

Pronósticos de inflación en Guatemala: ¿Modelos de series de tiempo o algoritmos de *machine learning*?*

Gabriel A. Fuentes[†]

7 de abril, 2022

Resumen

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque nibh ex, mollis nec turpis ac, ultrices mollis quam. Aliquam sed tortor eget dolor dignissim ornare quis in nibh. Duis non nisl convallis, ornare erat et, sagittis mauris. Mauris sit amet sapien vehicula, volutpat mi laoreet, viverra tellus. Aliquam id placerat risus, sed convallis urna. Phasellus varius iaculis tellus ac venenatis. Pellentesque molestie libero orci, in hendrerit dolor semper quis. Nunc sed dolor facilisis, faucibus odio semper, accumsan tellus. Proin pretium felis et urna commodo, in tempor sapien iaculis. Quisque feugiat tincidunt porta. Nulla vestibulum vel nibh non rhoncus. Nunc eu urna velit. In imperdiet magna eget ex auctor, porta aliquam lorem cursus. Nullam et commodo mauris. In ullamcorper egestas erat, quis faucibus lectus. Nullam pellentesque quam eget nibh condimentum, in vestibulum neque posuere. Sed sed lorem vitae lacus laoreet aliquet. Curabitur egestas enim a odio lacinia sodales. Vestibulum fermentum, dui.

Palabras clave: Machine learning, series de tiempo, pronósticos, inflación

Clasificación JEL: E31, C22, C45, C53

*Con el propósito de maximizar la reproducibilidad del presente estudio, las bases de datos utilizadas y el código en R se encuentran disponibles en: <https://github.com/gafnts/Inflation-forecasting>

[†]Séptimo semestre, Teoría y Política Monetaria. Correo electrónico: gafnts@gmail.com

1 Introducción

El artículo número 3 de la ley orgánica del Banco de Guatemala establece que su objetivo fundamental consiste en propiciar las condiciones monetarias, cambiarias y crediticias que promuevan la estabilidad en el nivel general de precios. En el 2005, como parte de su esfuerzo por lograr este fin, dicha entidad optó por conducir sus acciones a través de un esquema de metas inflacionarias.

Debido a que naturalmente existe un periodo de tiempo entre el ajuste de las condiciones en el mercado de dinero y el efecto que este cambio genera en las variables del sector real, la autoridad monetaria—bajo un régimen de metas explícitas de inflación—actúa en el presente considerando sus pronósticos sobre el comportamiento de los precios en el futuro.

Concretamente, un cambio en la tasa de interés de política monetaria puede demorar entre 12 y 24 meses en ejercer una influencia en la demanda agregada y, por tanto, en tener un impacto en la evolución del ritmo inflacionario interno. En consecuencia, el banco central deberá pronosticar esta última variable durante dicho horizonte temporal y efectuar sus decisiones de política monetaria de modo que pueda estar seguro de que los pronósticos y expectativas de inflación se mantendrán anclados a la meta establecida durante el mediano plazo (Moenjak, 2014).

En este sentido, los pronósticos de variables macroeconómicas pueden llevarse a cabo por medio de dos enfoques distintos: Métodos estructurales y métodos no estructurales (Diebold, 1998). Los primeros informan a la especificación de sus modelos a través de una teoría económica determinada, mientras que los últimos se valen de las correlaciones en forma reducida que subyacen en las series de tiempo, sin depender explícitamente de una teoría económica (Pratap & Sengupta, 2019).

Este estudio en particular se ocupará del segundo enfoque, que al mismo tiempo puede ser subdividido en dos metodologías diferentes. Por un lado se encuentran los métodos econométricos de series de tiempo, los cuales son considerados como modelos relativamente tradicionales, de los que sobresalen los modelos univariados autorregresivos integrados de medias móviles (ARIMA) y los modelos multivariados de vectores autorregresivos (VAR).

Por el otro, algoritmos de aprendizaje estadístico¹ comienzan a ser cada vez más populares, principalmente por la creciente disponibilidad de grandes bases de datos y poder de cómputo, así como un mayor acceso a *software* especializado (Rodríguez-Vargas, 2020), aunque—tal y como mencionan Coulombe et al. (2020)—los métodos de aprendizaje de máquina en realidad tienen una historia dentro de la literatura macroeconómica que se remonta a inicios de la década de los noventas.

En general, al emplear modelos de series de tiempo, se asume que las variables utilizadas durante la estimación de los parámetros se vinculan entre sí a través de una dinámica intrínseca regida por relaciones lineales que únicamente conducen a soluciones que oscilan periódicamente o que exhiben un comportamiento exponencial, de modo que la totalidad en la conducta irregular del sistema es atribuida únicamente a una entrada que es tanto exógena como estocástica (Kantz & Schreiber, 2004).

Sin embargo, esta entrada aleatoria puede no ser la fuente exclusiva de irregularidad. Evidencia apunta que tanto series macroeconómicas como datos financieros exhiben interesantes estructuras no-lineales que se originan debido al impacto de perturbaciones durante las fases del ciclo económico (Granger et al., 1993; LeBaron, 1994).

¹A lo largo de la presente investigación los términos “aprendizaje estadístico,” “aprendizaje de máquina” y “aprendizaje automático” serán empleados indistintamente.

Dada la posible naturaleza no-lineal entre la relación de las variables económicas, existe cada vez un mayor énfasis en la aplicación de modelos de aprendizaje automático, los cuales—a diferencia de sus contrapartes—son particularmente buenos para explotar este tipo de relaciones (así como interacciones de alto orden) durante el proceso de predicción de nuevos valores basados en los valores de sus regresores (Athey & Imbens, 2019). Precisamente, tal es el propósito del presente estudio.

Mediante una competencia de pronósticos fuera de muestra (*pseudo out-of-sample*) del ritmo inflacionario en Guatemala, esta investigación buscará evaluar si modelos de aprendizaje estadístico como máquinas de vectores de soporte (SVM), bosques aleatorios (RF), algoritmos de potenciación de gradiente extremo (XGB), perceptrones multicapa (MLP) y redes neuronales recurrentes de larga memoria de corto plazo (LSTM) son capaces de generar pronósticos más eficientes—evaluados durante 4 horizontes temporales (3, 9, 12, y 24 meses) y comparados con la ayuda de 3 medidas de precisión (MAE, rMAE y MASE)—al contrastarlos con las predicciones de un repertorio de modelos estimados a partir de métodos econométricos tales como un modelo de camina aleatoria estacional (*snaiïve*), un modelo de suavizamiento exponencial (*Holt-Winters*), un modelo autorregresivo (AR), uno autorregresivo integrado de medias móviles estacionales (SARIMA) y un modelo de vectores autorregresivos (VAR).

El resto del documento será organizado de la siguiente manera. La próxima sección hace un repaso de la literatura correspondiente al uso y evaluación de modelos paramétricos y no paramétricos en tareas de predicción para series de tiempo. La sección 3 se ocupa de presentar un resumen estadístico de la base de datos, la descripción y configuración computacional de los modelos empleados, además de la metodología de evaluación para las predicciones de los mismos. Por último, la sección 4 y 5 se enfocan en discutir los resultados y conclusiones del estudio, respectivamente.

