con software R, librería ggplot2.



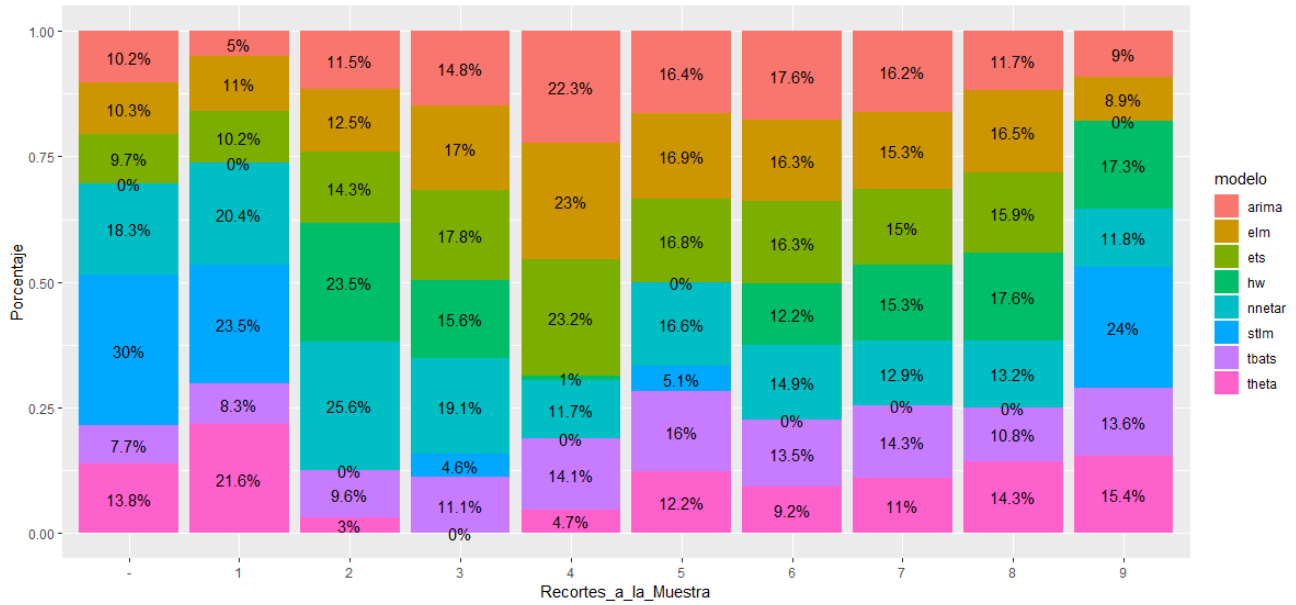
Gráfica 8. Histograma de Estabilidad Secuencial para la Inflación General: Porcentajes de ponderación de los modelos utilizados, calculados con el algoritmo AFTER, para 9 recortes a la muestra.¹⁹



Gráfica 9. Histograma de Estabilidad Secuencial para la Inflación Subyacente: Porcentajes de ponderación de los modelos utilizados, calculados con el algoritmo AFTER, para 9 recortes a la muestra.²⁰

¹⁹ Gráfico de elaboración propia, realizado con software R, librería ggplot2.

