

Grado en Estadística

Título: “Impacto del Covid-19 en el crecimiento del PIB de países de la OECD: Una aplicación de técnicas de pronóstico tradicionales y de *machine learning*”

Autor: Ivan Moreno Tamayo

Director: Ignacio Garrón Vedia

Departamento: Dept. Econometria, Estadística i Economia Aplicada

Convocatoria: Junio de 2023



UNIVERSITAT DE
BARCELONA



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Facultat de Matemàtiques i Estadística

Resumen y palabras clave

En este estudio se evalúa el impacto del Covid-19 en el crecimiento del PIB de los países miembros de la Organización para la Cooperación y el Desarrollo Económicos (OCDE) para el año 2020. Para este fin, primero, se realizan pronósticos del crecimiento del PIB mediante modelos tradicionales y de *machine learning*. Segundo, de acuerdo con el modelo más preciso para cada país se calcula una medida de impacto del Covid-19 sobre el crecimiento del PIB para 2020. Los resultados muestran en general que la elección del modelo de pronóstico depende del país y el horizonte temporal analizados; no obstante, en muchos casos los modelos no lineales como Random Forests y Redes Neuronales son los más adecuados, demostrando su capacidad para capturar patrones complejos y no lineales en los datos. La adición de estos modelos en muchos casos ha sido fundamental para comprender el impacto específico del Covid-19 en el crecimiento del PIB de cada país de la OCDE.

Palabras claves: Series temporales, modelización, horizonte, RMSE.

This study assesses the impact of Covid-19 on the GDP growth of member countries of the Organization for Economic Co-operation and Development (OECD) by 2020. To this end, first, GDP growth forecasts are made using traditional and machine learning models. Second, according to the most accurate model for each country, a measure of Covid-19 impact on GDP growth for 2020 is calculated. The results show in general that the choice of forecasting model depends on the country and time horizon analyzed; however, in many cases nonlinear models such as Random Forests and Neural Networks are the most suitable, demonstrating their ability to capture complex and nonlinear patterns in the data. The addition of these models in many cases has been instrumental in understanding the specific impact of Covid-19 on GDP growth in each OECD country.

Keywords: Time series, modeling, horizon, RMSE.

Clasificación AMS

- 91B84 Economic time series analysis

Índice general

| | |
|---|----|
| Clasificación AMS | 3 |
| Índice de cuadros | 5 |
| Índice de figuras | 6 |
| Cuerpo del trabajo..... | 7 |
| I. INTRODUCCIÓN | 7 |
| II. DATOS..... | 9 |
| 2.1 Tratamiento de los datos y software..... | 9 |
| III. CARACTERÍSTICAS TEMPORALES DEL CRECIMIENTO DEL PIB | 10 |
| 3.1. Análisis descriptivo | 11 |
| 3.2 Estudio de la media y la varianza móviles | 12 |
| 3.3. Cálculo y representación del ACF | 14 |
| 3.4. Test de raíz unitaria | 16 |
| IV. METODOLOGÍA | 18 |
| 4.1. Ecuación general | 18 |
| 4.2. Mecánica de predicción | 18 |
| 4.3. Modelos de predicción | 21 |
| 4.3 Evaluación de pronósticos..... | 32 |
| 4.5. Cálculo impacto covid-19 | 32 |
| V. RESULTADOS..... | 33 |
| 5.1 ETS | 34 |
| 5.2 ARIMA..... | 37 |
| 5.3 ARIMA con factor económico global (ARIMA_X) | 40 |
| 5.4 NNETAR | 43 |
| 5.5 NNETAR con factor económico global (NNETAR_X)..... | 45 |
| 5.6 RF | 47 |
| 5.7 RF con factor económico global (RF_X)..... | 48 |
| 5.8 Comparación de pronósticos por horizonte de predicción..... | 49 |
| 5.9 Impacto del covid-19 en el año 2020 | 51 |
| VI. Conclusiones | 54 |
| VII. Bibliografía..... | 55 |
| Apéndice 1: Salidas de R | 56 |
| Apéndice 2: Código de R | 67 |

Índice de cuadros

| | |
|--|----|
| Cuadro 3.1: Análisis descriptivo | 11 |
| Cuadro 3.4: Estadísticos del test de Dickey-Fuller Aumentado | 17 |
| Cuadro 4.3.1: Comparativa ETS aditivo versus multiplicativo..... | 21 |
| Cuadro 4.3.2: Comparativa modelos ETS y ARIMA | 25 |
| Cuadro 5.1.1: Pronóstico RMSE ETS | 34 |
| Cuadro 5.1.2: Modelos ETS por país..... | 35 |
| Cuadro 5.2.1: Pronóstico RMSE ARIMA..... | 37 |
| Cuadro 5.2.2: Elección modelo ARIMA | 38 |
| Cuadro 5.3.1: Pronóstico RMSE ARIMA_X | 40 |
| Cuadro 5.3.2: Elección modelo ARIMA_X | 41 |
| Cuadro 5.4.1: Pronóstico RMSE NNETAR | 43 |
| Cuadro 5.4.2: Elección modelo NNETAR | 44 |
| Cuadro 5.5.1: Pronóstico RMSE NNETAR_X | 45 |
| Cuadro 5.5.2: Elección modelo NNETAR_X | 46 |
| Cuadro 5.6.1: Pronóstico RMSE RF..... | 47 |
| Cuadro 5.7.1: Pronóstico RMSE RF_X..... | 48 |
| Cuadro 5.8.1: Tabla resumen selección de modelos según su RMSE | 49 |
| Cuadro 5.8.2: Mejores modelos según país y horizonte | 50 |
| Cuadro 5.9: Predicciones h1 a h4 + cálculo del impacto del covid-19 en el año 2020 | 51 |

Índice de figuras

| | |
|---|----|
| Figura 3.2: Análisis de la media y la varianza móviles | 12 |
| Figura 3.3: Representación de los ACF de los países de la OECD | 14 |
| Figura 4.2: Esquema de modelo de ventana móvil | 19 |
| Figura 4.3. Marco de ventana extensible versus móvil | 20 |
| Figura 4.3.3.1: Vista de la regresión de árbol..... | 27 |
| Figura 4.3.3.2: Matriz de confusión..... | 27 |
| Figura 4.3.4.1: Una red neuronal simple equivalente a una regresión lineal | 30 |
| Figura 4.3.4.2: Una red neuronal con 4 inputs y una capa oculta con tres neuronas ocultas | 31 |
| Figura 5.1.3: AICc modelo ETS..... | 36 |
| Figura 5.2.3: AICc modelo ARIMA | 39 |
| Figura 5.3.3: AICc modelo ARIMA_X | 42 |
| Figura A1.1: MSE modelo ETS | 66 |

Cuerpo del trabajo

I. INTRODUCCIÓN

La pandemia del Covid-19 ha tenido un impacto dramático en la economía mundial (ver, International Monetary Fund, 2020; World Health Organization, 2020; World Bank, 2020). El brote global de esta enfermedad en el año 2020 ha generado incertidumbre y desafíos significativos para los países, afectando diversos aspectos de la sociedad, incluido el Producto Interno Bruto (PIB). Por lo tanto, es crucial evaluar y comprender el impacto que esta crisis sanitaria ha tenido sobre el crecimiento económico de los países.

El crecimiento del PIB es un indicador fundamental para medir el desarrollo económico de una nación. Este indicador proporciona información valiosa para los responsables de la toma de decisiones y las organizaciones, permitiendo comparaciones internacionales, evaluación de riesgos y oportunidades, y orientación en la formulación de políticas económicas. En especial, contar con pronósticos del crecimiento del PIB en distintos países proporciona información clave a los hacedores de política y organizaciones sobre el crecimiento económico futuro, permite comparaciones internacionales, brinda indicadores de bienestar, orienta la formulación de políticas económicas y ayuda a evaluar riesgos y oportunidades. Esta información es fundamental para comprender las dinámicas económicas globales y tomar decisiones informadas en diferentes ámbitos.

En este contexto, este estudio tiene como objetivo evaluar el impacto del Covid-19 en el crecimiento del PIB de los países miembros de la Organización para la Cooperación y el Desarrollo Económicos (OCDE) para el año 2020. A través de diversas técnicas de pronóstico emplearemos técnicas de pronósticos tradicionales y de *machine learning*, se busca proporcionar de una medida de impacto del Covid-19 sobre el crecimiento del PIB. De esta manera, se puede tener una mejor comprensión de las dinámicas económicas y tomar decisiones informadas en un contexto global afectado por una crisis sin precedentes.

La estructura de este documento se compone de varios apartados. En la sección de datos, abordamos el PIB como variable clave de estudio, explicando cómo hemos tratado los datos y mencionando el software utilizado para su análisis. A continuación, en la sección de características temporales del crecimiento del PIB, examinamos las dinámicas y patrones de crecimiento económico en el contexto de los países de la OCDE. En la parte de metodología, describimos los modelos empleados, incluyendo el test de Dickey-Fuller aumentado para evaluar la estacionariedad de las series temporales y el uso del Error Cuadrático Medio (RMSE) como métrica de evaluación de los pronósticos. Posteriormente, aplicamos cuatro técnicas de pronóstico: ETS (*Error, Trend, Seasonality*), ARIMA (*Autoregressive Integrated Moving Average*), *Random Forests* y Redes Neuronales. Por último, elegimos los mejores modelos para cada país para calcular el efecto del COVID-19 sobre el crecimiento del PIB en 2020.

Los resultados principales de este trabajo consisten en la comparación de distintos modelos de predicción para cada país miembro de la OCDE, identificando cuál de ellos estima mejor las predicciones. Además, utilizamos estos modelos para calcular el impacto que la crisis del Covid-19 tuvo en la evolución del PIB de cada país. Los resultados sugieren que el modelo de pronóstico óptimo cambia según el país y el periodo temporal examinados; no obstante, en

muchos escenarios los modelos no lineales como Random Forests y Redes Neuronales son los más idóneos, evidenciando su destreza para reconocer patrones complejos y no lineales en los datos. La inclusión de estos modelos en muchos casos ha sido clave para comprender el impacto específico del Covid-19 en el crecimiento del PIB de cada país de la OCDE.

El resto del documento se organiza de la siguiente forma. La Sección II describe las series de datos utilizadas y la Sección III muestra las características temporales de las series de estudio. La Sección IV explica la metodología de pronóstico empleada y la Sección V analiza los hallazgos del estudio. Por último, la Sección VI resume las conclusiones principales.

II. DATOS

Los datos utilizados en este documento corresponden al PIB desestacionalizados de manera cuatrimestral de una muestra de países miembros de la OECD. Los datos que se han utilizado para poder llevar a cabo este estudio se han conseguido a través de la página web oficial de la OECD.¹ Además, los datos empleados fueron recogidos a fecha de 15 de marzo de 2023. Este dato es sumamente importante ya que, debido a la constante actualización de estos, resultará imposible reproducir con exactitud este análisis, ya que los valores serán ligeramente distintos.

El producto interno bruto (PIB) es la medida estándar del valor agregado creado a través de la producción de bienes y servicios en un país durante un período determinado. Como tal, también mide los ingresos obtenidos de esa producción, o la cantidad total gastada en bienes y servicios finales (menos las importaciones). Si bien el PIB es el indicador individual más importante para capturar la actividad económica, no proporciona una medida adecuada del bienestar material de las personas para la cual los indicadores alternativos pueden ser más apropiados. Este indicador se basa en el PIB real (también llamado PIB a precios constantes o PIB en volumen), es decir, la evolución a lo largo del tiempo se ajusta a los cambios de precios. Los números también se ajustan por influencias estacionales. El indicador está disponible en diferentes medidas: variación porcentual respecto al trimestre anterior, variación porcentual respecto al mismo trimestre del año anterior e índice de volumen (2015=100). Todos los países de la OCDE compilan sus datos de acuerdo con el Sistema de Cuentas Nacionales (SCN) de 2008.

2.1 Tratamiento de los datos y software

El estudio utiliza el software R Studio, 4.1.2 (2021-11-01) - "Bird Hippie", y el paquete tidyverse para el tratamiento de datos. Además, seleccionamos solamente los países que contienen valores completos del PIB sin missings y descartamos el resto. Así como eliminamos países que son colonias de otros (Gibraltar, por ejemplo).

¹ Se encuentra disponible en: <https://data.oecd.org/gdp/quarterly-gdp.htm#indicator-chart>.

III. CARACTERÍSTICAS TEMPORALES DEL CRECIMIENTO DEL PIB

En este apartado se discute si el crecimiento del PIB es estacionario en el sentido débil. En términos generales, una serie de tiempo estacionaria débil es aquella en la que las propiedades estadísticas de la serie no varían con el tiempo. En otras palabras, la media, la varianza y la autocovarianza de la serie son constantes a lo largo del tiempo.

En una serie de tiempo estacionaria débil, la media de la serie es constante en el tiempo. Esto implica que la serie no tiene una tendencia sistemática o un patrón de crecimiento o declive a largo plazo. La varianza también se mantiene constante, lo que significa que la amplitud de las fluctuaciones en la serie no cambia con el tiempo.

Además, en una serie de tiempo estacionaria débil, la autocovarianza entre dos puntos de tiempo solo depende de la diferencia entre esos dos puntos y no de los puntos de tiempo en sí. Esto se conoce como estacionariedad en sentido débil o estacionariedad débil. En otras palabras, la estructura de dependencia temporal de la serie se mantiene constante a lo largo del tiempo.

La estacionariedad débil es un supuesto importante en muchos modelos y técnicas de análisis de series de tiempo. Permite utilizar métodos estadísticos y modelos que asumen estacionariedad, como el análisis espectral, modelos ARMA (AutoRegressive Moving Average) y modelos de pronóstico basados en regresión lineal.

Es importante destacar que la estacionariedad débil es solo uno de los conceptos de estacionariedad en series de tiempo. Existe también la estacionariedad estricta, que implica que la distribución conjunta de cualquier conjunto de observaciones de la serie es invariante ante traslaciones en el tiempo. La estacionariedad débil es una condición menos restrictiva que la estacionariedad estricta.

A lo largo del estudio se utiliza como variable de estudio la transformación logarítmica del PIB:

$$y_t = \ln(PIB_t/PIB_{t-1}).$$

3.1. Análisis descriptivo

Podemos ver que el país que contiene el valor más bajo se trata de Francia con un valor mínimo de -0.1456, mientras que el máximo lo tiene la variable OECD, con un 0.090, seguido de Canadá con un valor de 0.0863 (Cuadro 3.1).

Cuadro 3.1: Análisis descriptivo

| <i>País</i> | <i>Mínimo</i> | <i>Q1</i> | <i>Mediana</i> | <i>Media</i> | <i>Q3</i> | <i>Máximo</i> |
|-------------|---------------|-----------|----------------|--------------|-----------|---------------|
| <i>AUS</i> | -0.069967 | 0.004048 | 0.007454 | 0.007397 | 0.011199 | 0.037827 |
| <i>CAN</i> | -0.115728 | 0.001672 | 0.006379 | 0.005434 | 0.010962 | 0.086342 |
| <i>FRA</i> | -0.145565 | 0.001282 | 0.004911 | 0.004128 | 0.007118 | 0.007118 |
| <i>ITA</i> | -0.129122 | -0.001252 | 0.003196 | 0.002329 | 0.006483 | 0.135050 |
| <i>KOR</i> | -0.070616 | 0.007206 | 0.014315 | 0.014708 | 0.021283 | 0.074209 |
| <i>NOR</i> | -0.056226 | -0.001231 | 0.006220 | 0.005694 | 0.012929 | 0.042360 |
| <i>SWE</i> | -0.085963 | 0.001383 | 0.006541 | 0.005231 | 0.011298 | 0.071156 |
| <i>CHE</i> | -0.0730785 | 0.0003708 | 0.0051884 | 0.0042111 | 0.0083389 | 0.0706754 |
| <i>USA</i> | -0.088660 | 0.003630 | 0.006949 | 0.006299 | 0.009975 | 0.075613 |
| <i>OECD</i> | -0.109020 | 0.003672 | 0.006078 | 0.005661 | 0.008665 | 0.090298 |

Fuente: Elaboración propia

3.2 Estudio de la media y la varianza móviles

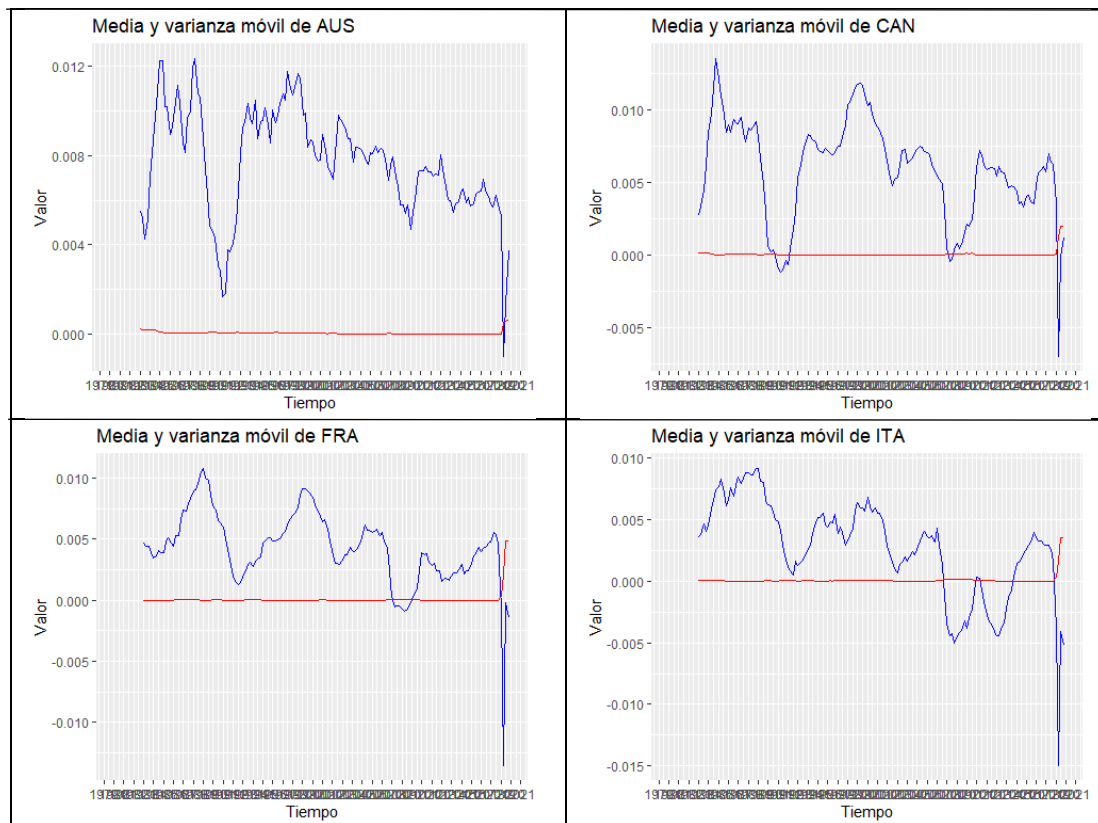
Para poder llevar a cabo un análisis con esta metodología, es necesario identificar qué componente de la serie se le tiene que aplicar una transformación. Es decir, se llevará a cabo un estudio de la media y varianza.