2 Revisión de la literatura

Algunos autores afirman que la literatura dedicada a la comparación de pronósticos entre modelos econométricos y de aprendizaje de máquina es relativamente reciente. Por ejemplo, Rodríguez-Vargas (2020) observa que, si bien desde inicios de los años noventas ha habido diferentes estudios en los que variables como tasas de inflación y tipos de cambio son pronosticadas utilizando k vecinos más cercanos (k -NN) y redes neuronales artificiales (ANN), en ningún momento estos autores se enfocaron en producir evaluaciones de las propiedades de sus predicciones de manera sistemática.

No obstante, ya desde esa década diversos estudios evalúan los resultados de pronósticos generados con redes neuronales artificiales al contrastarlos con las predicciones obtenidas por medio de modelos autorregresivos integrados de medias móviles y modelos de suavizamiento exponencial para series de tiempo en ámbitos micro y macroeconómicos.

Kang (1991) encuentra que las predicciones de modelos ARIMA son iguales o superiores a las de 18 redes neuronales con arquitecturas diferentes. Foster et al. (1992) obtienen resultados similares: modelos de suavizamiento exponencial y de mínimos cuadrados ordinarios producen pronósticos más eficientes que redes neuronales en series de tiempo anuales y mensuales, aunque semejantes en series trimestrales. Mientras tanto, Hill et al. (1996) concluyen que los pronósticos generados por redes neuronales son comparables con las predicciones de modelos econométricos tradicionales para series anuales e incluso superiores en series mensuales y trimestrales.

Recientemente, Ülke et al. (2018) evalúan la eficiencia de 2 modelos de series de tiempo univariados (RW y AR), 2 multivariados (VAR y ARDL) y 3 modelos de aprendizaje estadístico (k -NN, ANN y SVR) en la predicción de tasas de inflación en Estados Unidos para el periodo de 1984 a 2014.

Los autores pronostican dicha variable en 4 horizontes temporales (3, 6, 9 y 12 meses) para luego comprar los resultados mediante la raíz del error cuadrático medio (RSME) y el coeficiente de determinación (R^2) de cada uno de los modelos. Los resultados les permiten concluir que la regresión de vectores de soporte (SVR) supera al resto de métodos al pronosticar el ritmo de inflación subyacente medido con el índice de precios de gastos de consumo personal (PCE), mientras que el modelo autorregresivo distribuido rezagado (ARDL) proporciona la mayor precisión en la predicción de la inflación subyacente medida a través del índice de precios al consumidor (CPI).

Makridakis et al. (2018) también contrastan la eficiencia de pronóstico entre 8 algoritmos de aprendizaje automático y 8 modelos econométricos—utilizando un modelo de caminata aleatoria (RW) como punto de referencia—al generar predicciones para 3003 series que proceden de dominios micro y macroeconómicos. Los autores concluyen que los métodos tradicionales son más precisos que los de *machine learning*.

Sin embargo, Cerqueira et al. (2019) consideran que este último estudio se encuentra sesgado en lo que respecta al tamaño de muestra, debido a que la duración promedio de las series es de 116 observaciones. Curvas de aprendizaje elaboradas por medio de procedimientos precuenciales les permiten determinar que a medida que el tamaño de muestra se amplía, los pronósticos generados por modelos de aprendizaje de máquina superan en términos de eficiencia a los métodos tradicionales.

Por otra parte, Parmezan et al. (2019) realizan un meta-análisis de la literatura y concluyen que los 2 modelos más utilizados en 29 estudios que involucran métodos econométricos y de aprendizaje automático a lo largo de la última década son redes neuronales artificiales y modelos ARIMA. Posteriormente, comparan 7 modelos tradicionales con 4 modelos de aprendizaje estadístico durante la predicción de 40 series sintéticas y 55 series reales. Los resultados les permiten concluir que tanto SARIMA, SVM y k-NN son los métodos más efectivos para el modelado y predicción de series de tiempo, en ese orden.

Katris (2020) evalúa el desempeño de 3 modelos estadísticos (ARIMA, ARFIMA y ARFIMA/GARCH) con relación a 3 modelos de aprendizaje de máquina (ANN, SVM y MARS) al pronosticar las tasas de desempleo mensual para 22 economías desarrolladas en 3 horizontes temporales distintos. El autor concluye que los modelos autorregresivos fraccionalmente integrados de medias móviles (ARFIMA) resultan ser los procedimientos más adecuados durante pronósticos de un paso hacia delante (*one-step ahead*), mientras que, para el horizonte más largo (12 meses), redes neuronales logran resultados comparables con modelos ARFIMA. Puesto que ningún modelo es aceptado globalmente (en términos de RMSE y MAE), el autor recomienda tener en cuenta tanto el horizonte de pronóstico como la ubicación geográfica de la serie durante la selección de un enfoque.

A nivel latinoamericano, Rodríguez-Vargas (2020) pronostica la tasa de variación interanual del índice de precios al consumidor en Costa Rica con cuatro modelos de aprendizaje estadístico (RF, k-NN, XGB y LSTM) y contrasta sus resultados con el promedio de las predicciones de un conjunto de métodos econométricos univariantes. De tal forma, el autor encuentra que la combinación de pronósticos generados por los métodos de aprendizaje de máquina mejora la precisión al compararla con las predicciones individuales de cada uno de ellos y, al mismo tiempo, supera a los resultados producidos por los métodos univariados tradicionales.

Araujo & Gaglianone (2020), por otro lado, comparan los pronósticos de la tasa de inflación en Brasil para 16 modelos distintos—de los que sobresalen ARMA y VAR en la categoría de modelos econométricos y bosques aleatorios (RF) y modelos lineales generalizados (GLM) en la categoría de aprendizaje de máquina. Los autores encuentran que algunos algoritmos de aprendizaje automático superan consistentemente a los métodos tradicionales en términos de error cuadrático medio (MSE).

Por último, Peirano et al. (2021) evalúan los resultados de modelos SARIMA y LSTM durante la predicción del ritmo inflacionario mensual en Brasil, México, Chile, Colombia y Perú. Los autores concluyen que es posible alcanzar un mejor desempeño en el pronóstico al ensamblar ambos modelos, por lo que proponen un enfoque híbrido SARIMA-LSTM con el propósito de conservar las bondades de los modelos SARIMA, sin perder la posibilidad de capturar el efecto de fenómenos no-lineales por vía de una red neuronal recurrente de larga memoria de corto plazo. Sin embargo, observan los autores, la metodología propuesta únicamente mejora la eficiencia de las predicciones en algunos casos.