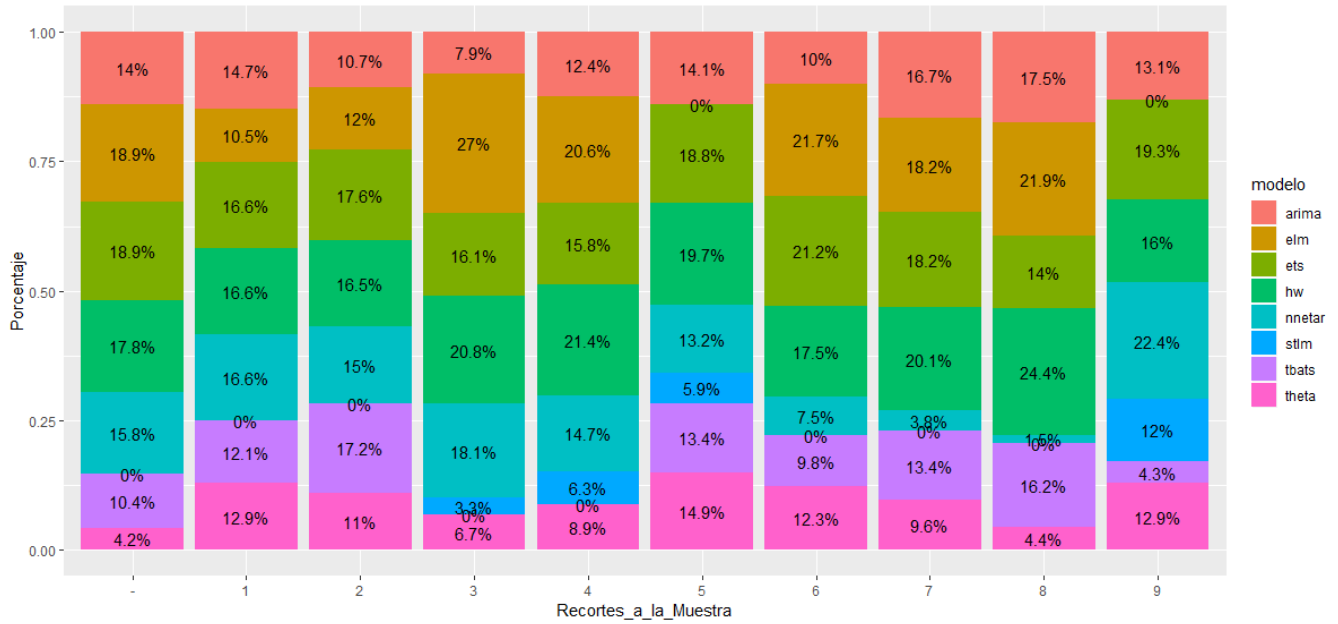
²⁰ Gráfico de elaboración propia, realizado con software R, librería ggplot2.



Gráfica 10. Histograma de Estabilidad Secuencial para Tipo de Cambio: Porcentajes de ponderación de los modelos utilizados, calculados con el algoritmo AFTER, para 9 recortes a la muestra.²¹

La Gráfica 11, para la serie del tipo de cambio, muestra heterogeneidades notables en la ponderación de los modelos y un patrón de estabilidad secuencial con mucha variabilidad. No se observan patrones, de modo que éste corresponde a un tercer caso de análisis para la selección de modelos: cuando no hay tendencias definidas en las ponderaciones AFTER ni estabilidad secuencial. Una buena práctica, similar al caso anterior, sería descartar los modelos que en los recortes más recientes a la muestra tengan el peor desempeño y realizar el proceso de *pooling* con el algoritmo AFTER.

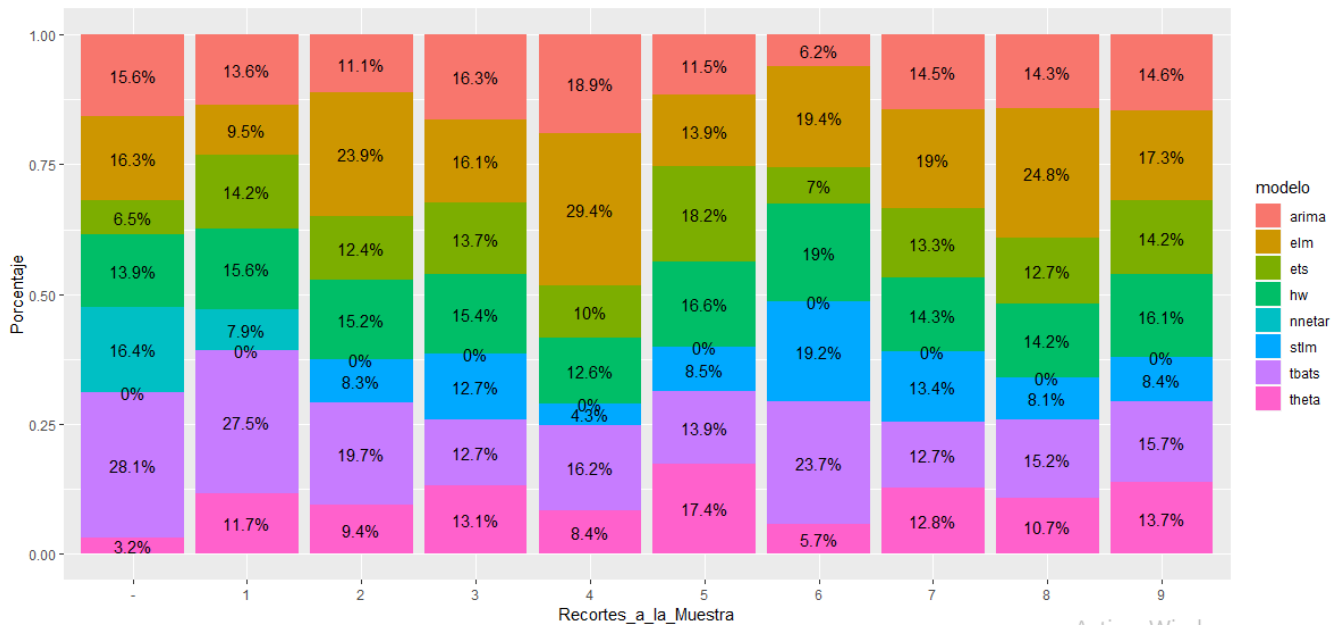
²¹ Gráfico de elaboración propia, realizado con software R, librería *ggplot2*.