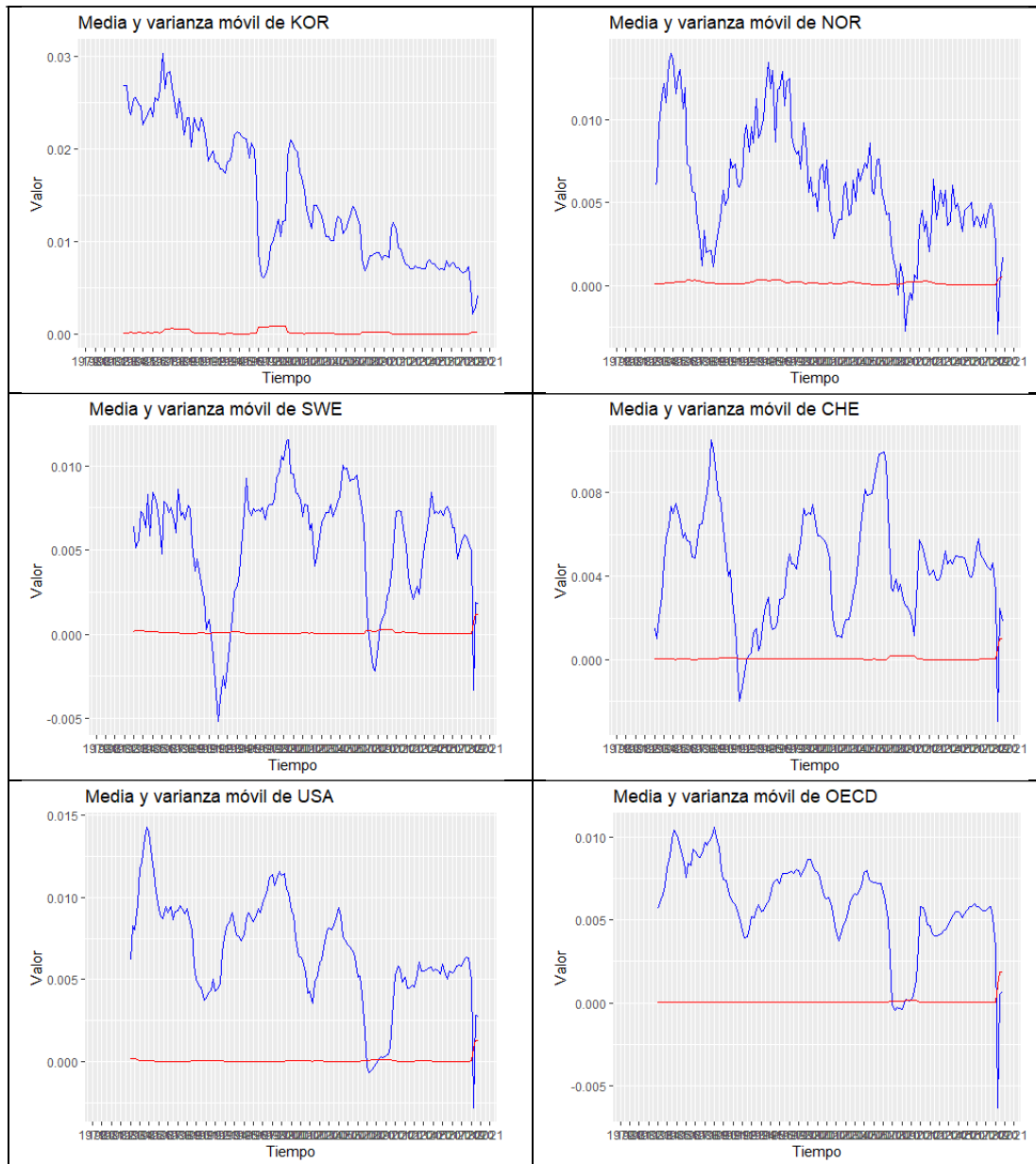
La Figura 3.2. muestra el análisis de la media y varianza para el crecimiento del PIB de los países considerados.

Para todos los casos, vemos que la media no se mantiene constante en el tiempo. Además, se pueden apreciar dos picos descendientes bastante pronunciados, que se corresponden con la crisis del 2008 y el covid-19.

En el caso de la varianza, vemos que esta sí se mantiene constante a lo largo del tiempo, a excepción del periodo de 2019 hasta la actualidad, seguramente causado por la ya mencionada crisis causada por el covid-19.

Figura 3.2: Análisis de la media y la varianza móviles



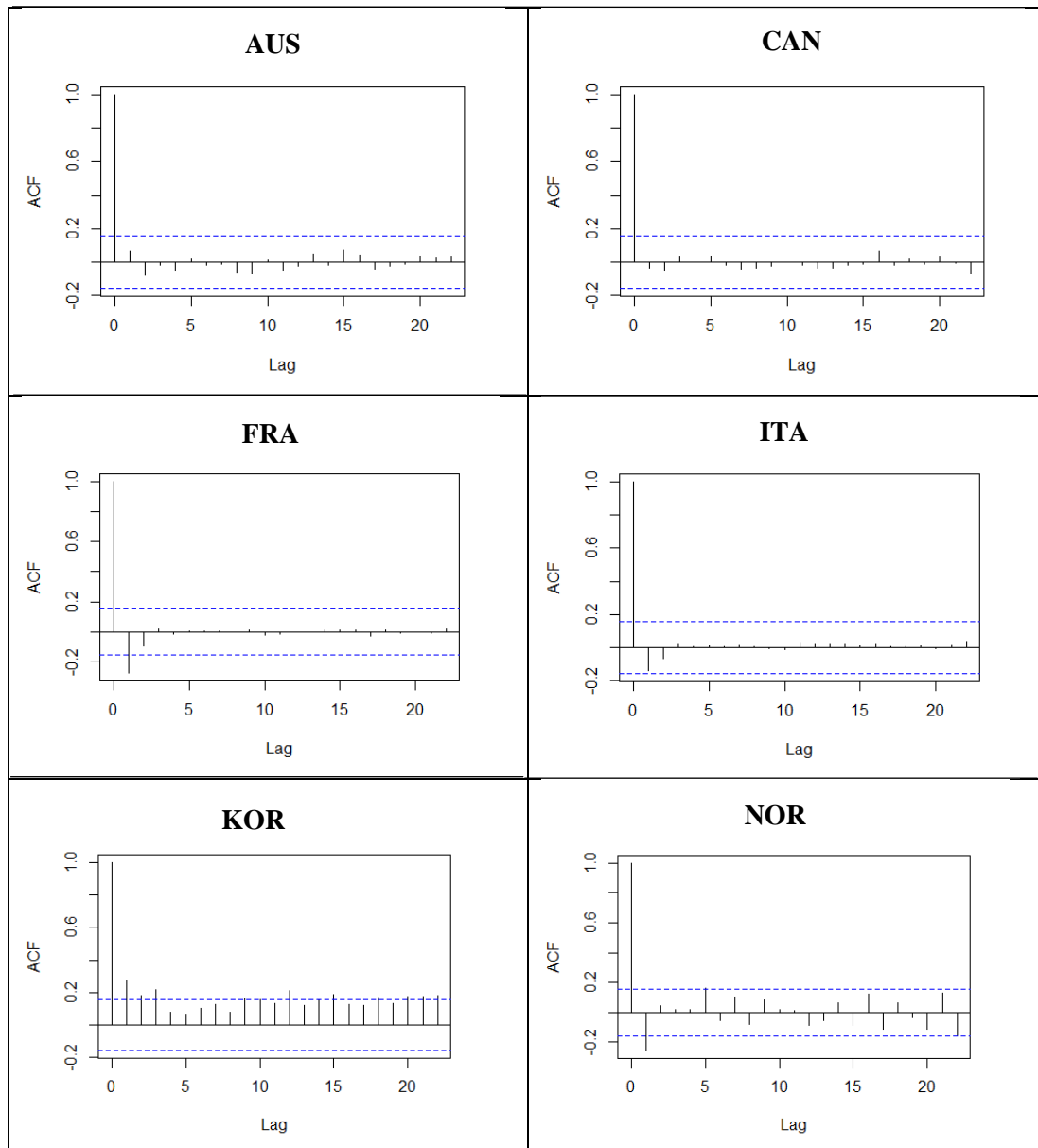


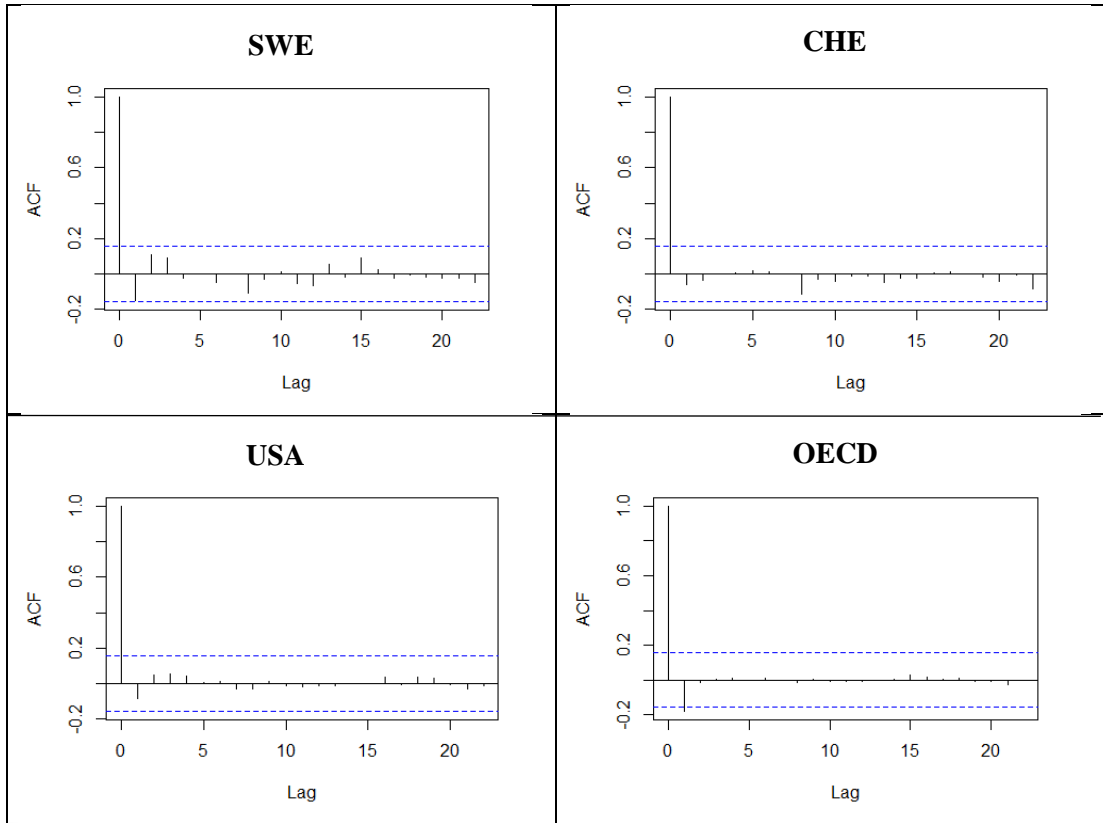
Fuente: Elaboración propia

3.3. Cálculo y representación del ACF

Mediante la Figura 3.3, se puede determinar si la serie realmente es estacionaria o no. En este caso, los ACF decaen rápidamente a cero. Por lo tanto, podemos determinar que las series son estacionarias. Pero para asegurarnos, realizaremos algún test para confirmar nuestras sospechas.

Figura 3.3: Representación de los ACF de los países de la OECD





Fuente: Elaboración propia

3.4. Test de raíz unitaria

Antes de realizar los modelos de pronóstico, tenemos que estudiar la estacionalidad de la serie. Para ello, realizaremos el test de raíz unitaria de Dickey-Fuller Augmentado.

El test de raíz unitaria de Dickey-Fuller Augmentado (ADF, por sus siglas en inglés) es una prueba estadística utilizada para determinar si una serie de tiempo tiene una raíz unitaria, lo que implica que la serie es no estacionaria. Fue desarrollado por los economistas David Dickey y Wayne Fuller en 1979 y posteriormente se amplió mediante la inclusión de términos rezagados o retardos en el modelo, dando lugar al ADF.

La hipótesis nula del test de ADF establece que la serie de tiempo tiene una raíz unitaria, es decir, es no estacionaria. La hipótesis alternativa, por otro lado, sugiere que la serie es estacionaria. La idea detrás del test es evaluar si el coeficiente de la variable rezagada es significativamente diferente de cero después de eliminar el efecto de los términos rezagados adicionales.

El procedimiento del test de ADF implica la estimación de un modelo de regresión lineal donde la variable dependiente es la serie de tiempo en cuestión y las variables independientes son la propia serie rezagada y sus retardos. Se utiliza un enfoque de mínimos cuadrados ordinarios para estimar los coeficientes de la regresión y se calcula un estadístico de prueba basado en la diferencia entre el valor estimado del coeficiente de la variable rezagada y su desviación estándar.

El estadístico de prueba se compara con valores críticos de la distribución de Dickey-Fuller para determinar si se rechaza o no la hipótesis nula. Si el estadístico de prueba es menor que el valor crítico, se puede rechazar la hipótesis nula y concluir que la serie de tiempo es estacionaria.

El test de ADF es ampliamente utilizado en el análisis de series de tiempo, especialmente en econometría y finanzas, para verificar la estacionariedad de una serie y determinar si se requieren transformaciones o modelos más adecuados para el análisis posterior. También se utiliza en el contexto de modelado de procesos estocásticos y pronóstico de series de tiempo.

Es importante tener en cuenta que el test de ADF tiene supuestos y limitaciones, y su interpretación debe realizarse en conjunto con un análisis más completo de la serie de tiempo y la consideración de otros factores relevantes.

Así pues, realizamos el test para cada país y obtenemos la siguiente tabla resumen:

Cuadro 3.4: Estadísticos del test de Dickey-Fuller Aumentado

| <i>País</i> | <i>Estadístico</i> |
|-------------|--------------------|
| <i>AUS</i> | -8.8112 |
| <i>CAN</i> | -8.6581 |
| <i>FRA</i> | -10.9066 |
| <i>ITA</i> | -9.2786 |
| <i>KOR</i> | -6.7974 |
| <i>NOR</i> | -9.9381 |
| <i>SWE</i> | -8.0303 |
| <i>CHE</i> | -8.7889 |
| <i>USA</i> | -7.9575 |
| <i>OECD</i> | -9.0686 |

Fuente: Elaboración propia

Como podemos ver, los estadísticos de todos los países son inferiores al valor crítico al 5% de -2.861. Como el estadístico de prueba es menor que el valor crítico, podemos rechazar la hipótesis nula de que las serie tienen una raíz unitaria. Esto sugiere que las series son estacionarias y tienen un orden de integración de 0, es decir, cada una es una serie I(0).

IV. METODOLOGÍA

4.1. Ecuación general

La ecuación general del modelo planteado para la variable pronosticada es la siguiente:

$$y_{t+1} = g(x_t) + e_t$$

Donde y_{t+1} es el crecimiento del PIB en logaritmos, x_t es un vector de covariantes que pueden ser rezagos de la variable dependiente, errores autorregresivos o un factor económico global. $G(\cdot)$ es una función que puede venir de un modelo lineal (ETS o ARIMA) o un modelo no lineal (RF y NN).

El pronóstico viene para $h=1$ dado por:

$$\widehat{y}_{t+1} = \widehat{g}(x_t)$$

Y para $h=2,3,4$ se considera un pronóstico iterativo, es decir que se toma en cuenta el valor previo de y_t para pronósticos hacia adelante.

4.2. Mecánica de predicción

El presente trabajo emplea una mecánica de *rolling window* para la modelización de las variables.

El modelo de predicción mediante la mecánica de "*rolling window*" es una técnica utilizada para realizar pronósticos en series temporales. También se conoce como "ventana deslizante" o "ventana móvil". Este enfoque implica ajustar repetidamente un modelo de pronóstico a medida que se avanza en el tiempo, utilizando solo los datos históricos disponibles hasta ese punto.

La mecánica de "*rolling window*" se basa en la idea de que los modelos de pronóstico pueden beneficiarse de una actualización periódica a medida que se dispone de nuevos datos. Esto permite que el modelo se adapte a los cambios en la serie temporal y mejore su precisión a medida que se obtienen observaciones adicionales.

El proceso básico de la mecánica de "*rolling window*" implica los siguientes pasos:

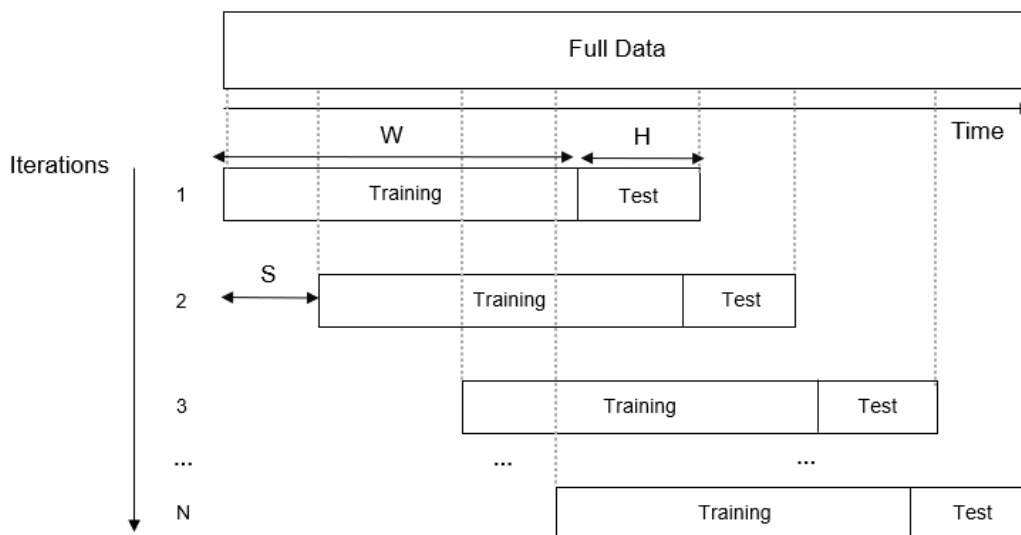
1. Seleccionar una ventana de tiempo inicial: Se elige un período inicial que defina el tamaño de la ventana deslizante. Por ejemplo, se puede seleccionar un año de datos iniciales.
2. Ajustar el modelo: Se ajusta un modelo de pronóstico a los datos dentro de la ventana inicial. Esto puede implicar el uso de técnicas como ARIMA, suavizado exponencial, regresión u otros modelos.
3. Generar el pronóstico: Se utiliza el modelo ajustado para generar un pronóstico para el siguiente período fuera de la ventana actual.

4. Actualizar la ventana: Se mueve la ventana deslizante hacia adelante, incluyendo el siguiente período de datos y excluyendo el período más antiguo.

5. Repetir los pasos 2-4: Se repiten los pasos de ajustar el modelo, generar el pronóstico y actualizar la ventana en cada iteración a medida que se avanza en el tiempo.

Este proceso se repite hasta que se haya pronosticado el período final deseado. A medida que se progresa en el tiempo, el modelo se recalibra continuamente para adaptarse a los cambios en los datos y mejorar la precisión del pronóstico.

Figura 4.2: Esquema de modelo de ventana móvil



Fuente Anselin et al. (2008).

La mecánica de "rolling window" permite evaluar el rendimiento del modelo de pronóstico a medida que se avanza en el tiempo, ya que se pueden comparar los pronósticos generados con los valores reales que se van revelando. Esto proporciona información valiosa sobre la calidad y precisión del modelo, lo que permite ajustes o mejoras si es necesario.

McCracken (2020) muestra por ejemplo, que el test Diebold y Mariano (DM) sigue una distribución normal estándar, cuando una variable aleatoria estacionaria en covarianza con autocovarianzas que decaen a cero bajo ciertas condiciones. En especial, el resultado anterior es válido, cuando los pronósticos se calculan en un marco de ventana móvil con longitud fija.