3 Datos y metodología

3.1 Análisis de datos

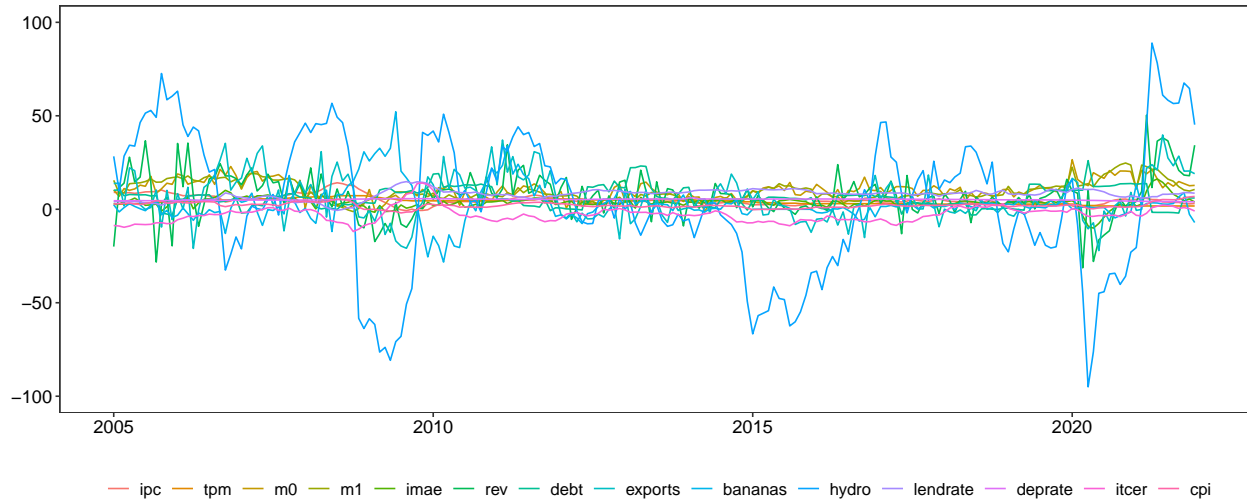
Con el objeto de pronosticar la tasa de variación interanual del índice de precios al consumidor en Guatemala, la especificación de los regresores ha sido propuesta a partir de un procedimiento en el que a 36 series de tiempo—seleccionadas considerando el canal de demanda agregada, el canal de tipo de cambio y las variables del sector externo del mecanismo de transmisión de política monetaria del país (Castillo, 2014)—les son aplicadas pruebas de causalidad de Wiener-Granger respecto a la variable objetivo para cada uno de los rezagos de mediano plazo y en el que, posteriormente, son filtradas tomando en cuenta (a) la cantidad de retardos que causan en el sentido de Granger al ritmo inflacionario y (b) la magnitud en el nivel de significancia del estadístico F de cada una de las pruebas del *test* previamente mencionado.

El cuadro 1 muestra a las 13 variables cuyos valores pasados son los más útiles, dado el vector inicial de posibles predictores, para pronosticar los valores futuros de la variación en el nivel de precios de la economía. Los datos corresponden al periodo que abarca de enero del 2005 a diciembre del 2021.

Cuadro 1: Variables empleadas durante el proceso de predicción

Variable	Descripción	Fuente
ipc	Tasa de variación interanual del índice de precios al consumidor	SECMCA
tpm	Tasa de interés de política monetaria	SECMCA
m0	Tasa de variación interanual de la base monetaria restringida	SECMCA
m1	Tasa de variación interanual del medio circulante	SECMCA
imae	Tasa de variación interanual del índice mensual de actividad económica	SECMCA
rev	Tasa de variación interanual de los ingresos totales del gobierno central	BANGUAT
debt	Tasa de variación interanual del saldo de la deuda pública externa	BANGUAT
exports	Tasa de variación interanual del valor (FOB) de las exportaciones totales	SECMCA
bananas	Tasa de variación interanual del valor promedio (dólares por tonelada) de las exportaciones de banano	SECMCA
hydro	Tasa de variación interanual del valor promedio (dólares por barril) de las importaciones de hidrocarburos	SECMCA
lendrate	Tasa de interés activa (real) en moneda nacional	SECMCA
deprate	Tasa de interés pasiva (nominal) en moneda nacional	SECMCA
itcer	Tasa de variación interanual del índice de tipo de cambio efectivo real (global)	SECMCA
cpi	Tasa de variación interanual del índice de precios al consumidor en Estados Unidos	FRED

Figura 1: Variable objetivo y predictores seleccionados



El cuadro 2 presenta estadísticas descriptivas para cada una de las variables. En ellas se incluyen el primer, tercer y cuarto momentos de sus funciones de densidad de probabilidad, así como la raíz cuadrada del segundo—para evidenciar la volatilidad de las series. En este último sentido, es posible observar que las tasas de cambio interanual del valor promedio de las importaciones de hidrocarburos, las exportaciones totales y los ingresos del gobierno central son los predictores que exhiben la mayor volatilidad durante el periodo del análisis.

Las estadísticas descriptivas también demuestran que esta investigación lidia mayoritariamente con variables que no se distribuyen normalmente, de las cuales 9 presentan distribuciones leptocúrticas, 2 platicúrticas y 3 una distribución aproximadamente normal, en términos de exceso de curtosis. Además, el estadístico de asimetría comprueba que 4 series se encuentran sesgadas hacia la izquierda y el resto de ellas hacia la derecha (exceptuando a la tasa de variación de los ingresos del gobierno central, cuya asimetría es cercana a 0).

Cuadro 2: Estadísticas descriptivas

Variable	Media	Mediana	Mínimo	Máximo	Desv. Est.	Asimetría	Curtosis	Obs.
ipc	4.92	4.36	-0.73	14.16	2.64	0.99	1.55	204
tpm	4.05	4.50	1.75	7.25	1.44	0.21	-0.77	204
m0	10.37	10.09	-1.32	26.52	5.04	0.30	0.05	204
m1	9.89	8.36	-1.25	24.74	5.13	0.76	0.14	204
imae	3.60	3.76	-10.10	16.21	2.94	-0.24	6.94	204
rev	7.35	7.14	-31.33	50.37	11.64	0.09	1.95	204
debt	6.35	6.14	-3.19	26.00	5.47	0.66	0.69	204
exports	5.85	3.65	-22.12	39.77	12.03	0.38	-0.08	204
bananas	2.69	1.27	-28.25	52.15	9.70	1.22	4.95	204
hydro	3.86	4.29	-94.99	88.91	36.00	-0.33	-0.36	204
lendrate	7.90	8.34	-0.59	14.79	2.72	-0.54	0.85	204
deprate	5.14	5.26	3.99	6.07	0.42	-0.81	0.05	204
itcer	-2.83	-3.11	-11.93	14.04	4.04	1.16	3.20	204
cpi	2.12	1.99	-1.98	6.86	1.46	0.31	0.73	204