Gráfica 11. Histograma de Estabilidad Secuencial para PIB: Porcentajes de ponderación de los modelos utilizados, calculados con el algoritmo AFTER, para 9 recortes a la muestra.²²

Convenientemente, un cuarto caso aparece en la Gráfica 12: varios modelos presentan tendencias de estabilidad secuencial parecidas. Más específicamente, los métodos ARIMA, la red neuronal con algoritmo ELM y ETS, tienen un crecimiento en su ponderación a la mitad del intervalo de recortes a la muestra. En este caso, lo más conveniente sería utilizar el método que mejor capture el proceso estocástico de la serie utilizando el criterio msMAPE.

²² Gráfico de elaboración propia, realizado con software R, librería ggplot2.



Gráfica 12. Histograma de Estabilidad Secuencial para Tasa de Desempleo: Porcentajes de ponderación de los modelos utilizados, calculados con el algoritmo AFTER, para 9 recortes a la muestra.²³

Como se puede observar en la Tabla 2, a pesar de que el modelo ARIMA aparece consistentemente como uno de los métodos más eficientes en las simulaciones dentro de muestra a diferentes cortes, otros modelos no-lineales generan, a lo largo de distintas series, menor varianza. Es necesario observar de cerca las similitudes en el desempeño de los métodos ARIMA, STLM y Theta, dado que, para esta muestra de series, el error promedio es muy similar.

²³ Gráfico de elaboración propia, realizado con software R, librería ggplot2.

Método	Promedio	IGAE	Inf. Gen.	Inf. Sub.	Tipo de Cambio	PIB	Tasa Desempleo
NNETAR	0.01735872	0.00938222	0.01468496	0.00613241	0.03362008	0.00844751	0.03188516
ETS	0.02121949	0.01031804	0.00849567	0.00654368	0.04877307	0.00625746	0.04692903
STLM	0.02303349	0.00894829	0.02821268	0.01779682	0.01264624	0.01968451	0.0509124
Theta	0.02453815	0.01021111	0.01269358	0.01396801	0.04158795	0.01667012	0.05209815
ARIMA	0.02751064	0.06395555	0.0047113	0.00560931	0.0479262	0.00975163	0.03310987
TBATS	0.03274045	0.10730986	0.00563711	0.00455695	0.05248021	0.01230293	0.01415564
ELM	0.03613789	0.11465335	0.01093169	0.01199402	0.04091077	0.00627271	0.03206482
Holt-Winters	0.06253969	0.0084405	0.01385049	0.24415057	0.06611314	0.00700414	0.03567929

Tabla 2. Error msMAPE dentro de muestra, promedio y para cada serie, de cada modelo, en un horizonte de training de $H=6$ para series mensuales y $H=4$ para series trimestrales.²⁴

²⁴ Gráfico de elaboración propia, basado en los resultados computados en Software R.

Capítulo 5: Conclusiones

A pesar de los antecedentes sobre el desempeño de los modelos lineales comparado con el de los no-lineales, como es el caso del estudio de Stock y Watson (1998), las especificaciones de modelos lineales con prueba de raíz unitaria, como lo es el ARIMA, no resultaron ser los más eficientes en términos del error msMAPE dentro de muestra. Este resultado se ve matizado en este estudio, con los diagramas de estabilidad secuencial, los cuales nos han permitido abstraer una taxonomía sobre los casos de estabilidad secuencial, así como un criterio de selección de modelos para cada tipo de patrón de comportamiento.

En general, el conjunto de datos nos muestra que cuando la estabilidad secuencial es heterogénea a través del tiempo y el desempeño del modelo ARIMA es bajo, hay un intervalo en el que otro modelo, usualmente no-lineal, captura mejor el comportamiento estocástico de la serie, por lo que es más conveniente utilizarlo.

Las tendencias sobre estabilidad secuencial son el criterio que justifica que los procesos de *pooling* estimen mejores pronósticos en horizontes $H > 1$ que los métodos que sólo estiman utilizando un método único, por lo que puede ser una herramienta muy útil al momento de seleccionar un portafolio de modelos para realizar pronósticos.