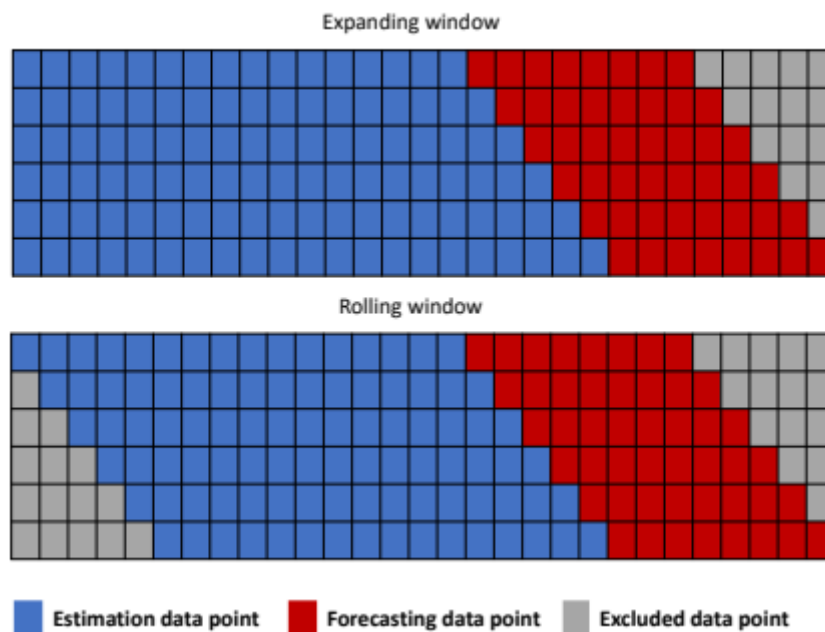
Por lo general, en la literatura de series de tiempo, los modelos de pronóstico se evalúan por su desempeño (pseudo) fuera de la muestra (OOS). Por lo tanto, la muestra se suele dividir en dos submuestras. Cuántas observaciones usamos para la estimación y cuántas dejamos para probar el modelo. El conjunto de observaciones utilizado para estimar los modelos de pronóstico suele denominarse ventana de estimación.

Expanding window: el pronosticador elige un tamaño de ventana inicial para estimar los modelos, digamos R . Después de la estimación, se calculan los pronósticos para $t = R + 1, \dots, R + h$. Cuando llega una nueva observación, el pronosticador incorpora los nuevos datos en la ventana de estimación, de modo que R se incrementa en una unidad. El proceso se repite hasta llegar al final de la muestra. Ver el panel superior en la Figura 4.1. En este caso, los modelos se estiman con un número creciente de observaciones a lo largo del tiempo.

Rolling window: el pronosticador elige un tamaño de ventana inicial para estimar los modelos, digamos R . Después de la estimación, se calculan los pronósticos para $R = R + 1, \dots, R + h$. Cuando llega una nueva observación, el primer punto de datos se elimina y el pronosticador incorpora la nueva observación en la ventana de estimación. Por lo tanto, el tamaño de la ventana se mantiene constante. Ver panel inferior en la Figura 4.3. En este caso, todos los modelos se estiman en una muestra con R observaciones.

La elección entre una ventana expandible o una ventana móvil no es necesariamente trivial. Por un lado, con ventanas que se expanden, el número de observaciones aumenta con el tiempo, lo que potencialmente genera estimadores más precisos de los modelos. Por otro lado, la estimación con ventanas en expansión es más susceptible a la presencia de rupturas estructurales y valores atípicos, por lo tanto, produce estimadores menos precisos. Los pronósticos basados en ventanas móviles están menos influenciados por rupturas estructurales y valores atípicos.

Figura 4.3. Marco de ventana extensible versus móvil



Fuente: Anselin et al. (2008).

4.3. Modelos de predicción

En esta sección se presentan los dos grupos de métodos de predicción utilizados. Modelos de regresión lineal: ETS y ARIMA; y Modelos de regresión no lineal: *Random Forest* y *Neural Networks*.

4.3.1 ETS

De acuerdo con la nomenclatura ETS, etiquetamos cada modelo de espacio de estado como ETS(\cdot, \cdot, \cdot) para (Error, Trend, Seasonal). Esta etiqueta también se puede considerar como suavizado exponencial. Las posibilidades para cada componente (o estado) son: Error = {A,M}, Tendencia = {N,A,Ad} y Estacionalidad = {N,A,M}.

Cuadro 4.3.1: Comparativa ETS aditivo versus multiplicativo

| ADDITIVE ERROR MODELS | | | |
|-----------------------------|--|--|--|
| Trend | Seasonal | | |
| | N | A | M |
| N | $y_t = \ell_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t$ | $y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$ | $y_t = \ell_{t-1} s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / \ell_{t-1}$ |
| A | $y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$ $b_t = b_{t-1} + \beta \varepsilon_t$ | $y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$ $b_t = b_{t-1} + \beta \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$ | $y_t = (\ell_{t-1} + b_{t-1}) s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $b_t = b_{t-1} + \beta \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} + b_{t-1})$ |
| Ad | $y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$ $b_t = \phi b_{t-1} + \beta \varepsilon_t$ | $y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$ $b_t = \phi b_{t-1} + \beta \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$ | $y_t = (\ell_{t-1} + \phi b_{t-1}) s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $b_t = \phi b_{t-1} + \beta \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} + \phi b_{t-1})$ |
| MULTIPLICATIVE ERROR MODELS | | | |
| Trend | Seasonal | | |
| | N | A | M |
| N | $y_t = \ell_{t-1} (1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} (1 + \alpha \varepsilon_t)$ | $y_t = (\ell_{t-1} + s_{t-m}) (1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \alpha (\ell_{t-1} + s_{t-m}) \varepsilon_t$ $s_t = s_{t-m} + \gamma (\ell_{t-1} + s_{t-m}) \varepsilon_t$ | $y_t = \ell_{t-1} s_{t-m} (1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} (1 + \alpha \varepsilon_t)$ $s_t = s_{t-m} (1 + \gamma \varepsilon_t)$ |
| A | $y_t = (\ell_{t-1} + b_{t-1}) (1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1}) (1 + \alpha \varepsilon_t)$ $b_t = b_{t-1} + \beta (\ell_{t-1} + b_{t-1}) \varepsilon_t$ | $y_t = (\ell_{t-1} + b_{t-1} + s_{t-m}) (1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha (\ell_{t-1} + b_{t-1} + s_{t-m}) \varepsilon_t$ $b_t = b_{t-1} + \beta (\ell_{t-1} + b_{t-1} + s_{t-m}) \varepsilon_t$ $s_t = s_{t-m} + \gamma (\ell_{t-1} + b_{t-1} + s_{t-m}) \varepsilon_t$ | $y_t = (\ell_{t-1} + b_{t-1}) s_{t-m} (1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1}) (1 + \alpha \varepsilon_t)$ $b_t = b_{t-1} + \beta (\ell_{t-1} + b_{t-1}) \varepsilon_t$ $s_t = s_{t-m} (1 + \gamma \varepsilon_t)$ |
| Ad | $y_t = (\ell_{t-1} + \phi b_{t-1}) (1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1}) (1 + \alpha \varepsilon_t)$ $b_t = \phi b_{t-1} + \beta (\ell_{t-1} + \phi b_{t-1}) \varepsilon_t$ | $y_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m}) (1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha (\ell_{t-1} + \phi b_{t-1} + s_{t-m}) \varepsilon_t$ $b_t = \phi b_{t-1} + \beta (\ell_{t-1} + \phi b_{t-1} + s_{t-m}) \varepsilon_t$ $s_t = s_{t-m} + \gamma (\ell_{t-1} + \phi b_{t-1} + s_{t-m}) \varepsilon_t$ | $y_t = (\ell_{t-1} + \phi b_{t-1}) s_{t-m} (1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1}) (1 + \alpha \varepsilon_t)$ $b_t = \phi b_{t-1} + \beta (\ell_{t-1} + \phi b_{t-1}) \varepsilon_t$ $s_t = s_{t-m} (1 + \gamma \varepsilon_t)$ |

Fuente: Hyndman y Athanasopoulos (2021).

Una gran ventaja del marco estadístico ETS es que se pueden utilizar criterios de información para la selección del modelo. El AIC, AICcc y BIC, pueden usarse aquí para determinar cuál de los modelos ETS es el más apropiado para una serie de tiempo determinada.

Para los modelos ETS, se define el criterio de información de Akaike (AIC)

$$AIC = -2\log(L) + 2k,$$

donde L es la verosimilitud del modelo y k es el número total de parámetros y estados iniciales que se han estimado (incluyendo la varianza residual).

El AIC corregido por sesgo de muestra pequeña (AICcc) se define como

$$AIC_c = AIC + \frac{2k(k+1)}{T-k-1},$$

y el Criterio de Información Bayesiano (BIC) es

$$BIC = AIC + k[\log(T) - 2].$$

Tres de las combinaciones de (Error, Tendencia, Estacionalidad) pueden generar dificultades numéricas. Específicamente, los modelos que pueden causar tales inestabilidades son ETS(A,N,M), ETS(A,A,M) y ETS(A,Ad,M), debido a la división por valores potencialmente cercanos a cero en el estado ecuaciones Normalmente no consideramos estas combinaciones particulares cuando seleccionamos un modelo.

Los modelos con errores multiplicativos son útiles cuando los datos son estrictamente positivos, pero no son numéricamente estables cuando los datos contienen ceros o valores negativos. Por tanto, no se considerarán modelos de error multiplicativo si la serie temporal no es estrictamente positiva. En ese caso, solo se aplicarán los seis modelos totalmente aditivos.

4.3.2 ARIMA

Para poder aplicar la metodología ARIMA, la serie que se está analizando debe cumplir dos importantes características (Pankratz, 2009):

- La serie tiene que ser estacionaria en media (tendencia constante).

$$\text{Media: } E(X_t) = E(X_{t+h}) = \mu$$

- La serie tiene que ser estacionaria en varianza (varianza constante).

$$\text{Varianza: } V(X_t) = V(X_{t+h}) = \sigma^2$$

Si la serie cumple estas condiciones, decimos que la serie es estacionaria. La serie puede ser estacionaria en media y/o varianza. Para comprobar la estacionariedad de la serie, se observa la gráfica de la función de autorregresión simple (ACF), que tiene la siguiente expresión:

$$\rho(h) = \rho(X_t, X_{t+h}) = \frac{E((X_t - \mu)(X_{t+h} - \mu))}{\sqrt{E(X_t - \mu)^2 E(X_{t+h} - \mu)^2}} = \frac{\gamma(h)}{\gamma(0)}$$

Si una serie es estacionaria, su función de autorregresión decae rápidamente hacia cero, en caso contrario, la serie no es estacionaria. Sin embargo, no todas las series cumplen estas características, pero mediante unas sencillas transformaciones podemos conseguir que la serie sea estacionaria.

Si la serie no cumple la estacionariedad en media, se debería de aplicar la transformación usando la diferenciación regular. Para diferenciar de manera regular una serie de datos, se define una nueva variable W_t , que es una transformación de la serie original de la siguiente forma:

$$W_t = Z_t - Z_{t-1}, t = 2, 3, \dots, n$$

Ahora, la serie W_t es la serie Z_t aplicando una diferenciación. En caso de que aplicando la primera diferenciación regular tampoco se consiga una media constante, se pueden aplicar tantas diferenciaciones como sean necesarias. Se define como una segunda diferenciación:

$$W_t = (Z_t - Z_{t-1}) - (Z_{t-1} - Z_{t-2}), t = 3, 4, \dots, n$$

Aparte de las diferenciaciones regulares, también están las diferenciaciones estacionales. Este tipo de diferenciación se aplica cuando una serie tiene una media no constante estacionalmente. En este caso, se aplicaría una diferenciación estacional de orden s , donde s representa el ciclo del periodo. Por ejemplo, si los datos son mensuales $s = 12$ observando una estacionalidad mensual, si los datos son diarios $s = 7$ representa una estacionalidad semanal.

A continuación, se muestra la fórmula de la primera diferenciación estacional:

$$W_t = Z_t - Z_{t-s}, t = 2, 3, \dots, n$$

Y si aún no se consigue una media estacional constante, se aplica una segunda diferenciación:

$$W_t = (Z_t - Z_{t-s}) - (Z_{t-1} - Z_{t-1-s}), t = 3, 4, \dots, n$$

A efectos prácticos, la primera diferenciación estacional se necesita con bastante frecuencia, la segunda solo ocasionalmente y una tercera o más diferenciaciones casi nunca es necesario. Una serie temporal, aparte de ser no estacionaria en media también, puede presentar no estacionariedad en la varianza. Para resolver este problema y poder trabajar con esta metodología, se aplica una transformación logarítmica a la serie original.

Modelo autorregresivo

El modelo AR(p), modelo Autorregresivo de orden p retrocede p periodos para realizar la predicción, es decir, el valor actual de la serie es una combinación lineal de la serie estudiando los p periodos anteriores. El modelo autorregresivo es un modelo de regresión en que tanto la variable dependiente como la variable explicativa, son la misma variable. La variable dependiente está explicada por las observaciones procedentes de la misma variable (Villavicencio, 2010). La fórmula matemática que define el modelo es:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + Z_t$$

Expresada en otros términos, donde B representa el operador retroceso ($BX_t = X_{t-1}$), la fórmula que define el modelo sería:

$$(1 - \phi_1 B - \phi_2 B - \dots - \phi_p B) X_t = Z_t, Z_t \sim N(0, \sigma_z^2)$$

Donde X_t es el valor resultante para el mes t, y los ϕ son los parámetros que multiplican a las observaciones del día anterior hasta el mes $t - p$ y Z_t representa el ruido blanco.

Modelo media móvil

El modelo MA(q) o el modelo *Moving Average* (Media móvil) de orden q es un modelo de regresión que depende de los errores pasados y tiene la siguiente expresión:

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \dots + \theta_q Z_{t-q}, Z_t \sim N(0, \sigma_z^2)$$

Expresada de otra manera, usando el operador retroceso, se puede definir como:

$$X_t = (1 + \theta_1 B + \theta_2 B + \dots + \theta_q B) Z_t, Z_t \sim N(0, \sigma_z^2)$$

En el Cuadro 4.3.2. podemos ver una tabla comparativa sobre las equivalencias de los modelos ETS y ARIMA.

Cuadro 4.3.2: Comparativa modelos ETS y ARIMA

| ETS model | ARIMA model | Parameters |
|--------------------------|--------------------------------------|--|
| ETS(A,N,N) | ARIMA(0,1,1) | $\theta_1 = \alpha - 1$ |
| ETS(A,A,N) | ARIMA(0,2,2) | $\theta_1 = \alpha + \beta - 2$ $\theta_2 = 1 - \alpha$ |
| ETS(A,A _d ,N) | ARIMA(1,1,2) | $\phi_1 = \phi$ $\theta_1 = \alpha + \phi\beta - 1 - \phi$ $\theta_2 = (1 - \alpha)\phi$ |
| ETS(A,N,A) | ARIMA(0,1,m)(0,1,0) _m | |
| ETS(A,A,A) | ARIMA(0,1,m + 1)(0,1,0) _m | |
| ETS(A,A _d ,A) | ARIMA(1,0,m + 1)(0,1,0) _m | |

Fuente: Hyndman y Athanasopoulos (2021).

4.3.3 Random Forest

Los métodos basados en árboles no son paramétricos. Como tal, asumen muy poco sobre el proceso de generación de datos, lo que ha llevado a su uso en una variedad de aplicaciones diferentes. Los métodos basados en árboles son similares a los árboles de decisión, donde la estructura jerárquica permite a los usuarios tomar una serie de decisiones secuenciales para llegar a una conclusión. Dada esta estructura, las decisiones pueden pensarse como entradas x , y la conclusión como salida, y . Curiosamente, uno de los primeros artículos sobre la construcción automática de árboles de decisión, Morgan y Sonquist (1963), fue escrito por un economista. Sin embargo, fue solo a partir de la publicación de Breiman, Friedman, Olshen y Stone (1984) que los métodos basados en árboles ganaron prominencia. Sobre la base de este trabajo seminal, dos tipos de árboles, los árboles de clasificación y regresión (CART), se han vuelto populares. Ambos tipos de árboles tienen un marco/estructura similar, y la principal diferencia es la naturaleza de la variable de salida, y . En el caso de los árboles de clasificación, y es categórico o nominal, mientras que para los árboles de regresión, y es numérico. En la literatura de ML, esta distinción se denomina cualitativa frente a cuantitativa; consulte James, Witten, Hastie y Tibshirani (2013) para obtener más información.

La popularidad de los árboles se debe principalmente a su marco centrado en decisiones, es decir, al generar un conjunto de entradas basadas en reglas x , los árboles pueden predecir un resultado y . La estructura resultante de un árbol es fácil de entender, interpretar y visualizar. Las entradas x más significativas son fácilmente identificables a partir de un árbol. Esto es útil con respecto a la exploración de datos y requiere conocimientos estadísticos limitados o nulos para interpretar los resultados. En la mayoría de los casos, los datos requieren relativamente menos preprocesamiento para generar un árbol en comparación con otros métodos de regresión y clasificación. Dadas estas ventajas, los árboles se utilizan para modelar datos numéricos y categóricos. A continuación, se proporciona un resumen del proceso de construcción del árbol. El proceso de construcción de un árbol implica la partición recursiva del espacio regresor (X_i) para predecir y_i . La partición produce J regiones distintas y no superpuestas R_1, R_2, \dots, R_J . Para todas las observaciones que caen en una región R_j , se aplica la misma predicción a todas ellas. En el caso de los árboles de regresión (y numérico), esto es simplemente la media de y para esa región. Para los árboles de clasificación, la predicción es el caso/categoría más frecuente en esa región. Una pregunta natural es '¿cómo se realiza la partición/división de la variable? El enfoque utilizado se conoce como división binaria recursiva. El árbol comienza seleccionando una variable X_j y un valor de umbral correspondiente s tal que dividir el espacio del regresor en regiones $\{X|X_j < s\}$ y $\{X|X_j \geq s\}$ conduce a la mayor reducción posible del error, donde 'error' es una medida de qué tan bien las predicciones coinciden con los valores reales. Este proceso de división en regiones considerando todos los umbrales posibles en todas las variables que minimiza el error continúa hasta que se alcanza un criterio de parada; por ejemplo, un número mínimo de observaciones en una región particular. En el caso de árboles de regresión (y numérico) se utiliza la suma residual de cuadrados (RSS), definida como

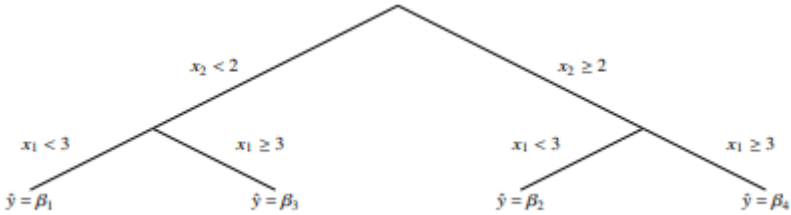
$$RSS = \sum_{j=1}^J \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2,$$

donde \hat{y}_{R_j} es la media de y para la región R_j . Minimizar RSS implica que todas las desviaciones al cuadrado para una región determinada se minimizan en todas las regiones. En

una configuración de clasificación, RSS no se puede usar como criterio para dividir variables, ya que no podemos usar los valores medios de y , ya que es categórico. Sin embargo, hay una serie de medidas alternativas, cada una con sus propias ventajas y limitaciones.

Una forma popular de evaluar el rendimiento de un árbol de clasificación binaria es mediante una matriz de confusión (Figura 2.12). Esta matriz compara directamente cada uno de los valores predichos con sus valores reales correspondientes. Como tal, esta matriz clasifica todas las observaciones en una de cuatro categorías. Los dos primeros son Verdaderos Positivos (TP) y Verdaderos Negativos (TN). Estas categorías cuentan todas las observaciones donde las predicciones son consistentes con el valor real/observado. Las dos categorías restantes Falsos positivos (FP) y Falsos negativos (FN) cuentan todas las observaciones donde las predicciones no son consistentes con los valores reales/observados. Según los recuentos de las cuatro categorías, se puede calcular una variedad de métricas para medir diferentes facetas del modelo, como la precisión general. Tenga en cuenta que la matriz de confusión puede ser calculada en cualquier modelo de clasificación. Esto incluye el modelo Logit o Probit utilizado en econometría. Esto permite al usuario no solo comparar diferentes versiones del mismo modelo, sino también comparar el rendimiento en diferentes tipos de modelos. Por ejemplo, una matriz de confusión basada en un modelo Logit se puede comparar con una matriz de confusión resultante de un árbol de clasificación.