Cuadro 3: Pruebas estadísticas

Variable	Autocorrelación	Normalidad	No-linealidad	Raíz unitaria	Hurst
ipc	850.5 (0.0000)	55.55 (0.0000)	0.78 (0.4611)	-4.52 (0.0100)	0.7883
tpm	1887.78 (0.0000)	6.21 (0.0448)	0.46 (0.6350)	-3.94 (0.0133)	0.8425
m0	607.64 (0.0000)	3.18 (0.2036)	0.87 (0.4223)	-2.21 (0.4863)	0.7836
m1	1024.67 (0.0000)	20.03 (0.0000)	2.2 (0.1130)	-2.46 (0.3823)	0.8171
imae	326.83 (0.0000)	422.88 (0.0000)	5.68 (0.0040)	-3.82 (0.0190)	0.6823
rev	125.43 (0.0000)	34.34 (0.0000)	1.59 (0.2069)	-3.52 (0.0423)	0.6729
debt	363.96 (0.0000)	19.53 (0.0001)	6.06 (0.0028)	-3.72 (0.0242)	0.6636
exports	272.33 (0.0000)	4.97 (0.0831)	3.25 (0.0408)	-3.54 (0.0402)	0.7240
bananas	568.35 (0.0000)	265.96 (0.0000)	4.38 (0.0137)	-5.68 (0.0100)	0.6513
hydro	576.63 (0.0000)	4.61 (0.0998)	0.07 (0.9365)	-3.69 (0.0266)	0.7236
lendrate	881.26 (0.0000)	16.9 (0.0002)	0.55 (0.5790)	-4.17 (0.0100)	0.7952
deprate	1566.44 (0.0000)	22.71 (0.0000)	12.7 (0.0000)	0.07 (0.9900)	0.8337
itcer	648.31 (0.0000)	136.98 (0.0000)	0.25 (0.7801)	-4.9 (0.0100)	0.6709
cpi	561.14 (0.0000)	8.36 (0.0153)	0.37 (0.6916)	-2.22 (0.4819)	0.7720

* Nivel de significancia entre paréntesis.

El cuadro 3 muestra un panel de pruebas estadísticas compuesto por un *test* de Ljung–Box para autocorrelación, uno de Jarque–Bera para normalidad, una prueba de red neuronal de White para comprobar si las variables exhiben comportamientos no-lineales, una prueba de Dickey–Fuller aumentada para verificar si existen raíces unitarias y la estimación del exponente de Hurst para medir la memoria de largo plazo de las series.

Como es de esperar, todas las variables exhiben correlación serial y—tal como evidencia el cuadro de estadísticas descriptivas—la mayor parte de estas no se distribuyen normalmente. La prueba de White confirma que 5 de las series (la tasa de interés pasiva nominal y las tasas de variación del índice mensual de actividad económica, del saldo de la deuda pública externa, de las exportaciones totales y del valor promedio de las exportaciones de banano) presentan un comportamiento no-lineal.

Además, los resultados de la prueba de raíz unitaria evidencian que la tasa de interés pasiva nominal y las tasas de cambio interanual de la base monetaria restringida, del medio circulante y del índice de precios al consumidor en Estados Unidos no son estacionarias. Por último, el exponente de Hurst demuestra que todas las series exhiben memoria de largo plazo.

3.2 Métodos empíricos

3.2.1 Modelos de series de tiempo

En la categoría de modelos tradicionales, este estudio considera un modelo de caminata aleatoria estacional (*snaiïve*), un modelo de suavizamiento exponencial (*Holt-Winters*), uno autorregresivo (AR), uno autorregresivo integrado de medias móviles estacionales (SARIMA) y un modelo de vectores autorregresivos (VAR). El primero de estos pronostica los valores futuros de una serie de tiempo asumiendo que cada predicción h es igual al valor observado del mismo mes durante el año anterior, por lo que sus pronósticos tienden a ser relativamente eficientes cuando la variable cuenta con un marcado componente estacional.

Mientras tanto, la metodología detrás de los diferentes sabores de modelos de suavizamiento exponencial consiste en predecir a los valores futuros de una serie a través de un promedio ponderado en el que los pesos de los ponderadores decrecen exponencialmente a medida que los valores observados se alejan en el tiempo. El método *Holt-Winters* extiende a este tipo de modelos con el propósito de capturar al componente estacional de la variable. De tal forma, la especificación del modelo comprende tres ecuaciones de suavizamiento: una para dar cuenta del nivel, otra para considerar la tendencia de la serie y la última para modelar su estacionalidad.

Los modelos autorregresivos utilizan una combinación lineal de los valores pasados de la variable objetivo para predecir su comportamiento en el futuro. De tal forma, una autorregresión de orden p es en esencia una regresión lineal múltiple en la que los predictores son los p valores rezagados de la variable que se desea pronosticar. Un modelo ARIMA combina a este último método con modelos de medias móviles y procedimientos de diferenciación, por lo que es capaz de lidiar con una serie no estacionaria al obtener su primera diferencia d veces (induciendo consistencia temporal en su media y su varianza) antes de modelarla por medio de un proceso ARMA (p, q) . Un modelo SARIMA extiende este procedimiento para poder considerar una alta gama de comportamientos estacionales al incluir términos adicionales que implican desplazamientos hacia atrás del periodo estacional y que luego son multiplicados por los términos (p, d, q) del modelo ARIMA convencional.

Por último, un modelo de vectores autorregresivos es un método multivariante en el que no se impone una relación causal unidireccional entre la variable de interés y sus predictores. Durante este procedimiento todas las variables son tratadas simétricamente al ser modeladas como si cada una tuviera una influencia sobre el resto de variables incluidas en el sistema. Es decir, todas las variables son consideradas endógenas. Un modelo VAR es en esencia una generalización de un proceso autorregresivo aplicado a un vector de series de tiempo.

3.2.2 Modelos de aprendizaje estadístico

3.3 Validación de modelos y calibración de hiperparámetros

La validación empírica de los modelos es una preocupación constante en la literatura de *machine learning*. Tal y como observan Kuhn & Silge (2022), debido a que la implementación de estos métodos conlleva una serie de pasos (estimación de parámetros, calibración de hiperparámetros, selección del modelo, evaluación de su desempeño) y, al mismo tiempo, una muestra finita de observaciones, es una práctica común y recomendada dividir al conjunto de datos existente en un subconjunto de observaciones de entrenamiento (el cual es utilizado para optimizar al modelo) y un subconjunto de datos de prueba (que se mantiene en reserva hasta que el método que tiene más probabilidades de éxito sea finalmente escogido).

Sin embargo, el investigador necesita entender qué tan efectivo es su modelo antes de poder emplear el conjunto de observaciones de evaluación. En este sentido, Gareth et al. (2021) mencionan que los métodos de remuestreo se han convertido en una herramienta esencial durante la aplicación de técnicas modernas de aprendizaje estadístico. Básicamente, estos procedimientos consisten en la extracción repetida de muestras aleatorias al conjunto de datos de entrenamiento con el fin de reajustar al modelo en cada una de ellas para obtener información adicional sobre el modelo en cuestión.

Ahora bien, cuando los datos contienen un fuerte elemento temporal, los métodos usuales de remuestreo tales como validación cruzada y *bootstrapping* no son factibles pues el orden de las observaciones tiene que ser preservado para que el modelo pueda aprender los distintos patrones cíclicos y

tendenciales que subyacen en la información: una versión más sofisticada de métodos de remuestreo es necesaria.