Bibliografía

1. Amisano, G., Geweke J. (2017). *Prediction Using Several Macroeconomic Models*. The Review of Economics and Statistics 99 (5), 912-925.
2. Assimakopoulos, V., Nikolopoulos K. (2000). *The theta model: a decomposition approach to forecasting*. International Journal of Forecasting 16, 521–530.
3. Bontempi, G., Le Borgne Y., Ben Taieb S. (2013). *Chapter: Machine Learning Strategies For Time Series Forecasting, Book: Lecture Notes In Business Information Processing*. 1865-1348.
4. Box, G., Jenkins, G., Reinsel, G., Ljung G. (2016). *Time Series Analysis: Forecasting and Control*. Hoboken, NJ: John Wiley y Sons.
5. Chen Z., Yang Y. (2004). *Assessing Forecast Accuracy Measures*.
6. Clement, M., Hendry, D. (1998). *Forecasting economic times series*. Cambridge University Press.
7. Cleveland R., Cleveland W., McRae, J. y Terpenning I. (1990). *STL: A Seasonal-Trend Decomposition Procedure Based on Loess*. Journal of Official Statistics, 6, 3-73.
8. De Livera, A., Hyndman, R., y Snyder, R. (2011). *Forecasting time series with complex seasonal patterns using exponential smoothing*. Journal of the American Statistical Association, 106(496), 1513-1527.
9. Ding, S., Zhao, H., Zhang, Y., Xu, X., Nie, R. (2013). *Extreme learning machine: algorithm, theory and applications*. Artificial Intelligence Review, 44(1), 103-115.
10. Hanke, J., Wichern D. (2015). *Business Forecasting*. [Chennai]: Pearson.
11. Holt, C. (1957). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20(1), 5-10.
12. Huang, B., Zhu, Q., Siew, C. (2004). Extreme learning machine: A new learning scheme of feedforward neural networks. *Neurocomputing*, 70, 489-501.

13. Hyndman, R., Koehler, A., Ord, K., Snyder, R. (2008). *Forecasting with exponential smoothing: the state space approach*. Berlin: Springer Berlin.
14. Hyndman, R., Koehler, A., Snyder, R., Grose, S. (2002). *A state space framework for automatic forecasting using exponential smoothing methods*. International Journal of Forecasting, 18(3), 439-454.
15. Makridakis, S., Michèle H. (2000). *The M3-Competition: Results, Conclusions and Implications*. International Journal of Forecasting 16 (4): 451-476.
16. Makridakis, S., Wheelwright S., Hyndman R. (1998). *Forecasting: methods and applications*, John Wiley y Sons: New York.
17. Stock, J., Watson M. (1998). *A Comparison Of Linear And Nonlinear Univariate Models For Forecasting Macroeconomic Time Series*. NBER Working Paper Series.
18. Winters, P. (1960). *Forecasting Sales by Exponentially Weighted Moving Averages*. Management Science, 6(3), 324-342.
19. Yang, Y. (2001). *Combining forecasting procedures: Some theoretical results*. Accepted by Econometric Theory.
20. Zou H., Yang Y. (2004). *Combining Time Series Models for Forecasting*. International Journal of Forecasting, 20(1), 69-84.

Bases de Datos:

21. Board of Governors of the Federal Reserve System (US), Mexico / U.S. Foreign Exchange Rate [DEXMXUS], obtenida de FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/DEXMXUS>. Consultada el 15 de mayo de 2019.
22. Instituto Nacional de Estadística, Geografía e Informática, (2018). Índice Nacional de Precios al Consumidor (INPC). México: INEGI. <https://www.inegi.org.mx/programas/inpc/2018/> Consultada el 15 de mayo de 2019.

23. Instituto Nacional de Estadística, Geografía e Informática, (2018). Indicador Global de Actividad Económica (IGAE). México: INEGI. <https://www.inegi.org.mx/temas/igae/> Consultada el 15 de mayo de 2019.
24. Instituto Nacional de Estadística, Geografía e Informática, (2018). Producto Interno Bruto (PI). México: INEGI. <https://www.inegi.org.mx/temas/pib/> Consultada el 15 de mayo de 2019.
25. Organización para la Cooperación y el Desarrollo Económicos, Unemployment Rate: Aged 15 and Over: All Persons for Mexico [LRUNTTTTMXQ156N], obtenida de: FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/LRUNTTTTMXQ156N>. Consultada el 15 de mayo de 2019.