Figura 4.3.3.1: Vista de la regresión de árbol



Fuente: Anselin et al. (2008)

Figura 4.3.3.2: Matriz de confusión

| | | Actual Values | |
|------------------|--------------|---------------|--------------|
| | | Positive (1) | Negative (0) |
| Predicted Values | Positive (1) | TP | FP |
| | Negative (0) | FN | TN |

Fuente: Anselin et al. (2008)

Como señaló Breiman (1996b), se sabe que los árboles son inestables, ya que un pequeño cambio en la muestra de entrenamiento puede hacer que la estructura del árbol resultante cambie drásticamente. Este cambio implica que las predicciones también cambiarán, lo que generará una gran variación en las predicciones. Dado esto, no es sorprendente que los árboles sean generalmente superados por otros métodos de clasificación y regresión con respecto a la precisión predictiva. Esto se debe al hecho de que los árboles individuales son propensos al crecimiento excesivo y la poda no garantiza la estabilidad. Otros dos métodos populares que se pueden aplicar a los árboles para mejorar su rendimiento predictivo son el embolsado y los bosques aleatorios. Ambos métodos se basan en la agregación, por lo que la predicción general se obtiene combinando las predicciones de muchos árboles. Los detalles breves de estos dos enfoques se discuten a continuación.

El embolsado (Bootstrap Aggregation) consiste en tomar muestras repetidamente y construir árboles a partir de cada muestra, y luego combinar las predicciones de cada árbol para obtener la predicción final. Los siguientes pasos describen este proceso:

1. Generar B diferentes muestras autocargadas de tamaño n (con reemplazo).
2. Construya un árbol en cada muestra y obtenga una predicción para un x dado, $f_b(x)$
3. Calcule el promedio de todas las predicciones B para obtener la predicción de embolsado final,

$$\hat{f}_{\text{bagging}}(\mathbf{x}) = \frac{1}{B} \sum_{b=1}^B \hat{f}_b(\mathbf{x}).$$

Los pasos anteriores funcionan bien para los árboles de regresión. Con respecto a los árboles de clasificación, la predicción final consiste en seleccionar la categoría/clase que más se presenta entre la predicción B

Como argumenta Breiman (1996a), el procedimiento de embolsado descrito anteriormente mejora las predicciones al reducir su varianza. El concepto de bosques aleatorios amplía el enfoque de embolsado al restringir el número de covariables en cada muestra de arranque. En otras palabras, solo se usa un subconjunto de todas las covariables para construir cada árbol. Como regla general, si hay un total de p covariables, entonces solo se seleccionan \sqrt{p} covariables para cada muestra de arranque. Los siguientes pasos describen el enfoque de bosques aleatorios:

1. Extraiga una muestra aleatoria (con reemplazo) de tamaño n.
2. Seleccionar aleatoriamente m covariables del conjunto completo de p covariables (donde $m \approx \sqrt{p}$).
3. Construya un árbol usando las m covariables seleccionadas y obtenga una predicción para un $f_{b,m}(x)$ dado. Esta es la predicción del árbol bth basada en m covariables seleccionadas.
4. Repita los pasos 1 a 3 para todas las muestras de arranque B.
5. Calcule el promedio de las predicciones de árboles individuales para obtener la predicción de bosque aleatorio, frf

$$\hat{f}_{\text{rf}}(\mathbf{x}) = \frac{1}{B} \sum_{b=1}^B \hat{f}_{b,m}(\mathbf{x}).$$

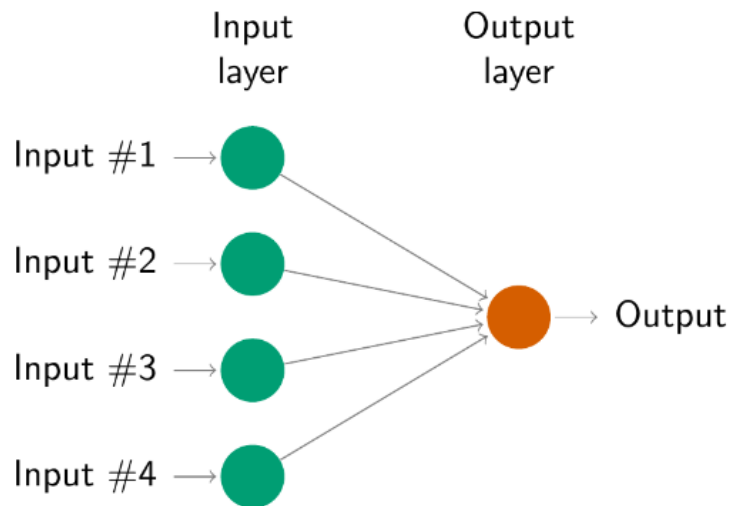
6. En la configuración de clasificación, cada árbol de arranque predice la categoría/clase que ocurre con mayor frecuencia y el predictor de bosque aleatorio selecciona la clase que ocurre con mayor frecuencia en todos los árboles B. Tenga en cuenta que el embolsado es un caso especial de bosque aleatorio, cuando $m = p$ (paso 2). Similar al enfoque de embolsado, el enfoque de bosques aleatorios reduce la varianza, pero a diferencia del embolsado, el enfoque

de bosque aleatorio también reduce la correlación entre los árboles con arranque (debido al hecho de que no todos los árboles tienen las mismas covariables). Al permitir una selección aleatoria de covariables, el proceso de división no es el mismo para cada árbol y esto parece mejorar la predicción general. Varios estudios numéricos muestran que los bosques aleatorios funcionan bien en muchas aplicaciones. Estos métodos de agregación están relacionados con el concepto de combinaciones de pronósticos, que es popular en la econometría convencional.

4.3.4 Neural Networks

Se puede pensar en una red neuronal como una red de "neuronas" que se organizan en capas. Los predictores (o entradas) forman la capa inferior y los pronósticos (o salidas) forman la capa superior. También puede haber capas intermedias que contengan "neuronas ocultas".

Figura 4.3.4.1: Una red neuronal simple equivalente a una regresión lineal

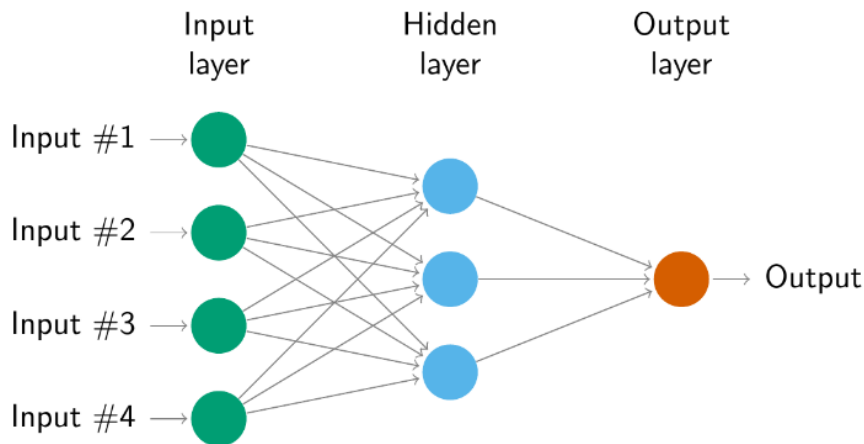


Fuente: Hyndman y Athanasopoulos (2021).

Las redes más simples no contienen capas ocultas y son equivalentes a las regresiones lineales. La figura 4.3.4.1 muestra la versión de red neuronal de una regresión lineal con cuatro predictores. Los coeficientes adjuntos a estos predictores se denominan "pesos". Los pronósticos se obtienen mediante una combinación lineal de las entradas. Los pesos se seleccionan en el marco de la red neuronal utilizando un "algoritmo de aprendizaje" que minimiza una "función de costo" como el MSE. Por supuesto, en este ejemplo simple, podemos usar la regresión lineal, que es un método mucho más eficiente para entrenar el modelo.

Una vez que agregamos una capa intermedia con neuronas ocultas, la red neuronal se vuelve no lineal. Un ejemplo simple se muestra en la Figura 4.3.4.2

Figura 4.3.4.2: Una red neuronal con 4 inputs y una capa oculta con tres neuronas ocultas



Fuente: Hyndman y Athanasopoulos (2021).

Esto se conoce como una red de avance de múltiples capas, donde cada capa de nodos recibe entradas de las capas anteriores. Las salidas de los nodos en una capa son entradas para la siguiente capa. Las entradas de cada nodo se combinan mediante una combinación lineal ponderada. Luego, el resultado se modifica mediante una función no lineal antes de generarse.

Con datos de series de tiempo, los valores retrasados de la serie de tiempo se pueden usar como entradas para una red neuronal, tal como usamos valores retrasados en un modelo de autorregresión lineal. A esto lo llamamos autorregresión de red neuronal o modelo NNAR.

Solo consideraremos redes feed-forward con una capa oculta y usaremos la notación NNAR(p,k) para indicar que hay p entradas retrasadas y k nodos en la capa oculta. Por ejemplo, un modelo NNAR(9,5) es una red neuronal con las últimas nueve observaciones ($y_{t-1}, y_{t-2}, \dots, y_{t-9}$) utilizadas como entradas para pronosticar la salida y_t , y con cinco neuronas en el capa oculta. Un modelo NNAR(p,P,0)_m es equivalente a un modelo ARIMA(p,0,0)(P,0,0)_m, pero sin las restricciones en los parámetros para asegurar la estacionariedad.

La función NNETAR() se ajusta a un modelo NNAR(p,P,k)_m. Si no se especifican los valores de p y P, se seleccionan automáticamente. Para series temporales no estacionales, el valor predeterminado es el número óptimo de retrasos (según el AIC) para un modelo AR(p) lineal. Para las series temporales estacionales, los valores predeterminados son P=1 y p se elige del modelo lineal óptimo ajustado a los datos ajustados estacionalmente. Si no se especifica k, se establece en $k=(p+P+1)/2=(p+P+1)/2$ (redondeado al entero más próximo).

Cuando se trata de pronósticos, la red se aplica iterativamente. Para pronosticar un paso adelante, simplemente usamos las entradas históricas disponibles. Para pronosticar dos pasos adelante, usamos el pronóstico de un paso como entrada, junto con los datos históricos. Este proceso continúa hasta que hayamos calculado todos los pronósticos requeridos.

4.3 Evaluación de pronósticos

El método más utilizado para evaluar los pronósticos es el error cuadrático medio (RMSE). Es una medida del tamaño de los errores de pronóstico y es simplemente la raíz cuadrada del error de pronóstico cuadrático medio. Por lo tanto, al calcular el RMSE para un solo error de pronóstico, derivaríamos,

$$\sqrt{\mathbb{E}_t \left[\{ \hat{\epsilon}_R(h) \}^2 \right]} = \sqrt{\mathbb{E}_t \left[\{ y_{R+h} - \hat{y}_R(h) \}^2 \right]}$$

La diferencia entre esta medida y la que se calculó para el sesgo, es que luego de elevar al cuadrado todos los errores, los errores positivos y negativos ya no se anularían entre sí, al calcular el promedio sobre pronósticos sucesivos. Por lo tanto, con el uso del vector de errores de pronóstico fuera de la muestra, podríamos calcular el RMSE como,

$$\text{RMSE}_h = \sqrt{\frac{1}{(T-R)} \sum_{i=R}^T \hat{\epsilon}_{R+i}^2(h)}$$

donde $\hat{\epsilon}_{R+i}(h)$ es el error de pronóstico definido anteriormente. Como tal, el RMSE es una función de pérdida simétrica, donde los pronósticos que son demasiado altos o demasiado bajos se ponderan por igual. Naturalmente, los errores de pronóstico más pequeños se consideran mejores que los más grandes y, como tal, un valor RMSE bajo indica un mejor rendimiento de pronóstico.

4.5. Cálculo impacto covid-19

Para el cálculo del impacto del covid-19 calcularemos primero la predicción acumulada (\widehat{PIB} denota la predicción) y haremos la diferencia entre el valor real y el acumulado. Esto se ve respaldado por las siguientes ecuaciones:

$$g_T = \ln \left(\frac{PIB_T}{G_{T-1}} \right)$$

$$\hat{g}_{T+1} = \ln \left(\frac{\widehat{PIB}_{T+1}}{G_T} \right)$$

$$\hat{g}_{T+4} = \ln \left(\frac{\widehat{PIB}_{T+4}}{PIB_T} \right)$$

$$\ln \left(\frac{\widehat{PIB}_{T+4}}{\widehat{PIB}_{T+3}} \right) + \ln \left(\frac{\widehat{PIB}_{T+3}}{\widehat{PIB}_{T+2}} \right) + \ln \left(\frac{\widehat{PIB}_{T+2}}{\widehat{PIB}_{T+1}} \right) + \ln \left(\frac{\widehat{PIB}_{T+1}}{PIB_T} \right) = \ln \left(\frac{\widehat{G}_{T+4}}{PIB_T} \right)$$

V. RESULTADOS

En este apartado se presentan los resultados de los pronósticos obtenidos para cada modelo utilizado en el estudio. Después de realizar los cálculos correspondientes, se procede a comparar los diferentes pronósticos generados por cada modelo para cada país de la muestra. El objetivo principal de esta comparación es determinar cuál de los modelos ofrece el menor Error Cuadrático Medio (RMSE, por sus siglas en inglés).

El Error Cuadrático Medio es una medida utilizada para evaluar la precisión de los pronósticos en relación con los valores reales. Al elegir el modelo con el RMSE más bajo para cada país, podemos identificar la opción que mejor se ajusta a los datos históricos y, por lo tanto, tiene mayor capacidad predictiva.

Una vez seleccionados los mejores modelos para cada país, se lleva a cabo el análisis del impacto. Este análisis implica comparar los valores reales registrados durante el período de 2020 con los valores pronosticados por los modelos seleccionados. La diferencia entre el valor real y el valor pronosticado se utiliza como indicador del impacto, permitiendo identificar en qué medida los pronósticos se desviaron de los resultados reales.

El análisis del impacto es fundamental para evaluar la calidad y precisión de los modelos utilizados en el estudio. Permite determinar si los modelos subestimaron o sobreestimaron los valores reales, y brinda información valiosa para futuras decisiones y ajustes en los procesos de pronóstico.

En resumen, este apartado presenta los resultados de los pronósticos generados por diferentes modelos, se compara su desempeño a través del RMSE y se selecciona el modelo con el menor RMSE para cada país. Luego, se realiza el análisis del impacto utilizando los mejores modelos, evaluando las diferencias entre los valores reales y los pronosticados durante el período de 2020. Este análisis proporciona información crucial para comprender la efectividad de los modelos y su capacidad para predecir con precisión eventos futuros.

5.1 ETS

5.1.1 Pronóstico RMSE individual.

En este apartado, se presentan los resultados detallados de los pronósticos generados por los distintos modelos utilizados en el estudio. Estos resultados se muestran en el Cuadro 5.1.1, donde se exhiben los valores del Error Cuadrático Medio (RMSE) correspondientes a cada país y horizonte de tiempo considerado.

El RMSE es una medida ampliamente utilizada en la evaluación de pronósticos, ya que proporciona una estimación cuantitativa de la precisión de los modelos. Al observar los valores en el Cuadro 5.1.1, podemos identificar las combinaciones de país y horizonte de tiempo en las que se obtuvo el menor RMSE, lo que indica la mayor precisión en los pronósticos.

Al analizar el cuadro, se destaca que el horizonte h1 es aquel que minimiza el RMSE en la mayoría de los países. Esto implica que, en general, los pronósticos generados para un período de tiempo más corto, es decir, a un horizonte de tiempo más cercano, se ajustaron mejor a los valores reales. Estos resultados respaldan la idea de que los modelos utilizados en el estudio tienen una mayor capacidad de pronóstico a corto plazo, brindando una mayor precisión en la predicción de los eventos más inmediatos.

La elección del horizonte de tiempo adecuado es crucial para obtener pronósticos confiables y útiles. En este caso, los resultados indican que para una mayoría de países, el horizonte h1 proporciona los pronósticos más precisos. Sin embargo, es importante tener en cuenta que esta elección puede variar dependiendo del país, el contexto específico y la naturaleza de los datos utilizados en el estudio.

En conclusión, se observa que el horizonte h1 tiende a minimizar el RMSE en la mayoría de los países, lo que indica una mayor precisión en los pronósticos a corto plazo.

Cuadro 5.1.1: Pronóstico RMSE ETS

| <i>PAÍS</i> | <i>h1</i> | <i>h2</i> | <i>h3</i> | <i>h4</i> |
|-------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <i>AUS</i> | <u>0.01005</u> | 0.01009 | 0.01023 | 0.01028 |
| <i>CAN</i> | <u>0.0167</u> | 0.01714 | 0.01727 | 0.01717 |
| <i>FRA</i> | <u>0.02026</u> | 0.02033 | 0.02052 | 0.02065 |
| <i>ITA</i> | 0.0195 | <u>0.01948</u> | 0.01961 | 0.01973 |
| <i>KOR</i> | 0.01048 | 0.01012 | 0.01044 | <u>0.00973</u> |
| <i>NOR</i> | 0.01283 | 0.01253 | <u>0.01246</u> | 0.01273 |
| <i>SWE</i> | 0.01588 | 0.01557 | 0.01586 | <u>0.01554</u> |
| <i>CHE</i> | <u>0.01231</u> | 0.01243 | 0.01252 | 0.01246 |
| <i>USA</i> | <u>0.01312</u> | 0.01344 | 0.01354 | 0.01358 |
| <i>OECD</i> | <u>0.01544</u> | 0.01562 | 0.016 | 0.01607 |

Fuente: Elaboración propia

5.1.2 Elección del modelo.

En el análisis de los resultados presentados en el Cuadro 5.1.2, se observa una tendencia consistente en la elección del modelo óptimo para todos los casos evaluados. El modelo seleccionado en todos los países y horizontes de tiempo es el ETS(A,N,N).

El ETS(A,N,N) se refiere a un modelo de Suavizamiento Exponencial Triple, donde se utilizan componentes aditivos (A) y se asume que no hay estacionalidad (N) ni tendencia (N). Este modelo es ampliamente utilizado en pronósticos debido a su capacidad para capturar patrones de variabilidad en los datos sin tener en cuenta tendencias o estacionalidades.

No obstante, es importante destacar que la elección del modelo ETS(A,N,N) como el más adecuado para todos los casos no implica que sea el modelo óptimo en todas las situaciones o para todos los conjuntos de datos. Cada contexto y conjunto de datos pueden requerir un enfoque de modelado diferente para obtener los pronósticos más precisos y relevantes.

Cuadro 5.1.2: Modelos ETS por país

| | | | | |
|------------|------------|------------|------------|-------------|
| AUS | CAN | FRA | ITA | KOR |
| ETS(A,N,N) | ETS(A,N,N) | ETS(A,N,N) | ETS(A,N,N) | ETS(A,N,N) |
| NOR | SWE | CHE | USA | OECD |
| ETS(A,N,N) | ETS(A,N,N) | ETS(A,N,N) | ETS(A,N,N) | ETS(A,N,N) |

Fuente: Elaboración propia

5.1.3 AICc

En la Figura 5.1.3 se representa la evolución del criterio de información de Akaike corregido (AICc) para el modelo ETS a lo largo del período de análisis. El AICc es una medida que permite evaluar la calidad relativa de los modelos estadísticos, teniendo en cuenta tanto la bondad de ajuste como la complejidad del modelo.