En este estudio la metodología de remuestreo empleada será *rolling forecast origin resampling*, un procedimiento en el que el conjunto de datos de entrenamiento es nuevamente dividido en dos subconjuntos de análisis/evaluación con tamaños específicos, de tal forma que una primera iteración utiliza estos tamaños comenzando desde el inicio de la serie. La segunda iteración utiliza los mismos tamaños pero se desplaza por un número determinado de muestras. El proceso continúa hasta que el conjunto de datos de entrenamiento es agotado por completo, por lo que el origen sobre el cual se basan los pronósticos va avanzando en el tiempo (Hyndman & Athanasopoulos, 2018).

Dicho método de remuestreo es utilizado específicamente con el objetivo de calibrar los hiperparámetros de los modelos de aprendizaje estadístico (en promedio, 3 hiperparámetros diferentes son ajustados por modelo). Para esto—y en línea con las metodologías de diseño de experimentos computacionales (Santner et al., 2018)—se construye un *space-filling parameter grid* a través de un método llamado *latin hypercube*, que es en esencia un procedimiento estadístico con el que se selecciona, dado un espacio de parámetros, una muestra aleatoria de los mismos.

De tal forma, con 100 combinaciones distintas de hiperparámetros y 6 conjuntos de remuestreo, 600 modelos son estimados y aquella combinación de parámetros que minimiza el error absoluto medio (MAE) en los subconjuntos de evaluación es seleccionada para predecir el ritmo inflacionario durante cada uno de los horizontes de pronóstico, de modo que en total se estiman 2400 modelos por cada método de aprendizaje de máquina (exceptuando a la red neuronal de larga memoria de corto plazo).

3.4 Medidas de precisión

Debido a que existe una amplia gama de medidas de precisión que pueden ser catalogadas en siete grupos distintos (Shcherbakov et al., 2013), es importante considerar que la clasificación relativa del resultado de los diversos métodos que se evalúan puede variar según la medida que se utilice (Makridakis & Hibon, 2000). Además, cada medida tiene desventajas que pueden conducir a una evaluación inexacta de los resultados, por lo que—así como Mathews & Diamantopoulos (1994) señalan—ninguna medida por sí sola brinda una pauta inequívoca de la eficiencia en las predicciones.

Con el fin de evaluar las diferencias en la precisión de los pronósticos entre métodos tradicionales y de aprendizaje estadístico, este estudio utilizará 3 medidas distintas. La primera de ellas, el error absoluto medio (MAE), es seleccionada no solo por ser una opción ampliamente utilizada al comparar distintos métodos que predicen un solo conjunto de datos, sino por contar con la ventaja de presentar una mayor robustez a valores atípicos que otras medidas dependientes de escala como la raíz del error cuadrático medio (RMSE) (Shcherbakov et al., 2013). Esta medida se define como,

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_t - \hat{y}_t|.$$

La segunda medida a considerar será el error absoluto medio relativo (rMAE), que es calculado como el ratio entre el MAE del método i que se desea evaluar y el MAE_{rw} de un modelo de caminata aleatoria, utilizado como punto de referencia. Hyndman & Koehler (2006) observan que una de las ventajas de esta medida es su interpretabilidad: Un $rMAE > 1$ significa que el método propuesto se desempeña peor que lo que lo hace el modelo de referencia, mientras que un $rMAE < 1$ denota lo opuesto. Formalmente,

$$rMAE = \frac{MAE_i}{MAE_{rw}}.$$

La última medida que será considerada en la presente investigación es el error medio absoluto escalado (MASE), un método propuesto por primera vez en Hyndman & Koehler (2006) con el objeto de brindar una mejor opción que el error porcentual absoluto medio simétrico (sMAPE). El error medio absoluto escalado se encuentra definido de la siguiente manera:

$$MASE = \frac{1}{k} \frac{\sum_{t=1}^k |y_t - \hat{y}_t|}{\frac{1}{n-m} \sum_{t=1}^k |y_t - \hat{y}_{t-m}|},$$

donde n es el número de observaciones históricas disponibles y m es la frecuencia de la serie de tiempo.

La interpretación de esta medida es similar a la interpretación del rMAE: un $MASE < 1$ indica que los pronósticos del modelo evaluado son, en promedio, más exactos que los pronósticos de un modelo de referencia (Makridakis et al., 2018). En su implementación original, esta medida utiliza el error absoluto medio de un modelo de caminata aleatoria dentro de la muestra para calcular los errores escalados. Kuhn & Vaughan (2021) observan que el procedimiento es llevado a cabo de tal modo (en lugar de considerar el error fuera de muestra) ya que existe la posibilidad de que el error fuera de muestra no pueda ser calculado al pronosticar un periodo hacia delante. No obstante, la implementación del MASE en R solo conoce los valores genuinos y estimados fuera de muestra, por lo que es este último error el que se utiliza durante el cómputo de esta medida de precisión.

4 Resultados

Esta investigación compara la precisión en los pronósticos durante 4 horizontes temporales entre métodos econométricos comúnmente utilizados y diversas alternativas basadas en modelos de aprendizaje estadístico. Como ha sido mencionado anteriormente, un modelo de caminata aleatoria es empleado como línea base con la que el resto de los métodos son evaluados.

El cuadro 4 muestra los resultados por modelo y horizonte de pronóstico en términos del error absoluto medio (MAE) y el error absoluto medio relativo (rMAE). En principio, es posible apreciar que todos los métodos se desempeñan mejor de lo que lo hace el modelo de referencia. Sin embargo, existe heterogeneidad en cuanto a la forma en la que cada uno se comporta durante la predicción de la variable objetivo en los distintos intervalos de tiempo.

Cuadro 4: MAE y rMAE para cada horizonte de pronóstico y modelo

	MAE				rMAE			
	h = 3	h = 9	h = 12	h = 24	h = 3	h = 9	h = 12	h = 24
snaïve	2.23	1.97	2.50	1.64	0.65	0.58	0.73	0.48
holt-winters	1.01	2.95	1.44	1.34	0.30	0.87	0.42	0.39
ar	0.89	1.74	1.16	1.33	0.26	0.51	0.34	0.39
sarima	0.27	0.31	0.45	1.28	0.08	0.09	0.13	0.38
var	0.36	0.92	2.20	1.13	0.11	0.27	0.65	0.33
ts ensemble	0.70	1.09	1.11	0.95	0.21	0.32	0.33	0.28
svm	0.77	0.95	0.67	1.19	0.23	0.28	0.20	0.35
rf	0.77	0.71	0.61	0.58	0.23	0.21	0.18	0.17
xgb	0.93	0.80	0.75	0.57	0.27	0.23	0.22	0.17
mlp	1.00	1.07	0.70	1.07	0.29	0.31	0.21	0.31
lstm	0.79	1.73	0.80	1.23	0.23	0.51	0.23	0.36
ml ensemble	0.84	0.90	0.53	0.60	0.25	0.26	0.16	0.18

En lo que respecta a los métodos de series temporales, el modelo de caminata aleatoria estacional es comparativamente el más ineficiente. Mientras tanto, el modelo de suavizamiento exponencial con estacionalidad aditiva y el modelo autorregresivo de segundo orden tienen un desempeño similar a lo largo de todos los horizontes temporales, aunque este último es más eficiente en cada uno de ellos. Además, el modelo SARIMA y el modelo VAR—cuya especificación utiliza únicamente tres variables: las tasas de variación interanual del IPC y del IMAE, así como la segunda diferencia del medio circulante, cada una de ellas rezagada por 24 periodos—son los métodos que, en general, cuentan con la eficiencia más alta durante los pronósticos de 3 meses hacia delante. De hecho, el excelente desempeño del modelo SARIMA se extiende hasta el horizonte temporal de 12 periodos.

Por otro lado, los pronósticos de la regresión de vectores de soporte (SVM) superan en todos los horizontes a los de los modelos econométricos más simples. Sin embargo, lo mismo no ocurre cuando su desempeño es comparado con el del modelo SARIMA (al que únicamente supera durante su predicción de 24 meses hacia delante) o con los pronósticos de 3 y 9 periodos hacia delante del modelo de vectores autorregresivos. Una situación similar acontece con los modelos de bosques aleatorios (RF) y de potenciación de gradiente extremo (XGB), los que, en la categoría de métodos de aprendizaje automático, presentan la mayor eficiencia de pronóstico—siendo RF el modelo con el mejor desempeño. Es importante mencionar que ambos modelos son extremadamente eficientes en lo que respecta al horizonte temporal de 24 meses.

Los dos modelos de redes neuronales (el perceptrón multicapa y la red neuronal recurrente de larga memoria de corto plazo, la cual, a diferencia del resto de modelos de aprendizaje estadístico en este estudio, utiliza como predictor únicamente el ritmo inflacionario rezagado por 35 periodos) son los modelos de *machine learning* con el peor desempeño en términos de error absoluto medio. Si bien ambos modelos superan consistentemente a los métodos econométricos más simples, lo contrario sucede al contrastar sus resultados con los de los modelos SARIMA y VAR—específicamente durante los pronósticos de corto plazo—y con los del resto de modelos de aprendizaje de máquina. Aun así, al considerar únicamente los horizontes de 12 y 24 meses, las redes neuronales superan en algunos casos al resto de sus contrapartes.

El cuadro 4 también presenta los resultados de una combinación de pronósticos por cada método, mientras que la figura 3 muestra las series de tiempo de ambas combinaciones durante la predicción de 24 meses hacia delante. Estos ensambles se realizan a través de una media ponderada en la que el peso relativo de los ponderadores está dado por:

$$w_i = 1 - \frac{MAE_h}{\sum_{h=1}^H MAE},$$

donde MAE_h es el error absoluto medio del pronóstico en el periodo h de un modelo específico y H representa al horizonte de predicción. De tal forma, w_i pondera con un mayor peso a las predicciones de los modelos que tienen un menor MAE. Este método de combinación de pronósticos es considerado como una técnica dinámica cuya ventaja principal es abstraer mayor complejidad estadística en comparación con estimar un vector constante de ponderación y aplicarlo durante todo el horizonte de pronóstico (Wang et al., 2018).

En uno de los estudios seminales en la literatura de combinación de pronósticos, Bates & Granger (1969) mencionan que dado a que diferentes métodos utilizan distintos supuestos acerca de la forma en la que las variables se relacionan, combinaciones de los resultados de distintos modelos pueden mejorar la precisión en sus predicciones, aunque no es necesariamente el caso que tal resultado pueda obtenerse siempre. En esta investigación, la combinación de los pronósticos de modelos econométricos genera predicciones más eficientes en el horizonte de 24 meses. Mientras que el ensamble de pronósticos de modelos de aprendizaje automático produce una mejora en la eficiencia para el horizonte de 12 periodos hacia delante.

Figura 2: Pronósticos de 24 meses hacia delante

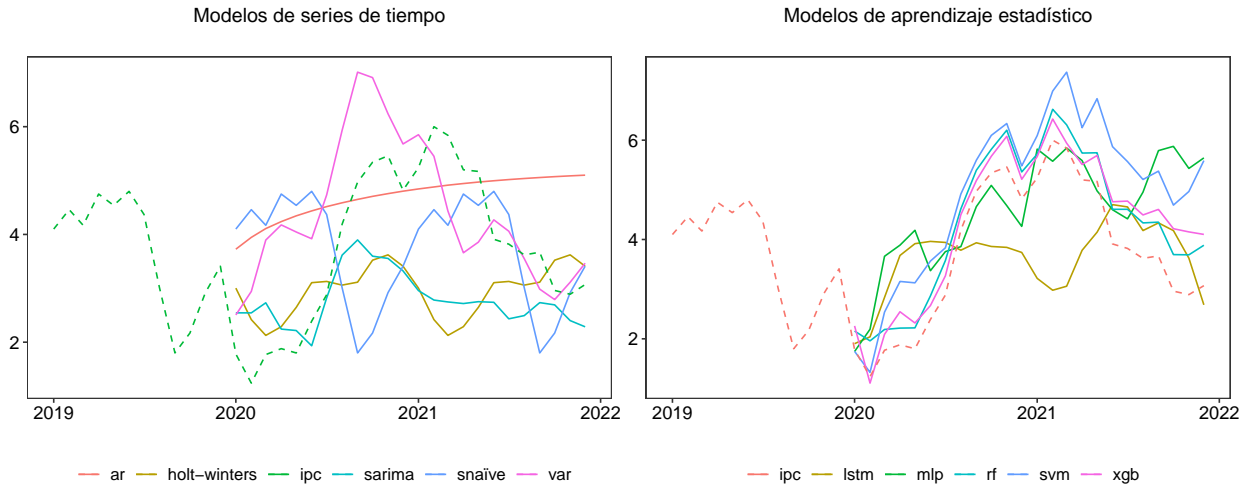
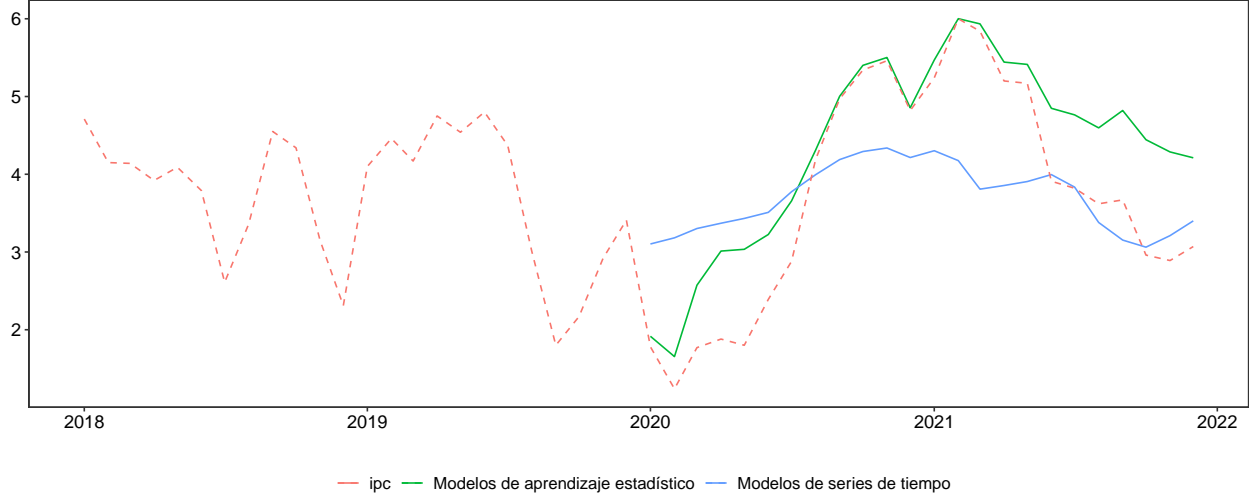


Figura 3: Combinación de pronósticos por clase de método



Como ha sido mencionado con anterioridad, el error absoluto medio escalado (MASE) contrasta los errores fuera de muestra entre un modelo de caminata aleatoria y el método que se busca evaluar. Si bien, en teoría un MASE menor que uno es la magnitud deseable del error en la predicción de un modelo particular durante el horizonte H , en la práctica esto rara vez ocurre, por lo que resulta problemático juzgar la precisión de los pronósticos sin considerar que los modelos pueden desempeñarse de formas diferentes según las características de los datos que se emplean.