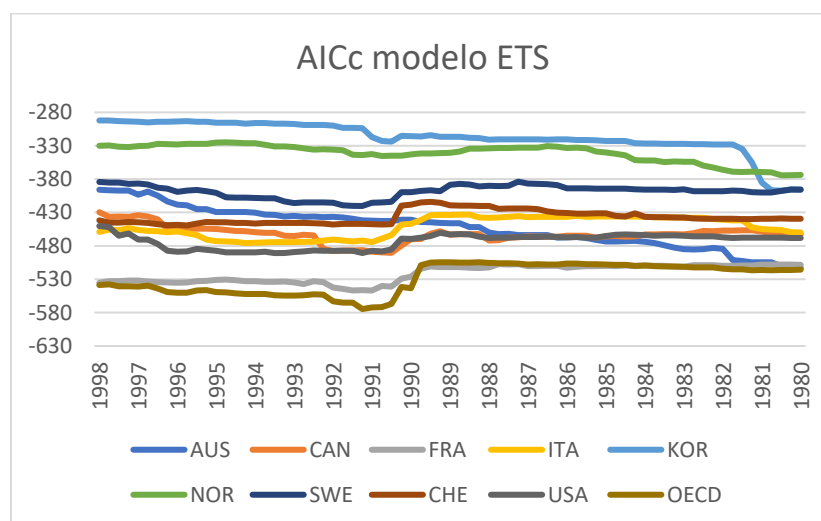
Al examinar la Figura 5.1.3, se observa un patrón interesante en la evolución del AICc a lo largo del tiempo. Inicialmente, el AICc muestra valores más altos, lo que indica una menor calidad de ajuste del modelo ETS. A medida que avanza el período de análisis, el AICc comienza a disminuir gradualmente, lo que sugiere una mejora en el ajuste del modelo a los datos observados.

Este comportamiento puede estar relacionado con la adaptación del modelo a los patrones y cambios en los datos a lo largo del tiempo. Al principio, el modelo puede tener dificultades para capturar de manera precisa las características subyacentes de los datos, lo que se refleja en valores más altos del AICc. Sin embargo, a medida que el modelo se ajusta mejor a los datos, el AICc disminuye, indicando una mayor precisión en la representación de los patrones y la estructura de los datos.

La evolución del AICc proporciona información valiosa sobre la calidad y la capacidad predictiva del modelo ETS a lo largo del tiempo. Al considerar esta evolución, los investigadores y analistas pueden identificar períodos en los que el modelo muestra un mejor rendimiento y confiar más en los pronósticos generados durante esos períodos.

Es importante tener en cuenta que el AICc es solo una de las medidas utilizadas para evaluar los modelos, y su interpretación debe realizarse en conjunto con otras métricas y análisis adicionales. Además, es recomendable considerar la naturaleza de los datos y el contexto específico del estudio al interpretar los resultados del AICc y tomar decisiones basadas en ellos.

Figura 5.1.3: AICc modelo ETS



Fuente: Elaboración propia

5.2 ARIMA

5.2.1 Pronóstico RMSE individual.

En este apartado, se presentan los resultados detallados de los pronósticos generados por los distintos modelos utilizados en el estudio. Estos resultados se muestran en el Cuadro 5.2.1, donde se exhiben los valores del Error Cuadrático Medio (RMSE) correspondientes a cada país y horizonte de tiempo considerado.

Al observar los valores en el Cuadro 5.2.1, podemos identificar las combinaciones de país y horizonte de tiempo en las que se obtuvo el menor RMSE, lo que indica la mayor precisión en los pronósticos.

Al analizar el cuadro, se destaca que el horizonte h1 es aquel que minimiza el RMSE en la mayoría de los países. Esto implica que, en general, los pronósticos generados para un período de tiempo más corto, es decir, a un horizonte de tiempo más cercano, se ajustaron mejor a los valores reales. Estos resultados respaldan la idea de que los modelos utilizados en el estudio tienen una mayor capacidad de pronóstico a corto plazo, brindando una mayor precisión en la predicción de los eventos más inmediatos.

La elección del horizonte de tiempo adecuado es crucial para obtener pronósticos confiables y útiles. En este caso, los resultados indican que, para una mayoría de países, el horizonte h1 proporciona los pronósticos más precisos. Sin embargo, es importante tener en cuenta que esta elección puede variar dependiendo del país, el contexto específico y la naturaleza de los datos utilizados en el estudio.

Cuadro 5.2.1: Pronóstico RMSE ARIMA

| <i>PAÍS</i> | <i>h1</i> | <i>h2</i> | <i>h3</i> | <i>h4</i> |
|-------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <i>AUS</i> | <u>0.01002</u> | 0.0101 | 0.01021 | 0.01026 |
| <i>CAN</i> | <u>0.01626</u> | 0.01646 | 0.01642 | 0.01638 |
| <i>FRA</i> | <u>0.02002</u> | 0.02004 | 0.02007 | 0.0202 |
| <i>ITA</i> | 0.01895 | <u>0.0188</u> | 0.01882 | 0.01888 |
| <i>KOR</i> | 0.0113 | 0.01085 | <u>0.01076</u> | 0.01091 |
| <i>NOR</i> | 0.01283 | 0.01254 | <u>0.01246</u> | 0.01275 |
| <i>SWE</i> | 0.01553 | 0.01537 | 0.01537 | <u>0.01495</u> |
| <i>CHE</i> | 0.01191 | <u>0.01171</u> | 0.01174 | 0.01175 |
| <i>USA</i> | <u>0.01288</u> | 0.0131 | 0.01309 | 0.01304 |
| <i>OECD</i> | <u>0.01529</u> | <u>0.01529</u> | 0.01534 | 0.01532 |

Fuente: Elaboración propia

5.2.2 Elección del modelo.

En el Cuadro 5.2.2 se presentan los resultados de la selección del modelo ARIMA utilizado para el pronóstico en cada país evaluado en el estudio. La elección de un modelo ARIMA específico se basa en diferentes criterios, como la calidad de ajuste, la significancia de los coeficientes y la capacidad predictiva del modelo.

Al analizar el Cuadro 5.2.2, se observa que se han seleccionado diferentes modelos ARIMA para el pronóstico en cada país. Esto indica que la estructura y las características de los datos pueden variar significativamente de un país a otro, lo que requiere la adaptación del modelo ARIMA para cada caso específico.

Es importante destacar que la selección del modelo ARIMA se basa en un análisis exhaustivo de los datos y en la aplicación de técnicas estadísticas para evaluar y comparar diferentes opciones. Los criterios utilizados para la selección del modelo pueden incluir el AIC (criterio de información de Akaike), el BIC (criterio de información bayesiano) u otros indicadores de bondad de ajuste.

La elección de un modelo ARIMA adecuado es fundamental para obtener pronósticos precisos y confiables. Al seleccionar el modelo ARIMA óptimo para cada país, se garantiza que se capturen las características particulares de los datos y se maximice la precisión del pronóstico.

Cuadro 5.2.2: Elección modelo ARIMA

| | | | | |
|--------------|--------------|--------------|--------------|--------------|
| AUS | CAN | FRA | ITA | KOR |
| ARIMA(0,0,0) | ARIMA(1,0,0) | ARIMA(2,0,0) | ARIMA(1,1,1) | ARIMA(1,0,0) |
| NOR | SWE | CHE | USA | OECD |
| ARIMA(1,0,0) | ARIMA(2,0,2) | ARIMA(0,0,1) | ARIMA(2,0,0) | ARIMA(2,0,0) |

Fuente: Elaboración propia

5.2.3 AICc

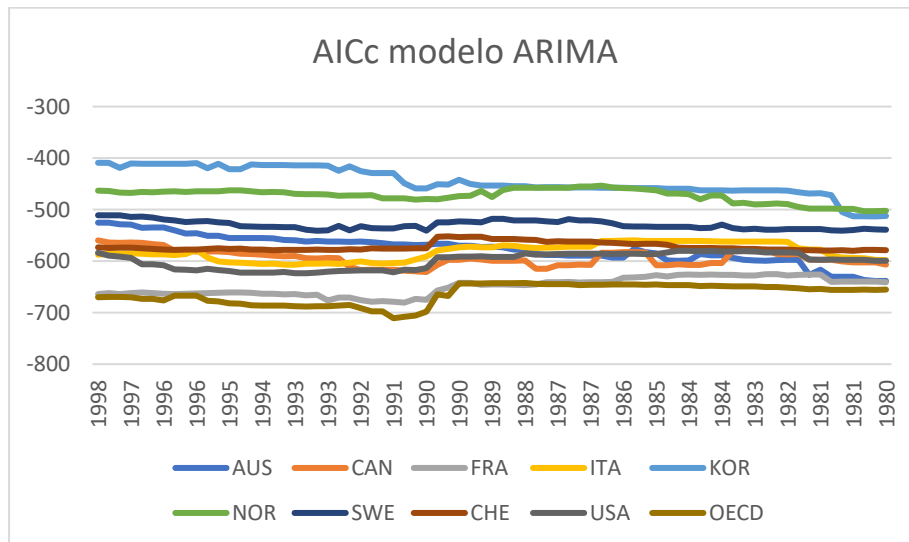
En la Figura 5.2.3 vemos la evolución del AICc de los modelos ARIMA de cada país a lo largo del tiempo. La Figura 5.2.3 ofrece una representación visual de la evolución del criterio de información de Akaike corregido (AICc) para los modelos ARIMA utilizados en el pronóstico de cada país a lo largo del tiempo.

Al examinar la Figura 5.2.3, se pueden identificar patrones y tendencias interesantes en la evolución del AICc para los diferentes modelos ARIMA. Estos patrones pueden revelar cambios en la calidad de ajuste y en la capacidad predictiva de los modelos a medida que avanza el tiempo.

En algunos países, se puede observar una disminución gradual del AICc a lo largo del tiempo, lo que indica una mejora en el ajuste y en la capacidad predictiva de los modelos ARIMA utilizados. Esta disminución puede ser el resultado de ajustes y refinamientos realizados en los modelos a medida que se obtuvieron nuevos datos y se profundizó en el análisis. Esta evolución positiva sugiere que los modelos ARIMA se adaptaron de manera más precisa a las características cambiantes de los datos en esos países específicos.

Por otro lado, en algunos casos, se pueden identificar fluctuaciones o variaciones en la evolución del AICc a lo largo del tiempo. Estas variaciones pueden reflejar cambios en los patrones o la estructura de los datos, lo que a su vez influye en la capacidad predictiva de los modelos ARIMA. Estas fluctuaciones pueden requerir una mayor exploración y análisis para comprender las causas subyacentes y ajustar los modelos en consecuencia.

Figura 5.2.3: AICc modelo ARIMA



Fuente: Elaboración propia

5.3 ARIMA con factor económico global (ARIMA_X)

5.3.1 Pronóstico RMSE individual.

En este apartado, se presentan los resultados detallados de los pronósticos generados por los distintos modelos utilizados en el estudio. Estos resultados se muestran en el Cuadro 5.3.1, donde se exhiben los valores del Error Cuadrático Medio (RMSE) correspondientes a cada país y horizonte de tiempo considerado.

Al observar los valores en el Cuadro 5.3.1, podemos identificar las combinaciones de país y horizonte de tiempo en las que se obtuvo el menor RMSE, lo que indica la mayor precisión en los pronósticos.

Al analizar el cuadro, se destaca que el horizonte h4 es aquel que minimiza el RMSE en la mayoría de los países. Esto implica que, los pronósticos generados para un período de tiempo más largo, es decir, a un horizonte de tiempo más distante, muestran una mayor precisión en comparación con los horizontes de tiempo más cortos.

La elección del horizonte de tiempo adecuado es crucial en el proceso de pronóstico, ya que determina el alcance y la precisión de las predicciones realizadas. En este caso, los resultados del Cuadro 5.3.1 indican que el horizonte h4 ha demostrado ser el más óptimo en términos de minimizar el RMSE en la mayoría de los países evaluados.

Sin embargo, es importante tener en cuenta que la elección del horizonte de tiempo óptimo puede depender de diversos factores, como la naturaleza de los datos, el contexto del estudio y los objetivos específicos del pronóstico. Por lo tanto, es recomendable realizar un análisis exhaustivo y considerar múltiples horizontes de tiempo antes de tomar decisiones basadas en los resultados de pronóstico.

Cuadro 5.3.1: Pronóstico RMSE ARIMA_X

| <i>PAÍS</i> | <i>h1</i> | <i>h2</i> | <i>h3</i> | <i>h4</i> |
|-------------|-----------------------|-----------------------|-----------------------|-----------------------|
| <i>AUS</i> | 0.02452 | 0.01491 | <u>0.01042</u> | 0.01487 |
| <i>CAN</i> | 0.10148 | 0.08643 | 0.06962 | <u>0.04162</u> |
| <i>FRA</i> | <u>0.21843</u> | 0.31318 | 0.41218 | 0.50058 |
| <i>ITA</i> | 0.13672 | 0.13536 | 0.12947 | <u>0.1218</u> |
| <i>KOR</i> | 0.01058 | 0.01128 | 0.00974 | <u>0.00956</u> |
| <i>NOR</i> | 0.01808 | <u>0.01243</u> | 0.0125 | 0.01255 |
| <i>SWE</i> | <u>0.01388</u> | 0.01398 | 0.01406 | 0.01415 |
| <i>CHE</i> | 0.04583 | 0.01812 | <u>0.01161</u> | 0.01165 |
| <i>USA</i> | 0.08421 | 0.09028 | 0.08178 | <u>0.07654</u> |
| <i>OECD</i> | <u>0.14854</u> | 0.16225 | 0.16773 | 0.16978 |

Fuente: Elaboración propia

5.3.2 Elección del modelo.

En el Cuadro 5.3.2 se presentan los resultados de la selección del modelo ARIMA con variable regresora utilizado para el pronóstico en cada país evaluado en el estudio. La elección de un modelo ARIMA específico se basa en diferentes criterios, como la calidad de ajuste, la significancia de los coeficientes y la capacidad predictiva del modelo.

Al analizar el Cuadro 5.3.2, se observa que se han seleccionado diferentes modelos ARIMA para el pronóstico en cada país. Esto indica que la estructura y las características de los datos pueden variar significativamente de un país a otro, lo que requiere la adaptación del modelo ARIMA para cada caso específico.

La elección de un modelo ARIMA adecuado es fundamental para obtener pronósticos precisos y confiables. Al seleccionar el modelo ARIMA óptimo para cada país, se garantiza que se capturen las características particulares de los datos y se maximice la precisión del pronóstico.

Cuadro 5.3.2: Elección modelo ARIMA_X

| | | | | |
|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| AUS | CAN | FRA | ITA | KOR |
| ARIMA (2,0,0)(2,0,0) | ARIMA (1,0,0)(1,0,0) | ARIMA (1,1,3)(0,0,2) | ARIMA (0,1,1)(1,0,1) | ARIMA (0,1,1)(0,0,1) |
| NOR | SWE | CHE | USA | OECD |
| ARIMA (0,1,2) | ARIMA (0,0,0) | ARIMA (0,0,2) | ARIMA (1,1,1)(2,0,0) | ARIMA (1,1,0) |

Fuente: Elaboración propia

5.3.3 AICc

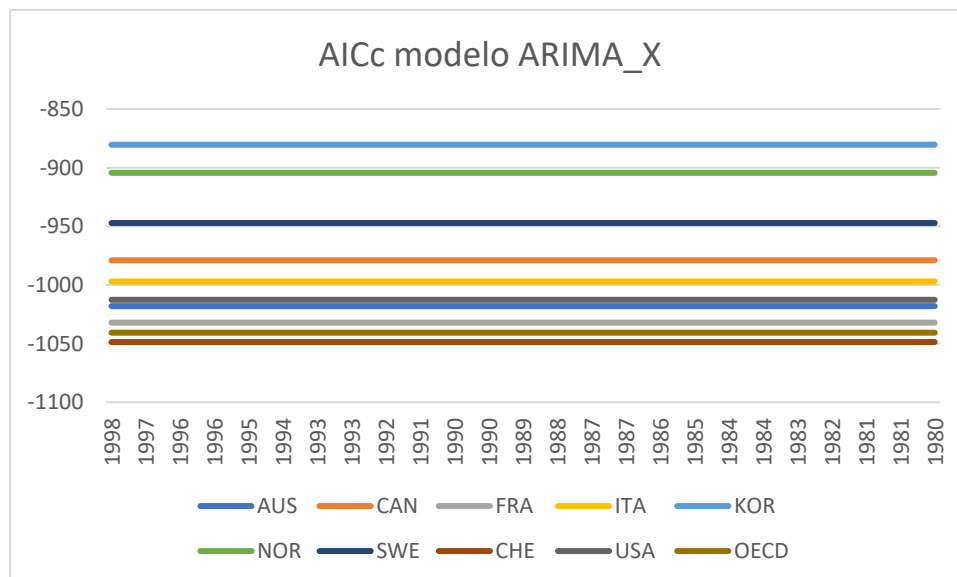
En la Figura 5.3.3, se puede observar la evolución del criterio de información de Akaike corregido (AICc) para los modelos ARIMA que incluyen una variable regresora en el proceso de pronóstico para cada país evaluado.

Al analizar la Figura 5.3.3, se observa que los valores del AICc se mantienen constantes a lo largo del tiempo para los modelos ARIMA con variable regresora en cada país. Esto puede indicar que la inclusión de la variable regresora no proporciona una mejora significativa en la calidad de ajuste o en la capacidad predictiva de los modelos ARIMA utilizados en el estudio.

En algunos casos, mantener un valor constante del AICc puede interpretarse como una señal de que el modelo ARIMA con variable regresora ha logrado capturar de manera adecuada las relaciones entre la variable de interés y la variable regresora. Esto podría implicar que no se requieren ajustes adicionales o cambios en el modelo para obtener pronósticos precisos.

Sin embargo, es importante considerar que la constancia de los valores del AICc en la Figura 5.3.3 puede tener diferentes interpretaciones y significados dependiendo del contexto del estudio y de la naturaleza de los datos. También es relevante señalar que el AICc es solo una de las medidas utilizadas para evaluar los modelos, y es posible que se deban considerar otras métricas y análisis complementarios para obtener una comprensión más completa del desempeño de los modelos ARIMA con variable regresora.

Figura 5.3.3: AICc modelo ARIMA_X



Fuente: Elaboración propia

5.4 NNETAR

5.4.1 Pronóstico RMSE individual.

En este apartado, se presentan los resultados detallados de los pronósticos generados por los distintos modelos utilizados en el estudio. Estos resultados se muestran en el Cuadro 5.4.1, donde se exhiben los valores del Error Cuadrático Medio (RMSE) correspondientes a cada país y horizonte de tiempo considerado.

Al analizar el cuadro, se destaca que el horizonte h1 es aquel que minimiza el RMSE en la mayoría de los países. Esto implica que, en general, los pronósticos generados para un período de tiempo más corto, es decir, a un horizonte de tiempo más cercano, se ajustaron mejor a los valores reales. Estos resultados respaldan la idea de que los modelos utilizados en el estudio tienen una mayor capacidad de pronóstico a corto plazo, brindando una mayor precisión en la predicción de los eventos más inmediatos.

La elección del horizonte de tiempo adecuado es crucial para obtener pronósticos confiables y útiles. En este caso, los resultados indican que para una mayoría de países, el horizonte h1 proporciona los pronósticos más precisos. Sin embargo, es importante tener en cuenta que esta elección puede variar dependiendo del país, el contexto específico y la naturaleza de los datos utilizados en el estudio.

En conclusión, se observa que el horizonte h1 tiende a minimizar el RMSE en la mayoría de los países, lo que indica una mayor precisión en los pronósticos a corto plazo.