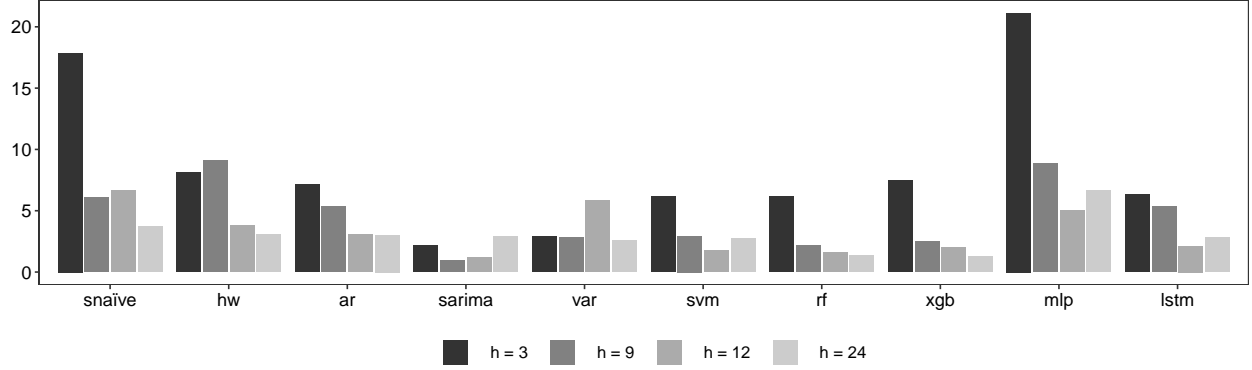
Por ejemplo, Athanasopoulos et al. (2011)—al comparar las predicciones de diversos métodos de suavizamiento exponencial con pronósticos de modelos ARIMA—utilizan como punto de referencia el promedio del MASE en el modelo ARIMA durante todos los horizontes temporales en lugar de la unidad. En línea con dicha investigación, el presente estudio utilizará el valor 8.58 (el MASE promedio del modelo de caminata aleatoria estacional durante los cuatro horizontes de pronóstico) como *benchmark* para el resto de modelos. Tanto el cuadro 5 como la figura 4 presentan los resultados obtenidos por cada uno de los métodos al ser evaluados por esta medida de precisión.

La comparación del MASE promedio durante todos los horizontes de pronóstico de los modelos de series de tiempo muestra que cada uno de ellos supera, en términos de eficiencia, al modelo de caminata aleatoria estacional. Además, exceptuando al pronóstico del modelo de suavizamiento exponencial durante el horizonte de 9 meses, la eficiencia en la predicción de cada uno de los periodos individuales se mantiene por debajo del promedio del modelo de referencia.

Cuadro 5: MASE por modelo y horizonte de pronóstico

	MASE									
	snaïve	holt-winters	ar	sarima	var	svm	rf	xgb	mlp	lstm
h = 3	17.87	8.11	7.13	2.17	2.90	6.16	6.17	7.47	21.12	6.29
h = 9	6.08	9.12	5.36	0.95	2.83	2.94	2.18	2.47	8.84	5.34
h = 12	6.63	3.82	3.08	1.19	5.83	1.78	1.61	1.99	5.04	2.12
h = 24	3.73	3.05	3.02	2.91	2.56	2.71	1.32	1.29	6.62	2.80
Media	8.58	6.02	4.65	1.80	3.53	3.40	2.82	3.31	10.41	4.14

Figura 4: MASE por modelo y horizonte de pronóstico

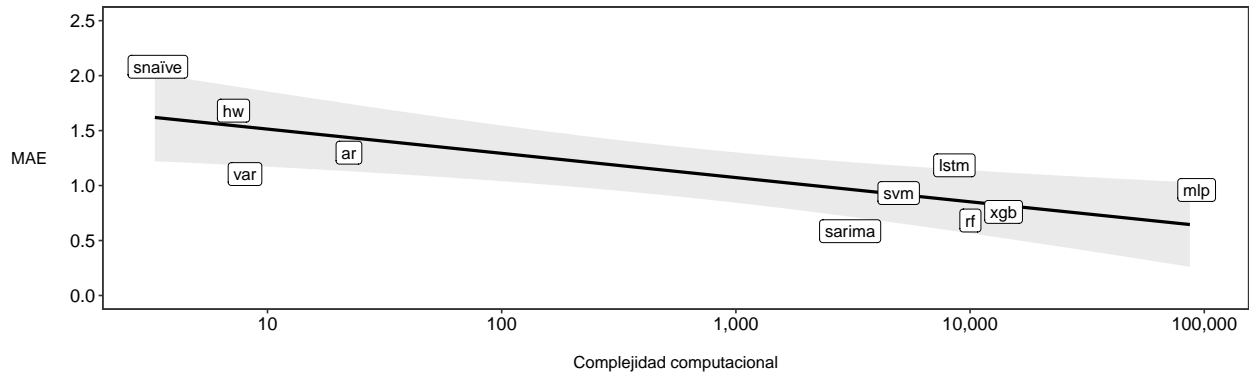


SARIMA vuelve a desempeñarse mejor que el resto de métodos, llegando durante el horizonte de 9 meses a superar al pronóstico del modelo ingenuo fuera de muestra (lo que lo convierte en el único modelo considerado en el estudio que consigue obtener un MASE menor que la unidad para algún horizonte de predicción). El modelo VAR logra resultados comparables, aunque aún lejanos a los del modelo SARIMA.

Mientras tanto, al considerar el desempeño de los diferentes métodos de aprendizaje estadístico, el perceptrón multicapa (MLP) es el único modelo cuyas predicciones, en promedio, no logran superar la eficiencia de pronóstico del modelo de caminata aleatoria estacional. Los resultados del resto de modelos son muy similares a los de las predicciones evaluadas a través de error absoluto medio. El modelo de bosques aleatorios, el de potenciación de gradiente extremo y la regresión de máquinas de vectores de soporte son los más eficientes en la categoría de *machine learning*, en ese orden. Sin embargo, mientras el error absoluto medio cataloga al promedio de las predicciones del perceptrón multicapa como pronósticos más eficientes que los de la red neuronal de larga memoria de corto plazo durante todos los horizontes de pronóstico, lo contrario ocurre al comparar sus resultados por medio del error medio absoluto escalado.