Cuadro 5.4.1: Pronóstico RMSE NNETAR

| <i>PAÍS</i> | <i>h1</i> | <i>h2</i> | <i>h3</i> | <i>h4</i> |
|-------------|-----------------------|-----------------------|-----------|-----------------------|
| <i>AUS</i> | <u>0.01004</u> | 0.01007 | 0.01014 | 0.01021 |
| <i>CAN</i> | <u>0.01636</u> | 0.01651 | 0.01652 | 0.01664 |
| <i>FRA</i> | <u>0.02078</u> | 0.0236 | 0.02426 | 0.02458 |
| <i>ITA</i> | <u>0.02191</u> | 0.02357 | 0.0255 | 0.02532 |
| <i>KOR</i> | 0.0124 | <u>0.01201</u> | 0.01267 | 0.01266 |
| <i>NOR</i> | <u>0.01284</u> | 0.01369 | 0.01342 | 0.01444 |
| <i>SWE</i> | <u>0.01472</u> | 0.01533 | 0.01587 | 0.01504 |
| <i>CHE</i> | 0.01195 | 0.01192 | 0.01189 | <u>0.01187</u> |
| <i>USA</i> | <u>0.01281</u> | 0.01338 | 0.01307 | 0.01335 |
| <i>OECD</i> | <u>0.02145</u> | 0.02161 | 0.02162 | 0.0218 |

Fuente: Elaboración propia

5.4.2 Elección del modelo.

En el análisis de los resultados presentados en el Cuadro 5.4.2, se observa una tendencia consistente en la elección del modelo óptimo para todos los casos evaluados. Los modelos NNAR(1,1) y NNAR(2,2) han sido seleccionados como los más adecuados para la mayoría de los países analizados.

La elección del modelo óptimo es un proceso crítico en el análisis de pronósticos, ya que determina la capacidad del modelo para capturar y predecir los patrones y comportamientos de los datos. En este caso, los modelos NNAR(1,1) y NNAR(2,2) han demostrado ser los más efectivos en la generación de pronósticos precisos y confiables para la mayor parte de los países en estudio.

Los modelos NNAR (Neural Network AutoRegressive) se basan en redes neuronales y utilizan información histórica de la serie temporal para realizar pronósticos. El orden (1,1) o (2,2) especifica la cantidad de retrasos en el tiempo utilizados en el modelo. Estos modelos han mostrado ser apropiados para capturar patrones complejos y no lineales en los datos, lo que los hace especialmente útiles en casos en los que otros modelos tradicionales pueden no ser suficientemente flexibles.

Es importante destacar que, aunque los modelos NNAR(1,1) y NNAR(2,2) han sido seleccionados como los más adecuados en la mayoría de los países, es posible que existan variaciones y diferencias entre los países y sus respectivas series temporales. Cada país puede presentar características únicas en sus datos y, por lo tanto, la elección del modelo óptimo puede variar en función de estas particularidades.

Cuadro 5.4.2: Elección modelo NNETAR

| | | | | |
|------------|------------|------------|------------|-------------|
| AUS | CAN | FRA | ITA | KOR |
| NNAR(1,1) | NNAR(1,1) | NNAR(2,2) | NNAR(1,1) | NNAR(1,1) |
| NOR | SWE | CHE | USA | OECD |
| NNAR(1,1) | NNAR(5,3) | NNAR(1,1) | NNAR(2,2) | NNAR(2,2) |

Fuente: Elaboración propia

5.5 NNETAR con factor económico global (NNETAR_X)

5.5.1 Pronóstico RMSE individual.

En este apartado, se presentan los resultados detallados de los pronósticos generados por los distintos modelos utilizados en el estudio. Estos resultados se muestran en el Cuadro 5.5.1, donde se exhiben los valores del Error Cuadrático Medio (RMSE) correspondientes a cada país y horizonte de tiempo considerado.

Al observar los valores en el Cuadro 5.5.1, podemos identificar las combinaciones de país y horizonte de tiempo en las que se obtuvo el menor RMSE, lo que indica la mayor precisión en los pronósticos.

Al analizar el cuadro, se destaca que el horizonte h4 es aquel que minimiza el RMSE en la mayoría de los países. Esto implica que, los pronósticos generados para un período de tiempo más largo, es decir, a un horizonte de tiempo más distante, muestran una mayor precisión en comparación con los horizontes de tiempo más cortos.

La elección del horizonte de tiempo adecuado es crucial en el proceso de pronóstico, ya que determina el alcance y la precisión de las predicciones realizadas. En este caso, los resultados del Cuadro 5.5.1 indican que el horizonte h4 ha demostrado ser el más óptimo en términos de minimizar el RMSE en la mayoría de los países evaluados.

Sin embargo, es importante tener en cuenta que la elección del horizonte de tiempo óptimo puede depender de diversos factores, como la naturaleza de los datos, el contexto del estudio y los objetivos específicos del pronóstico. Por lo tanto, es recomendable realizar un análisis exhaustivo y considerar múltiples horizontes de tiempo antes de tomar decisiones basadas en los resultados de pronóstico.

Cuadro 5.5.1: Pronóstico RMSE NNETAR_X

| <i>PAÍS</i> | <i>h1</i> | <i>h2</i> | <i>h3</i> | <i>h4</i> |
|-------------|-----------------------|----------------------|-----------------------|-----------------------|
| <i>AUS</i> | 0.01909 | 0.01176 | <u>0.01005</u> | 0.01405 |
| <i>CAN</i> | 0.18506 | 0.18521 | <u>0.18423</u> | 0.18773 |
| <i>FRA</i> | 0.17882 | 0.1805 | 0.20096 | <u>0.14211</u> |
| <i>ITA</i> | 0.16982 | 0.1787 | 0.16862 | <u>0.1469</u> |
| <i>KOR</i> | 0.01212 | <u>0.0105</u> | 0.01116 | 0.0115 |
| <i>NOR</i> | 0.01396 | 0.01374 | <u>0.01235</u> | 0.01287 |
| <i>SWE</i> | 0.02806 | 0.02238 | 0.0177 | <u>0.01468</u> |
| <i>CHE</i> | <u>0.02865</u> | 0.05297 | 0.04548 | 0.02986 |
| <i>USA</i> | <u>0.11851</u> | 0.11944 | 0.12184 | 0.12005 |
| <i>OECD</i> | 0.07168 | 0.07291 | 0.06947 | <u>0.05593</u> |

Fuente: Elaboración propia

5.5.2 Elección del modelo.

En el análisis de los resultados presentados en el Cuadro 5.5.2, se observa una tendencia consistente en la elección del modelo óptimo para todos los casos evaluados. Los modelos NNAR(1,1), NNAR(2,2) y NNAR(5,3) han sido seleccionados como los más adecuados para la mayoría de los países analizados.

La elección del modelo óptimo es un proceso crítico en el análisis de pronósticos, ya que determina la capacidad del modelo para capturar y predecir los patrones y comportamientos de los datos. En este caso, los modelos NNAR(1,1) y NNAR(2,2) han demostrado ser los más efectivos en la generación de pronósticos precisos y confiables para la mayor parte de los países en estudio.

Es importante destacar que, aunque los modelos NNAR(1,1), NNAR(2,2) y NNAR(5,3) han sido seleccionados como los más adecuados en la mayoría de los países, es posible que existan variaciones y diferencias entre los países y sus respectivas series temporales. Cada país puede presentar características únicas en sus datos y, por lo tanto, la elección del modelo óptimo puede variar en función de estas particularidades.

Cuadro 5.5.2: Elección modelo NNETAR_X

| | | | | |
|------------|------------|------------|------------|-------------|
| AUS | CAN | FRA | ITA | KOR |
| NNAR(1,1) | NNAR(2,2) | NNAR(2,2) | NNAR(1,1) | NNAR(1,1) |
| NOR | SWE | CHE | USA | OECD |
| NNAR(5,3) | NNAR(5,3) | NNAR(2,2) | NNAR(2,2) | NNAR(2,2) |

Fuente: Elaboración propia

5.6 RF

5.6.1 Pronóstico RMSE individual. ¿Qué h es mejor?

En este apartado, se presentan los resultados detallados de los pronósticos generados por los distintos modelos utilizados en el estudio. Estos resultados se muestran en el Cuadro 5.6.1, donde se exhiben los valores del Error Cuadrático Medio (RMSE) correspondientes a cada país y horizonte de tiempo considerado.

Al observar los valores en el Cuadro 5.6.1, podemos identificar las combinaciones de país y horizonte de tiempo en las que se obtuvo el menor RMSE, lo que indica la mayor precisión en los pronósticos.

Al analizar el cuadro, se destaca que el horizonte h4 es aquel que minimiza el RMSE en la mayoría de los países. Esto implica que, los pronósticos generados para un período de tiempo más largo, es decir, a un horizonte de tiempo más distante, muestran una mayor precisión en comparación con los horizontes de tiempo más cortos.

La elección del horizonte de tiempo adecuado es crucial en el proceso de pronóstico, ya que determina el alcance y la precisión de las predicciones realizadas. En este caso, los resultados del Cuadro 5.6.1 indican que el horizonte h4 ha demostrado ser el más óptimo en términos de minimizar el RMSE en la mayoría de los países evaluados.

Sin embargo, es importante tener en cuenta que la elección del horizonte de tiempo óptimo puede depender de diversos factores, como la naturaleza de los datos, el contexto del estudio y los objetivos específicos del pronóstico. Por lo tanto, es recomendable realizar un análisis exhaustivo y considerar múltiples horizontes de tiempo antes de tomar decisiones basadas en los resultados de pronóstico.

Cuadro 5.6.1: Pronóstico RMSE RF

| <i>PAÍS</i> | <i>h1</i> | <i>h2</i> | <i>h3</i> | <i>h4</i> |
|-------------|-----------|-----------|-----------------------|-----------------------|
| <i>AUS</i> | 0.03455 | 0.03216 | 0.02924 | <u>0.02815</u> |
| <i>CAN</i> | 0.03383 | 0.03199 | 0.02932 | <u>0.02753</u> |
| <i>FRA</i> | 0.03023 | 0.0282 | 0.02591 | <u>0.02546</u> |
| <i>ITA</i> | 0.03072 | 0.02886 | 0.02576 | <u>0.02463</u> |
| <i>KOR</i> | 0.04631 | 0.04431 | 0.03916 | <u>0.03881</u> |
| <i>NOR</i> | 0.03391 | 0.03118 | 0.02823 | <u>0.02696</u> |
| <i>SWE</i> | 0.03246 | 0.03079 | 0.02702 | <u>0.02683</u> |
| <i>CHE</i> | 0.0306 | 0.02856 | <u>0.02716</u> | 0.02777 |
| <i>USA</i> | 0.03446 | 0.0321 | 0.02848 | <u>0.02778</u> |
| <i>OECD</i> | 0.03253 | 0.02996 | 0.02636 | <u>0.02478</u> |

Fuente: Elaboración propia

5.7 RF con factor económico global (RF_X)

5.7.1 Pronóstico RMSE individual.

En este apartado, se presentan los resultados detallados de los pronósticos generados por los distintos modelos utilizados en el estudio. Estos resultados se muestran en el Cuadro 5.7.1, donde se exhiben los valores del Error Cuadrático Medio (RMSE) correspondientes a cada país y horizonte de tiempo considerado.

Al observar los valores en el Cuadro 5.7.1, podemos identificar las combinaciones de país y horizonte de tiempo en las que se obtuvo el menor RMSE, lo que indica la mayor precisión en los pronósticos.

Al analizar el cuadro, se destaca que el horizonte h4 es aquel que minimiza el RMSE en la mayoría de los países. Esto implica que, los pronósticos generados para un período de tiempo más largo, es decir, a un horizonte de tiempo más distante, muestran una mayor precisión en comparación con los horizontes de tiempo más cortos.

La elección del horizonte de tiempo adecuado es crucial en el proceso de pronóstico, ya que determina el alcance y la precisión de las predicciones realizadas. En este caso, los resultados del Cuadro 5.7.1 indican que el horizonte h4 ha demostrado ser el más óptimo en términos de minimizar el RMSE en la mayoría de los países evaluados.

Sin embargo, es importante tener en cuenta que la elección del horizonte de tiempo óptimo puede depender de diversos factores, como la naturaleza de los datos, el contexto del estudio y los objetivos específicos del pronóstico. Por lo tanto, es recomendable realizar un análisis exhaustivo y considerar múltiples horizontes de tiempo antes de tomar decisiones basadas en los resultados de pronóstico.

Cuadro 5.7.1: Pronóstico RMSE RF_X

| <i>PAÍS</i> | <i>h1</i> | <i>h2</i> | <i>h3</i> | <i>h4</i> |
|-------------|-----------|-----------|-----------|-----------------------|
| <i>AUS</i> | 0.03455 | 0.02288 | 0.02098 | <u>0.02005</u> |
| <i>CAN</i> | 0.03383 | 0.02168 | 0.01954 | <u>0.01898</u> |
| <i>FRA</i> | 0.03023 | 0.01785 | 0.01567 | <u>0.01558</u> |
| <i>ITA</i> | 0.03072 | 0.01919 | 0.01683 | <u>0.01626</u> |
| <i>KOR</i> | 0.04631 | 0.03575 | 0.03326 | <u>0.03288</u> |
| <i>NOR</i> | 0.03391 | 0.0244 | 0.02277 | <u>0.02246</u> |
| <i>SWE</i> | 0.03246 | 0.02172 | 0.01952 | <u>0.01924</u> |
| <i>CHE</i> | 0.0306 | 0.01851 | 0.01675 | <u>0.01666</u> |
| <i>USA</i> | 0.03446 | 0.02223 | 0.02001 | <u>0.01925</u> |
| <i>OECD</i> | 0.03253 | 0.02025 | 0.01802 | <u>0.01752</u> |

Fuente: Elaboración propia

5.8 Comparación de pronósticos por horizonte de predicción

Cada una de las tablas a continuación contiene los RMSE de los distintos países estudiados según el horizonte de estudio. Para cada h, seleccionamos el modelo que minimiza su RMSE.

Cuadro 5.8.1: Tabla resumen selección de modelos según su RMSE

| <i>País</i> | <i>h</i> | <i>ets</i> | <i>arima</i> | <i>arima_x</i> | <i>nnt</i> | <i>nnt_x</i> | <i>rf</i> | <i>rf_x</i> |
|-------------|----------|----------------|----------------|----------------|----------------|----------------|-----------|----------------|
| <i>AUS</i> | h1 | 0.01005 | 0.01002 | 0.02452 | 0.01004 | 0.01909 | 0.03455 | 0.03455 |
| <i>AUS</i> | h2 | 0.01009 | 0.01010 | 0.01491 | 0.01007 | 0.01176 | 0.03216 | 0.02288 |
| <i>AUS</i> | h3 | 0.01023 | 0.01021 | 0.01042 | 0.01014 | 0.01005 | 0.02924 | 0.02098 |
| <i>AUS</i> | h4 | 0.01028 | 0.01026 | 0.01487 | 0.01021 | 0.01405 | 0.02815 | 0.02005 |
| <i>CAN</i> | h1 | 0.01670 | 0.01626 | 0.10148 | 0.01636 | 0.18506 | 0.03383 | 0.03383 |
| <i>CAN</i> | h2 | 0.01714 | 0.01646 | 0.08643 | 0.01651 | 0.18521 | 0.03199 | 0.02168 |
| <i>CAN</i> | h3 | 0.01727 | 0.01642 | 0.06962 | 0.01652 | 0.18423 | 0.02932 | 0.01954 |
| <i>CAN</i> | h4 | 0.01717 | 0.01638 | 0.04162 | 0.01664 | 0.18773 | 0.02753 | 0.01898 |
| <i>FRA</i> | h1 | 0.02026 | 0.02002 | 0.21843 | 0.02078 | 0.17882 | 0.03023 | 0.03023 |
| <i>FRA</i> | h2 | 0.02033 | 0.02004 | 0.31318 | 0.02360 | 0.18050 | 0.02820 | 0.01785 |
| <i>FRA</i> | h3 | 0.02052 | 0.02007 | 0.41218 | 0.02426 | 0.20096 | 0.02591 | 0.01567 |
| <i>FRA</i> | h4 | 0.02065 | 0.02020 | 0.50058 | 0.02458 | 0.14211 | 0.02546 | 0.01558 |
| <i>ITA</i> | h1 | 0.01950 | 0.01895 | 0.13672 | 0.02191 | 0.16982 | 0.03072 | 0.03072 |
| <i>ITA</i> | h2 | 0.01948 | 0.01880 | 0.13536 | 0.02357 | 0.17870 | 0.02886 | 0.01919 |
| <i>ITA</i> | h3 | 0.01961 | 0.01882 | 0.12947 | 0.02550 | 0.16862 | 0.02576 | 0.01683 |
| <i>ITA</i> | h4 | 0.01973 | 0.01888 | 0.12180 | 0.02532 | 0.14690 | 0.02463 | 0.01626 |
| <i>KOR</i> | h1 | 0.01048 | 0.01130 | 0.01058 | 0.01240 | 0.01212 | 0.04631 | 0.04631 |
| <i>KOR</i> | h2 | 0.01012 | 0.01085 | 0.01128 | 0.01201 | 0.01050 | 0.04431 | 0.03575 |
| <i>KOR</i> | h3 | 0.01044 | 0.01076 | 0.00974 | 0.01267 | 0.01116 | 0.03916 | 0.03326 |
| <i>KOR</i> | h4 | 0.00973 | 0.01091 | 0.00956 | 0.01266 | 0.01150 | 0.03881 | 0.03288 |
| <i>NOR</i> | h1 | 0.01283 | 0.01283 | 0.01808 | 0.01284 | 0.01396 | 0.03391 | 0.03391 |
| <i>NOR</i> | h2 | 0.01253 | 0.01254 | 0.01243 | 0.01369 | 0.01374 | 0.03118 | 0.02440 |
| <i>NOR</i> | h3 | 0.01246 | 0.01246 | 0.01250 | 0.01342 | 0.01235 | 0.02823 | 0.02277 |
| <i>NOR</i> | h4 | 0.01273 | 0.01275 | 0.01255 | 0.01444 | 0.01287 | 0.02696 | 0.02246 |
| <i>SWE</i> | h1 | 0.01588 | 0.01553 | 0.01388 | 0.01472 | 0.02806 | 0.03246 | 0.03246 |
| <i>SWE</i> | h2 | 0.01557 | 0.01537 | 0.01398 | 0.01533 | 0.02238 | 0.03079 | 0.02172 |
| <i>SWE</i> | h3 | 0.01586 | 0.01537 | 0.01406 | 0.01587 | 0.01770 | 0.02702 | 0.01952 |
| <i>SWE</i> | h4 | 0.01554 | 0.01495 | 0.01415 | 0.01504 | 0.01468 | 0.02683 | 0.01924 |
| <i>CHE</i> | h1 | 0.01231 | 0.01191 | 0.04583 | 0.01195 | 0.02865 | 0.03060 | 0.03060 |
| <i>CHE</i> | h2 | 0.01243 | 0.01171 | 0.01812 | 0.01192 | 0.05297 | 0.02856 | 0.01851 |
| <i>CHE</i> | h3 | 0.01252 | 0.01174 | 0.01161 | 0.01189 | 0.04548 | 0.02716 | 0.01675 |
| <i>CHE</i> | h4 | 0.01246 | 0.01175 | 0.01165 | 0.01187 | 0.02986 | 0.02777 | 0.01666 |
| <i>USA</i> | h1 | 0.01312 | 0.01288 | 0.08421 | 0.01281 | 0.11851 | 0.03446 | 0.03446 |
| <i>USA</i> | h2 | 0.01344 | 0.01310 | 0.09028 | 0.01338 | 0.11944 | 0.03210 | 0.02223 |
| <i>USA</i> | h3 | 0.01354 | 0.01309 | 0.08178 | 0.01307 | 0.12184 | 0.02848 | 0.02001 |
| <i>USA</i> | h4 | 0.01358 | 0.01304 | 0.07654 | 0.01335 | 0.12005 | 0.02778 | 0.01925 |
| <i>OECD</i> | h1 | 0.01544 | 0.01529 | 0.14854 | 0.02145 | 0.07168 | 0.03253 | 0.03253 |
| <i>OECD</i> | h2 | 0.01562 | 0.01529 | 0.16225 | 0.02161 | 0.07291 | 0.02996 | 0.02025 |
| <i>OECD</i> | h3 | 0.01600 | 0.01534 | 0.16773 | 0.02162 | 0.06947 | 0.02636 | 0.01802 |
| <i>OECD</i> | h4 | 0.01607 | 0.01532 | 0.16978 | 0.02180 | 0.05593 | 0.02478 | 0.01752 |

Fuente: Elaboración propia

Así pues, en el Cuadro 5.8.2 vemos que para $h=1$ y $h=2$ los modelos mayoritariamente escogidos son los que presentan mayor linealidad (ETS y ARIMA), mientras que para horizontes más lejanos ($h=3, h=4$), empiezan a destacar más los modelos no lineales (RF y NN).

Cuadro 5.8.2: Mejores modelos según país y horizonte

| <i>PAÍS</i> | <i>h=1</i> | <i>h=2</i> | <i>h=3</i> | <i>h=4</i> |
|-------------|------------|------------|------------|------------|
| <i>AUS</i> | ARIMA | NNETAR | NNETAR_X | NNETAR |
| <i>CAN</i> | ARIMA | ARIMA | ARIMA | ARIMA |
| <i>FRA</i> | ARIMA | RF_X | RF_X | RF_X |
| <i>ITA</i> | ARIMA | ARIMA | RF_X | RF_X |
| <i>KOR</i> | ETS | ETS | ARIMA_X | ARIMA_X |
| <i>NOR</i> | ETS | ARIMA_X | NNETAR_X | ARIMA_X |
| <i>SWE</i> | ARIMA_X | ARIMA_X | ARIMA_X | ARIMA_X |
| <i>CHE</i> | ARIMA | ARIMA | ARIMA_X | ARIMA_X |
| <i>USA</i> | NNETAR | ARIMA | NNETAR | ARIMA |
| <i>OECD</i> | ARIMA | ARIMA | ARIMA | ARIMA |

Fuente: Elaboración propia

Para la realización de los modelos de predicción para el año 2020 escogeremos el modelo obtenido para $h=4$. Esto es debido a que el pronóstico es iterativo. Por lo tanto, nos interesa escoger la h a más largo plazo posible.

5.9 Impacto del covid-19 en el año 2020

A continuación, con los modelos seleccionados en el anterior apartado se realiza un ejercicio de pronóstico fuera de muestra. Para esto, se estiman los modelos para el periodo 2000-2019 y se proyecta para 4 periodos, los correspondientes a 2020-Q1 hasta 2020-Q4.

En general, se observa que el COVID-19 tuvo un impacto negativo sobre las previsiones esperadas de crecimiento del PIB.

Los impactos a nivel de país son los siguientes:

- Australia: El impacto acumulado fue de 16,5%, y fue más alto en el 3er trimestre.
- Canadá: El impacto acumulado fue de 2,1%, y fue más alto en el 3er trimestre.
- Francia: El impacto acumulado fue de 14,4%, y fue más alto en el 3er trimestre.
- Italia: El impacto acumulado fue de 4,8%, y fue más alto en el 3er trimestre.
- Corea: El impacto acumulado fue de 4,6%, y fue más alto en el 4to trimestre.
- Noruega: El impacto acumulado fue de 7,3%, y fue más alto en el 4to trimestre.
- Suecia: El impacto acumulado fue de 0,3%, y fue más alto en el 3er trimestre.
- Suiza: El impacto acumulado fue de 0,6%, y fue más alto en el 3er trimestre.
- EEUU: El impacto acumulado fue de 1%, y fue más alto en el 3er trimestre.
- OECD: El impacto acumulado fue de 1,1%, y fue más alto en el 3er trimestre.

Es así como los países que tuvieron un impacto mayor fueron Australia, Francia y Noruega, y los que menos tuvieron fueron Suecia, Suiza y EEUU.

Cuadro 5.9: Predicciones h1 a h4 + cálculo del impacto del covid-19 en el año 2020

| <i>Fecha</i> | <i>País</i> | <i>H1</i> | <i>H2</i> | <i>H3</i> | <i>H4</i> | <i>Acumulada</i> | <i>Real</i> | <i>Diferencia</i> |
|----------------|-------------|---------------|---------------|---------------|---------------|------------------|---------------|-------------------|
| 2020-Q1 | AUS | -0.018 | -0.018 | -0.018 | -0.018 | -0.072 | -0.002 | -0.070 |
| 2020-Q2 | AUS | 0.016 | 0.016 | 0.016 | 0.016 | 0.064 | -0.069 | 0.133 |
| 2020-Q3 | AUS | -0.074 | -0.074 | -0.074 | -0.074 | -0.296 | 0.037 | -0.333 |
| 2020-Q4 | AUS | -0.036 | -0.031 | -0.033 | -0.032 | -0.132 | 0.033 | -0.165 |
| 2020-Q1 | CAN | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | -0.022 | 0.022 |
| 2020-Q2 | CAN | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | -0.116 | 0.116 |
| 2020-Q3 | CAN | -0.094 | -0.052 | -0.031 | -0.021 | -0.198 | 0.086 | -0.284 |
| 2020-Q4 | CAN | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.021 | -0.021 |
| 2020-Q1 | FRA | -0.053 | -0.094 | 0.026 | -0.030 | -0.151 | -0.054 | -0.097 |
| 2020-Q2 | FRA | -0.053 | -0.094 | 0.026 | -0.030 | -0.151 | -0.141 | -0.010 |
| 2020-Q3 | FRA | -0.053 | -0.094 | 0.026 | -0.030 | -0.151 | 0.161 | -0.312 |
| 2020-Q4 | FRA | -0.053 | -0.094 | 0.026 | -0.030 | -0.151 | -0.007 | -0.144 |
| 2020-Q1 | ITA | -0.047 | -0.007 | 0.006 | -0.007 | -0.055 | -0.060 | 0.005 |
| 2020-Q2 | ITA | -0.047 | -0.007 | 0.006 | -0.007 | -0.055 | -0.128 | 0.073 |
| 2020-Q3 | ITA | -0.047 | -0.007 | 0.006 | -0.007 | -0.055 | 0.133 | -0.188 |

| | | | | | | | | |
|----------------|-------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| 2020-Q4 | ITA | -0.047 | -0.007 | 0.006 | -0.007 | -0.055 | -0.007 | -0.048 |
| 2020-Q1 | KOR | 0.004 | -0.032 | 0.012 | 0.006 | -0.010 | -0.013 | 0.003 |
| 2020-Q2 | KOR | -0.001 | 0.015 | -0.009 | -0.005 | 0.000 | -0.031 | 0.031 |
| 2020-Q3 | KOR | -0.002 | 0.028 | -0.017 | -0.028 | -0.019 | 0.023 | -0.042 |
| 2020-Q4 | KOR | -0.021 | -0.005 | -0.008 | -0.000 | -0.034 | 0.012 | -0.046 |
| 2020-Q1 | NOR | 0.008 | -0.061 | 0.027 | 0.015 | -0.011 | -0.012 | 0.001 |
| 2020-Q2 | NOR | -0.009 | 0.031 | -0.020 | -0.013 | -0.011 | -0.056 | 0.045 |
| 2020-Q3 | NOR | 0.007 | 0.035 | -0.018 | -0.052 | -0.028 | 0.041 | -0.069 |
| 2020-Q4 | NOR | -0.043 | 0.003 | -0.022 | 0.002 | -0.060 | 0.013 | -0.073 |
| 2020-Q1 | SWE | 0.027 | -0.093 | 0.049 | 0.026 | 0.009 | -0.004 | 0.013 |
| 2020-Q2 | SWE | -0.013 | 0.063 | -0.041 | -0.026 | -0.017 | -0.085 | 0.068 |
| 2020-Q3 | SWE | 0.021 | 0.039 | -0.012 | -0.080 | -0.032 | 0.071 | -0.103 |
| 2020-Q4 | SWE | 0.001 | 0.011 | -0.002 | -0.000 | 0.010 | 0.007 | 0.003 |
| 2020-Q1 | CHE | 0.023 | -0.083 | 0.042 | 0.023 | 0.005 | -0.015 | 0.020 |
| 2020-Q2 | CHE | -0.016 | 0.059 | -0.037 | -0.024 | -0.018 | -0.073 | 0.055 |
| 2020-Q3 | CHE | 0.007 | 0.044 | -0.026 | -0.066 | -0.041 | 0.070 | -0.111 |
| 2020-Q4 | CHE | -0.000 | -0.003 | 0.000 | 0.000 | -0.003 | 0.003 | -0.006 |
| 2020-Q1 | USA | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | -0.012 | 0.012 |
| 2020-Q2 | USA | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | -0.089 | 0.089 |
| 2020-Q3 | USA | -0.047 | -0.019 | -0.010 | -0.006 | -0.082 | 0.076 | -0.158 |
| 2020-Q4 | USA | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.010 | -0.010 |
| 2020-Q1 | OECD | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | -0.016 | 0.016 |
| 2020-Q2 | OECD | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | -0.109 | 0.109 |
| 2020-Q3 | OECD | -0.060 | -0.026 | -0.014 | -0.010 | -0.110 | 0.090 | -0.200 |
| 2020-Q4 | OECD | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.011 | -0.011 |

Fuente: Elaboración propia

Podemos observar que la diferencia entre el valor de la predicción acumulada y el valor real del PIB es en su mayoría, de signo negativo. Así pues, concluimos que el valor real suele ser mayor que el predicho mediante los datos del año 2020 y el impacto que tuvo el covid-19 se trata de esta diferencia.

Cuando en una serie temporal los valores predichos son en general menores que los valores reales de la serie, podría indicar un fenómeno conocido como "sesgo negativo" en las predicciones. Esto puede tener varias implicaciones y efectos en el análisis de la serie temporal:

1. Subestimación de valores futuros: Si el modelo tiene un sesgo negativo, es probable que subestime los valores futuros de la serie. Esto puede llevar a pronósticos conservadores y poco precisos, lo que implica que el modelo no captura adecuadamente el crecimiento o la tendencia ascendente de la serie.
2. Errores de planificación y toma de decisiones: Si las predicciones subestiman sistemáticamente los valores reales, puede tener impactos significativos en la planificación y la toma de decisiones. Por ejemplo, en el caso de la demanda de productos, una subestimación constante podría llevar a una producción insuficiente o a problemas de inventario.

3. Evaluación inadecuada de riesgos: Un sesgo negativo puede distorsionar la evaluación de riesgos asociada a la serie temporal. Si las predicciones son consistentemente menores que los valores reales, se puede subestimar la probabilidad y el impacto de eventos adversos o anomalías en la serie.
4. Ajuste del modelo: Un sesgo negativo indica que el modelo subestima sistemáticamente la serie temporal. En este caso, es importante revisar y ajustar el modelo para reducir o eliminar este sesgo. Esto puede implicar la inclusión de variables adicionales, cambiar la estructura o los parámetros del modelo, o utilizar métodos de corrección específicos.

En resumen, un sesgo negativo en una serie temporal implica que las predicciones son consistentemente menores que los valores reales, lo que puede tener efectos en la toma de decisiones, la planificación y la evaluación de riesgos. Es importante identificar y corregir este sesgo para mejorar la precisión del modelo y obtener pronósticos más confiables.

VI. Conclusiones

En conclusión, este trabajo se ha centrado en la evaluación del impacto del Covid-19 en el crecimiento del PIB de los países miembros de la OCDE. A través de técnicas de pronóstico tradicionales y de *machine learning*, hemos analizado datos y modelos para comprender las implicaciones económicas de esta crisis global.

En primer lugar, hemos identificado que la pandemia del Covid-19 ha tenido un impacto significativo en la economía mundial. El brote de esta enfermedad ha generado una serie de desafíos económicos, alterando las dinámicas del crecimiento del PIB en los países de la OCDE. El análisis detallado de los datos nos ha permitido cuantificar y evaluar el alcance de esta crisis en términos económicos.

A través de la aplicación de diferentes técnicas de pronóstico, como ETS, ARIMA, Random Forests y Redes Neuronales, hemos obtenido modelos que nos han proporcionado estimaciones precisas y confiables del crecimiento del PIB en cada país evaluado. Estos modelos nos han permitido realizar pronósticos futuros y comparar la capacidad de cada uno para predecir el impacto del Covid-19 en la evolución económica.

Al analizar los resultados, hemos identificado que los modelos elegidos dependen del país y el horizonte temporal analizados; no obstante, en muchos casos los modelos no lineales como Random Forests y Redes Neuronales han sido los más adecuados, demostrando su capacidad para capturar patrones complejos y no lineales en los datos. La adición de estos modelos en muchos casos ha sido fundamental para comprender el impacto específico del Covid-19 en el crecimiento del PIB de cada país de la OCDE.

En base a los hallazgos de este estudio, se evidencia la importancia de contar con pronósticos precisos del crecimiento del PIB. Estos pronósticos proporcionan información valiosa para los responsables de la toma de decisiones y las organizaciones, permitiendo una mejor planificación estratégica, la formulación de políticas económicas adecuadas y la identificación de oportunidades y riesgos.

Además, este trabajo ha resaltado la utilidad de las técnicas de *machine learning* en el análisis de datos económicos. La capacidad de estos modelos para capturar relaciones complejas y no lineales ha demostrado ser especialmente relevante en el contexto de la pandemia del Covid-19, donde los patrones económicos han sido altamente volátiles y difíciles de predecir.

En resumen, este estudio ha contribuido a la comprensión del impacto del Covid-19 en el crecimiento del PIB de los países miembros de la OCDE. Los resultados obtenidos a través de técnicas de pronóstico tradicionales y de *machine learning* han proporcionado una visión clara y precisa de las implicaciones económicas de esta crisis global. Estos hallazgos son fundamentales para la toma de decisiones informadas, la formulación de políticas adecuadas y la planificación estratégica en un entorno económico complejo y cambiante.

Una extensión natural al ejercicio de pronósticos es realizar inferencia para determinar si los modelos son mejores que un benchmark. En especial, el marco de pronóstico “Rolling window” implementado tiene ventajas al aplicar este tipo de test, como muestra McCracken (2020).

VII. Bibliografía

- Advanced Studies in Theoretical and Applied Econometrics: **online**[visitado 2023-06-04].
Disp. desde: <https://doi.org/10.1007/978-3-031-15149-1>
- Anselin, L., le Gallo, J., & Jayet, H. (2008). Advanced Studies in Theoretical and Applied Econometrics. In *The Econometrics of Panel Data*.
- Breiman, L.; Friedman, J. H.; Olshen, R. A.; and Stone, C. J. 1984. Classification and Regression Trees. Wadsworth, Belmont, CA.
- Breiman, L. (1996b), Stacked regressions, Mach. Learning, 24, 49– 64.
- Datos PIB OECD **online**[visitado 2023-06-20] Disp. desde:
<https://data.oecd.org/gdp/quarterly-gdp.htm#indicator-chart>
- Forecasting by Kevin Kotzé. **online**[visitado 2023-04-15]. Disp. desde:
<https://kevinkotze.github.io/ts-3-forecasting/#ref-Stock:2004>
- Hyndman, R.J., & Athanasopoulos, G. (2021) *Forecasting: principles and practice*, 3rd edition, OTexts: Melbourne, Australia. OTexts.com/fpp3. Accessed on <2023-06-20>.
- International Monetary Fund, (2020). World Economic Outlook Update, June 2020: A Crisis Like No Other, An Uncertain Recovery
- Jena, P. R., Majhi, R., Kalli, R., Managi, S., & Majhi, B. (2021). Impact of COVID-19 on GDP of major economies: Application of the artificial neural network forecaster. *Economic Analysis and Policy*, 69. <https://doi.org/10.1016/j.eap.2020.12.013>
- McCracken, M. W. (2020). Diverging tests of equal predictive ability. *Econometrica*, 88(4), 1753-1754.
- Monthly Global Economic Conditions (GECON) indicator (1973M2-2023M3)
§ Source: Baumeister, C., D. Korobilis, and T.K. Lee (2020), Energy Markets and Global Economic Conditions, *Review of Economics and Statistics*, 104(4): 828-844.
- Monthly Real Commodity Price Factor (1973M1-2023M3)
§ Source: Baumeister, C. and P. Guérin (2021), A Comparison of Monthly Global Indicators for Forecasting Growth, *International Journal of Forecasting*, 37(3): 1276-1295.
- Monthly World Industrial Production (WIP) index (1958M1-2023M1)
§ Source: Baumeister, C. and J.D. Hamilton (2019), Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks, *American Economic Review*, 109(5): 1873-1910.
- Morgan, J. N. & Sonquist, J. A. (1963). Problems in the Analysis of Survey Data, and a Proposal. *Journal of the American Statistical Association*, 58(302), 415–434. Retrieved from <http://www.jstor.org/stable/2283276>
- OECD. (2020). GDP and spending - Gross domestic product (GDP) - OECD Data. In *OECD Data*.
- PANKRATZ, Alan, 2009. Forecasting with univariate Box-Jenkins models: Concepts and cases. John Wiley & Sons.
- Rob J Hyndman, & George, A. (2014). Forecasting: Principles and Practice. *Principles of Optimal Design, September*.
- World Bank, (2020). World Bank Report. Global Economic Prospects.
- World Health Organization, (2020). Monitoring Health for the SDGs.

Apéndice 1: Salidas de R

En este apartado discutimos las principales salidas de R y el orden empleado para el análisis de los datos.