Por último, la figura 5 muestra la relación que existe entre la precisión de pronóstico evaluada a través del error absoluto medio y la complejidad computacional, a la que Makridakis et al. (2018) definen como $cc = S_i/S_{rw}$, donde el numerador representa la cantidad promedio de segundos que el modelo i tarda en generar sus predicciones para los cuatro horizontes de pronóstico considerados y el denominador denota la misma medida, pero para el modelo de referencia.

Figura 5: Relación entre precisión de pronóstico y complejidad computacional



5 Conclusiones

6 Referencias

- Araujo, G. S., & Gaglianone, W. P. (2020). *Machine learning methods for inflation forecasting in brazil: New contenders versus classical models*. Mimeo.
- Athanasopoulos, G., Hyndman, R. J., Song, H., & Wu, D. C. (2011). The tourism forecasting competition. *International Journal of Forecasting*, 27(3), 822–844. <https://doi.org/10.1016/j.ijforecast.2010.04.009>
- Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, 11(1), 685–725. <https://doi.org/10.1146/annurev-economics-080217-053433>
- Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Journal of the Operational Research Society*, 20(4), 451–468. <https://doi.org/10.1057/jors.1969.103>
- Castillo, C. (2014). Inflation targeting and exchange rate volatility smoothing: A two-target, two-instrument approach. *Economic Modelling*, 43, 330–345. <https://doi.org/10.1016/j.econmod.2014.08.011>
- Cerqueira, V., Torgo, L., & Soares, C. (2019). Machine learning vs statistical methods for time series forecasting: Size matters. *arXiv:1909.13316 [Cs, Stat]*. <http://arxiv.org/abs/1909.13316>
- Coulombe, P. G., Leroux, M., Stevanovic, D., & Surprenant, S. (2020). How is machine learning useful for macroeconomic forecasting? *arXiv:2008.12477 [Econ, Stat]*. <http://arxiv.org/abs/2008.12477>
- Diebold, F. X. (1998). The Past, Present, and Future of Macroeconomic Forecasting. *Journal of Economic Perspectives*, 12(2), 175–192. <https://doi.org/10.1257/jep.12.2.175>
- Foster, W. R., Collopy, F., & Ungar, L. H. (1992). Neural network forecasting of short, noisy time series. *Computers & Chemical Engineering*, 16(4), 293–297. [https://doi.org/10.1016/0098-1354\(92\)80049-F](https://doi.org/10.1016/0098-1354(92)80049-F)
- Gareth, J., Daniela, W., Trevor, H., & Robert, T. (2021). *An introduction to statistical learning: With applications in r*. Springer.
- Granger, C. W., Teräsvirta, T., & Anderson, H. M. (1993). Modeling nonlinearity over the business cycle. *Business Cycles, Indicators and Forecasting, NBER Chapters*, 311–326.
- Hill, T., O’Connor, M., & Remus, W. (1996). Neural Network Models for Time Series Forecasts. *Management Science*. <https://doi.org/10.1287/mnsc.42.7.1082>
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice*. OTexts.
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688. <https://doi.org/10.1016/j.ijforecast.2006.03.001>
- Kang, S. Y. (1991). *An investigation of the use of feedforward neural networks for forecasting* [PhD thesis]. Kent State University.
- Kantz, H., & Schreiber, T. (2004). *Nonlinear time series analysis* (Vol. 7). Cambridge university press.
- Katris, C. (2020). Prediction of Unemployment Rates with Time Series and Machine Learning Techniques. *Computational Economics*, 55(2), 673–706. <https://doi.org/10.1007/s10614-019-09908-9>
- Kuhn, M., & Silge, J. (2022). *Tidy modeling with r*. O’Reilly Media, Inc.
- Kuhn, M., & Vaughan, D. (2021). *Yardstick: Tidy characterizations of model performance*. <https://CRAN.R-project.org/package=yardstick>
- LeBaron, B. (1994). Chaos and nonlinear forecastability in economics and finance. *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences*. <https://doi.org/10.1098/rsta.1994.0099>
- Makridakis, S., & Hibon, M. (2000). The M3-Competition: results, conclusions and implications.

- International Journal of Forecasting*, 16(4), 451–476. [https://doi.org/10.1016/S0169-2070\(00\)00057-1](https://doi.org/10.1016/S0169-2070(00)00057-1)
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLOS ONE*, 13(3), e0194889. <https://doi.org/10.1371/journal.pone.0194889>
- Mathews, B. P., & Diamantopoulos, A. (1994). Towards a taxonomy of forecast error measures a factor-comparative investigation of forecast error dimensions. *Journal of Forecasting*, 13(4), 409–416. <https://doi.org/10.1002/for.3980130406>
- Moenjok, T. (2014). *Central banking: Theory and practice in sustaining monetary and financial stability*. John Wiley.
- Parmezan, A. R. S., Souza, V. M. A., & Batista, G. E. A. P. A. (2019). Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. *Information Sciences*, 484, 302–337. <https://doi.org/10.1016/j.ins.2019.01.076>
- Peirano, R., Kristjanpoller, W., & Minutolo, M. C. (2021). Forecasting inflation in Latin American countries using a SARIMA–LSTM combination. *Soft Computing*, 25(16), 10851–10862. <https://doi.org/10.1007/s00500-021-06016-5>
- Pratap, B., & Sengupta, S. (2019). *Macroeconomic Forecasting in India: Does Machine Learning Hold the Key to Better Forecasts?* <https://doi.org/10.2139/ssrn.3852945>
- Rodríguez-Vargas, A. (2020). Forecasting Costa Rican inflation with machine learning methods. *Latin American Journal of Central Banking*, 1(1), 100012. <https://doi.org/10.1016/j.latchb.2020.100012>
- Santner, T. J., Williams, B. J., & Notz, W. I. (2018). *Space-Filling Designs for Computer Experiments* (T. J. Santner, B. J. Williams, & W. I. Notz, Eds.; pp. 145–200). Springer. https://doi.org/10.1007/978-1-4939-8847-1_5
- Shcherbakov, M., Brebels, A., Shcherbakova, N. L., Tyukov, A., Janovsky, T. A., & Kamaev, V. A. (2013). A survey of forecast error measures. *World Applied Sciences Journal*, 24, 171–176. <https://doi.org/10.5829/idosi.wasj.2013.24.itmies.80032>
- Ülke, V., Sahin, A., & Subasi, A. (2018). A comparison of time series and machine learning models for inflation forecasting: empirical evidence from the USA. *Neural Computing and Applications*, 30(5), 1519–1527. <https://doi.org/10.1007/s00521-016-2766-x>
- Wang, L., Wang, Z., Qu, H., & Liu, S. (2018). Optimal Forecast Combination Based on Neural Networks for Time Series Forecasting. *Applied Soft Computing*, 66, 1–17. <https://doi.org/10.1016/j.asoc.2018.02.004>