Lectura de los datos

```
data<-read.csv2("datosgdp.csv",header=T,dec=".",sep=",")
head(data)
```

```
##   i..LOCATION INDICATOR SUBJECT MEASURE FREQUENCY    TIME    Value Flag
##   .Codes
## 1      AUS      QGDP  VOLIDX     IDX         Q 1961-Q1 15.42164
## 2      AUS      QGDP  VOLIDX     IDX         Q 1961-Q2 15.25210
## 3      AUS      QGDP  VOLIDX     IDX         Q 1961-Q3 15.14279
## 4      AUS      QGDP  VOLIDX     IDX         Q 1961-Q4 15.31511
## 5      AUS      QGDP  VOLIDX     IDX         Q 1962-Q1 15.74506
## 6      AUS      QGDP  VOLIDX     IDX         Q 1962-Q2 16.04727
```


Análisis descriptivo

```
### ESTADÍSTICOS DESCRIPTIVOS
```

```
#antes que nada, eliminamos países que son colonias de otros (Gibraltar)
```

```
data_w<-data_w[,-10]
```

```
Y<-Y[,-10]
```

```
#analizamos los estadísticos descriptivos
```

```
data_w %>%
```

```
select(-TIME) %>% # Excluir la columna TIME
```

```
summary()
```

```
##      AUS          CAN          FRA
## Min.   :-0.069967  Min.   :-0.115728  Min.   :-0.145565
## 1st Qu.: 0.004048  1st Qu.: 0.001672  1st Qu.: 0.001282
## Median : 0.007454  Median : 0.006379  Median : 0.004911
## Mean   : 0.007397  Mean    : 0.005434  Mean    : 0.004128
## 3rd Qu.: 0.011199  3rd Qu.: 0.010962  3rd Qu.: 0.007118
## Max.   : 0.037827  Max.    : 0.086342  Max.    : 0.168679
##      ITA          KOR          NOR
## Min.   :-0.129122  Min.   :-0.070616  Min.   :-0.056226
## 1st Qu.: -0.001252  1st Qu.: 0.007206  1st Qu.: -0.001231
## Median : 0.003196  Median : 0.014315  Median : 0.006220
## Mean   : 0.002329  Mean    : 0.014708  Mean    : 0.005694
## 3rd Qu.: 0.006483  3rd Qu.: 0.021283  3rd Qu.: 0.012929
## Max.   : 0.135050  Max.    : 0.074209  Max.    : 0.042360
##      SWE          CHE          USA
## Min.   :-0.085963  Min.   :-0.0730785  Min.   :-0.088660
## 1st Qu.: 0.001383  1st Qu.: 0.0003708  1st Qu.: 0.003630
## Median : 0.006541  Median : 0.0051884  Median : 0.006949
## Mean   : 0.005231  Mean    : 0.0042111  Mean    : 0.006299
## 3rd Qu.: 0.011298  3rd Qu.: 0.0083389  3rd Qu.: 0.009975
## Max.   : 0.071156  Max.    : 0.0706754  Max.    : 0.075613
##      OECD
## Min.   :-0.109020
## 1st Qu.: 0.003672
## Median : 0.006078
## Mean   : 0.005661
## 3rd Qu.: 0.008665
## Max.   : 0.090298
```

Cálculo y representación del ACF

El cálculo y la representación de los ACF de los respectivos países ha sido realizado mediante el siguiente código:

```
### CÁLCULO Y REPRESENTACIÓN DEL ACF
# Calcula el ACF para todas las variables del dataframe data_w
lapply(data_w[, 2:11], acf)
```

El resultado de este código será una lista que contiene los resultados de la función `acf()` aplicada a cada columna del *dataframe* `data_w`. Cada elemento de la lista corresponderá a la función de autocorrelación de una variable específica en el *dataframe*.

```
## $AUS
## Autocorrelations of series 'X[[i]]', by lag
##
##      0      1      2      3      4      5      6      7      8      9
10
##  1.000  0.065 -0.077 -0.017 -0.045  0.017 -0.019 -0.014 -0.060 -0.068
0.014
##      11      12      13      14      15      16      17      18      19      20
21
## -0.045 -0.024  0.048 -0.015  0.073  0.045 -0.039 -0.023 -0.014  0.038
0.024
##      22
##  0.032
##
## $CAN
## Autocorrelations of series 'X[[i]]', by lag
##
##      0      1      2      3      4      5      6      7      8      9
10
##  1.000 -0.036 -0.046  0.033 -0.002  0.037 -0.019 -0.040 -0.035 -0.023 -
0.002
##      11      12      13      14      15      16      17      18      19      20
21
## -0.018 -0.034 -0.034 -0.019 -0.012  0.065 -0.017  0.018 -0.012  0.033 -
0.004
##      22
## -0.066
##
## $FRA
## Autocorrelations of series 'X[[i]]', by lag
##
##      0      1      2      3      4      5      6      7      8      9
10
##  1.000 -0.272 -0.096  0.019 -0.014  0.006  0.003  0.006  0.000  0.010 -
0.018
##      11      12      13      14      15      16      17      18      19      20
21
```

```

## -0.012 -0.004 0.000 0.009 0.014 0.015 -0.029 0.011 -0.006 -0.004 -
0.010
##      22
## 0.016
##
## $ITA
## Autocorrelations of series 'X[[i]]', by lag
##
##      0      1      2      3      4      5      6      7      8      9
10
## 1.000 -0.136 -0.065 0.024 0.006 0.013 0.006 0.017 0.008 -0.006 -
0.009
##      11      12      13      14      15      16      17      18      19      20
21
## 0.030 0.024 0.022 0.025 0.011 0.025 0.008 0.007 0.012 -0.006
0.019
##      22
## 0.035
##
## $KOR
## Autocorrelations of series 'X[[i]]', by lag
##
##      0      1      2      3      4      5      6      7      8      9      10      11
12
## 1.000 0.273 0.178 0.220 0.079 0.069 0.100 0.128 0.079 0.162 0.159 0.136
0.212
##      13      14      15      16      17      18      19      20      21      22
## 0.121 0.158 0.188 0.129 0.119 0.170 0.135 0.172 0.173 0.180
##
## $NOR
## Autocorrelations of series 'X[[i]]', by lag
##
##      0      1      2      3      4      5      6      7      8      9
10
## 1.000 -0.255 0.043 0.018 0.018 0.160 -0.054 0.105 -0.078 0.087
0.017
##      11      12      13      14      15      16      17      18      19      20
21
## 0.010 -0.084 -0.053 0.065 -0.085 0.122 -0.114 0.067 -0.036 -0.112
0.129
##      22
## -0.160
##
## $SWE
## Autocorrelations of series 'X[[i]]', by lag
##
##      0      1      2      3      4      5      6      7      8      9
10
## 1.000 -0.148 0.108 0.092 -0.024 0.002 -0.049 0.003 -0.105 -0.032
0.015
##      11      12      13      14      15      16      17      18      19      20
21
## -0.056 -0.066 0.055 -0.016 0.088 0.027 -0.021 -0.005 -0.019 -0.023 -

```

```

0.023
##      22
## -0.048
##
## $CHE
## Autocorrelations of series 'X[[i]]', by lag
##
##      0      1      2      3      4      5      6      7      8      9
10
##  1.000 -0.057 -0.035 -0.002  0.007  0.020  0.014  0.001 -0.113 -0.032 -
0.041
##      11      12      13      14      15      16      17      18      19      20
21
## -0.011 -0.014 -0.048 -0.025 -0.022  0.006  0.015  0.001 -0.018 -0.043 -
0.007
##      22
## -0.085
##
## $USA
## Autocorrelations of series 'X[[i]]', by lag
##
##      0      1      2      3      4      5      6      7      8      9
10
##  1.000 -0.086  0.048  0.056  0.040  0.008  0.012 -0.030 -0.031  0.010 -
0.011
##      11      12      13      14      15      16      17      18      19      20
21
## -0.016 -0.013 -0.012  0.002 -0.002  0.034 -0.006  0.039  0.028 -0.004 -
0.027
##      22
## -0.015
##
## $OECD
## Autocorrelations of series 'X[[i]]', by lag
##
##      0      1      2      3      4      5      6      7      8      9
10
##  1.000 -0.179 -0.017  0.005  0.008  0.001  0.010 -0.003 -0.017  0.006 -
0.010
##      11      12      13      14      15      16      17      18      19      20
21
## -0.007 -0.005  0.000  0.005  0.031  0.018  0.003  0.012 -0.007 -0.011 -
0.026
##      22
## -0.004

```

Test de raíz unitaria

```
library(urca)
ur.df(data_w$AUS, type = "drift")

##
## #####
## # Augmented Dickey-Fuller Test Unit Root / Cointegration Test #
## #####
##
## The value of the test statistic is: -8.8112 38.8364
```

En el resultado del test de Dickey-Fuller aumentado, el valor del estadístico de prueba es -8.8112 y el valor crítico al 5% es -2.861. Como el estadístico de prueba es menor que el valor crítico, podemos rechazar la hipótesis nula de que la serie tiene una raíz unitaria. Esto sugiere que la serie es estacionaria y tiene un orden de integración de 0, es decir, es una serie I(0).

```
ur.df(data_w$CAN, type = "drift")

##
## #####
## # Augmented Dickey-Fuller Test Unit Root / Cointegration Test #
## #####
##
## The value of the test statistic is: -8.6581 37.5644

ur.df(data_w$FRA, type = "drift")

##
## #####
## # Augmented Dickey-Fuller Test Unit Root / Cointegration Test #
## #####
##
## The value of the test statistic is: -10.9066 59.5739

ur.df(data_w$ITA, type = "drift")

##
## #####
## # Augmented Dickey-Fuller Test Unit Root / Cointegration Test #
## #####
##
## The value of the test statistic is: -9.2786 43.1117

ur.df(data_w$KOR, type = "drift")

##
## #####
## # Augmented Dickey-Fuller Test Unit Root / Cointegration Test #
## #####
##
## The value of the test statistic is: -6.7974 23.115

ur.df(data_w$NOR, type = "drift")
```

```

##
## #####
## # Augmented Dickey-Fuller Test Unit Root / Cointegration Test #
## #####
##
## The value of the test statistic is: -9.9381 49.4311
ur.df(data_w$SWE, type = "drift")
##
## #####
## # Augmented Dickey-Fuller Test Unit Root / Cointegration Test #
## #####
##
## The value of the test statistic is: -8.0303 32.2755
ur.df(data_w$CHE, type = "drift")
##
## #####
## # Augmented Dickey-Fuller Test Unit Root / Cointegration Test #
## #####
##
## The value of the test statistic is: -8.7889 38.653
ur.df(data_w$USA, type = "drift")
##
## #####
## # Augmented Dickey-Fuller Test Unit Root / Cointegration Test #
## #####
##
## The value of the test statistic is: -7.9575 31.7201
ur.df(data_w$OECD, type = "drift")
##
## #####
## # Augmented Dickey-Fuller Test Unit Root / Cointegration Test #
## #####
##
## The value of the test statistic is: -9.0686 41.1971

```

A continuación, añadimos una variable factor global que nos servirá como variable regresora para elaborar futuros pronósticos. Además, le hacemos algunos ajustes para que se adapte al periodo de tiempo de los datos y la incluimos al *dataframe* dd.

```
library(readxl)
GECON<-read_excel("GECON_indicator.xlsx")
head(GECON)

## # A tibble: 6 x 5
##   TIME   `Global Economic Conditions Indicator` Standardized G~1 Smoot~
##   <chr>          <dbl>          <dbl> <chr>
##   <chr>
## 1 1973M2          0.599          1.29  NaN
##   NaN
## 2 1973M3          0.406          0.871 NaN
##   NaN
## 3 1973M4        -0.0880        -0.191 0.3056
##   ~ 0.7535~
## 4 1973M5          0.123          0.263 0.1469
##   ~ 0.3618~
## 5 1973M6          0.0528          0.112 2.9262
##   ~ 7.1456~
## 6 1973M7        -0.116         -0.250 2.0084
##   ~ 4.8801~
## # ... with abbreviated variable names 1: `Standardized GECON`,
## # 2: `Smoothed GECON3`, 3: `Standardized GECON3`
```

Ahora eliminamos las fechas que no utilizamos en *data_w*

```
# Convertir La variable TIME de data_w a formato de fecha
data_w$TIME <- as.Date(data_w$TIME, format = "%Y-%m-%d")

# Extraer solo las filas de GECON con fechas coincidentes con data_w
GECON_subset <- subset(GECON, TIME %in% data_w$TIME)
```

Aplicamos el mismo *modus operandi* para 2 variables más.

En el último apartado del código se muestra la formulación de los modelos de predicción de cada tipo.

En primer lugar, creamos las tablas vacías donde almacenaremos la información.

```
#FORECASTING
```

Matrices de info

```
table_ets_for_h1=c() #distinto para cada h
table_ets_for_h2=c() #distinto para cada h
table_ets_for_h3=c() #distinto para cada h
table_ets_for_h4=c() #distinto para cada h
table_ets_model=c() #mismo para todos los h
table_ets_aicc=c() #mismo para todos los h
```

Seguidamente, establecemos el marco de predicción y creamos la lista donde almacenaremos estas predicciones.

```
# Convertir la variable TIME a un objeto de clase "ts"
ts_time <- ts(dd$TIME, start = c(1980, 2), frequency = 4)
```

```
# Crear una lista para almacenar los pronósticos
forecast_list <- list()
```

Y finalmente, en un bucle for() realizamos la predicción para cada variable, calculamos su predicción en h1,h2,h3,h4, guardamos el modelo, su respectivo AICc y creamos una tabla resumen que contiene las predicciones en los 4 horizontes.

```
for (i in 1:length(window(ts_time, start = c(2001, 4), end = c(2019, 4))))
{
  # Iterar sobre todas las columnas del conjunto de datos, excepto la columna TIME
  for (col in names(dd)[c(-1,-12:-17)]) {
    # Crear una serie de tiempo para la variable en los últimos 20 años
    ts_col <- ts(dd[[col]], start = c(1980, 4), end = c(2019, 4), frequency = 4)

    ts_col_last_20_years <- ts_col[(i):(i+79)]

    # Ajustar un modelo ETS a la serie de tiempo en los últimos 20 años
    model_col <- ets(ts_col_last_20_years)

    # Hacer un pronóstico para los próximos 4 períodos
    forecast_col <- forecast(model_col, h = 4)

    # Agregar la columna al pronóstico y guardarlo en la lista
    forecast_col$col <- col
    forecast_list[[col]] <- forecast_col
  }
}
```



```

table_ets_for_h1=rbind(table_ets_for_h1,Reduce(cbind,lapply(forecast_list,
function(x) x$mean[1])))
table_ets_for_h2=rbind(table_ets_for_h2,Reduce(cbind,lapply(forecast_list,
function(x) x$mean[2])))
table_ets_for_h3=rbind(table_ets_for_h3,Reduce(cbind,lapply(forecast_list,
function(x) x$mean[3])))
table_ets_for_h4=rbind(table_ets_for_h4,Reduce(cbind,lapply(forecast_list,
function(x) x$mean[4])))

table_ets_model=rbind(table_ets_model,Reduce(cbind,lapply(forecast_list,fun
ction(x) x$model$method)))
table_ets_aicc <- rbind(table_ets_aicc,Reduce(cbind,lapply(forecast_list,f
unction(x) x$model$aicc)))

}

# Lista de pronósticos
# OJO: no contar pronóstico desde 2020q1-2020q4 (osea los que se estiman en
2019)
table_ets_for=list("h1"=table_ets_for_h1[1:(length(table_ets_for_h1[,1])-1
),],
                  "h2"=table_ets_for_h2[1:(length(table_ets_for_h2[,1])-2
),],
                  "h3"=table_ets_for_h3[1:(length(table_ets_for_h3[,1])-3
),],
                  "h4"=table_ets_for_h4[1:(length(table_ets_for_h4[,1])-4
),])

```

Finalmente, calculamos el RMSE

###Cálculo del RMSE

```

# Crear una lista para almacenar Los RMSE
mse_ets<- data.frame(matrix(nrow=10,ncol=4))
row.names(mse_ets)<-names(forecast_list)
colnames(mse_ets)<-c("h1","h2","h3","h4")

# Para cada h
for (h in 1:4){
  # Iterar sobre los elementos de la lista forecast_list
  i<-0
  for (col in names(forecast_list)){
    i<-i+1
    ts_col <- ts(dd[[col]], start = c(1980, 4), end = c(2019, 4), frequenc
y = 4)
    actual_values <- ts_col[(length(ts_col)-length(table_ets_for[[h]][,i]
+1):length(ts_col)]
    mse_ets[i,h]<- round(sqrt(mean((table_ets_for[[h]][,i] - actual_values
)^2)),5)
  }
}
# MIRA EL RMSE! A medida que nos alejamos en h, el error de predicción aum

```

```

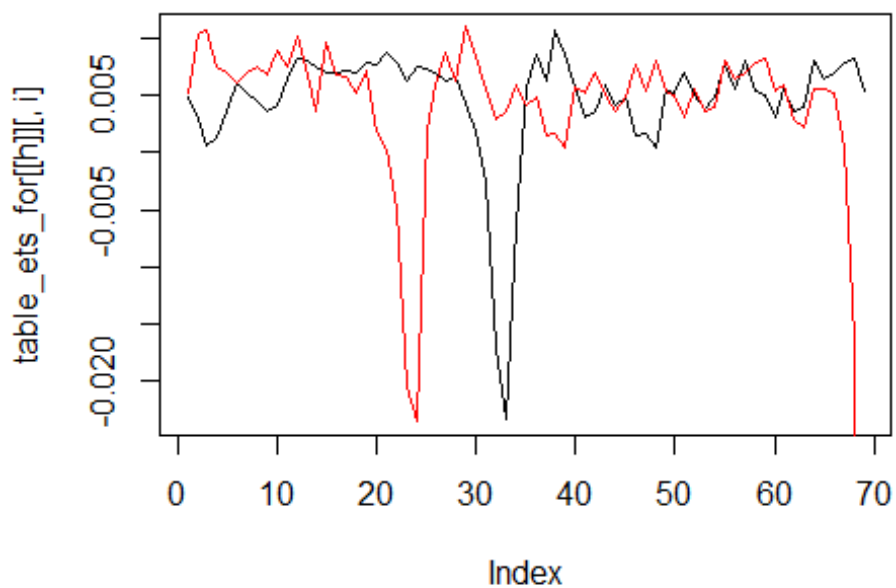
enta (correcto!)
mse_ets

##          h1      h2      h3      h4
## AUS  0.01005 0.01009 0.01023 0.01028
## CAN  0.01670 0.01714 0.01727 0.01717
## FRA  0.02026 0.02033 0.02052 0.02065
## ITA  0.01950 0.01948 0.01961 0.01973
## KOR  0.01048 0.01012 0.01044 0.00973
## NOR  0.01283 0.01253 0.01246 0.01273
## SWE  0.01588 0.01557 0.01586 0.01554
## CHE  0.01231 0.01243 0.01252 0.01246
## USA  0.01312 0.01344 0.01354 0.01358
## OECD 0.01544 0.01562 0.01600 0.01607

#check h=4 i=10
plot(table_ets_for[[h]][,i],t="l")
lines(actual_values,t="l",col="red") # Perfecto! ETS debería pronosticar e
n h=4 algo parecido que 4 periodos antes

```

Figura A1.1: MSE modelo ETS



Fuente: Elaboración propia

Repetimos el mismo proceso para el resto de modelos.

Apéndice 2: Código de R

En el Apéndice 2, titulado "Código de R", se proporciona información detallada sobre el código utilizado en el estudio, específicamente utilizando el lenguaje de programación R. Este apéndice es una sección complementaria del trabajo principal y tiene como objetivo mostrar y explicar el código utilizado para llevar a cabo análisis, cálculos y visualizaciones relacionadas con el pronóstico del crecimiento del PIB y la evaluación del impacto del Covid-19.

En este apartado, se incluirá el código fuente de R utilizado en el estudio, organizado de manera estructurada y comentado para facilitar su comprensión. El código puede abarcar desde la carga y manipulación de datos hasta la implementación de diferentes modelos de pronóstico, cálculos de métricas de evaluación y generación de gráficos relevantes.

La inclusión del código de R en el apéndice es útil para aquellos lectores interesados en replicar los resultados del estudio, comprender en detalle los procedimientos técnicos utilizados o utilizar el código como base para futuras investigaciones o aplicaciones.

Además del código en sí, este apéndice también contiene explicaciones adicionales y comentarios sobre el código, destacando aspectos clave, explicando la lógica detrás de ciertas líneas de código y brindando información adicional para comprender y utilizar correctamente el código presentado.

```
# IMPORTAR DATOS

library(tidyverse)

set.seed(12345)
rm(list = ls()) # Limpiar environment
setwd(getwd())
data<-read.csv2("datosgdp.csv", header=T, dec=".", sep=",")
head(data)

# ¿CUÁNTOS PAÍSES HAY EN LA OECD?

table(data$i..LOCATION)

países<-names(table(data$i..LOCATION))
length(países)

#####

### TRATAMIENTO DE DATOS

library(tidyverse)
library(forecast)

data<-read.csv2("datosgdp.csv", sep=",", dec=".")
names(data)<-c("LOCATION", "INDICATOR", "SUBJECT", "MEASURE", "FREQUENCY", "TIME", "Value", "Flag.Codes")
data_w<-data[,c(1,6,7)] %>% pivot_wider(names_from = "LOCATION", values_from
```

```

= "Value" )

data_w<-data_w[81:240, ]%>%
  select_if(~ !any(is.na(.))) %>%
  mutate_at(vars(-"TIME"), # rest I(1)
            .funs = function(x) log(x/lag(x,1))) %>%
  mutate (TIME=seq(as.Date("1980-1-1"), by = "quarter",
                  length.out = 240-81+1, n=seq(1,by=1,240-81+1)))

data_w %>% pivot_longer(cols = -"TIME",names_to = "Country",values_to = "V
alue") %>%
  group_by(Country) %>%
  summarise(n=n(),mean=mean(Value,na.rm = T))

data_w<-data_w[-1,]

Y<-data_w %>%
  mutate_at(vars(-"TIME"), # rest I(1)
            .funs = function(x) log(x/lag(x,12)))

# urca::ur.ers
colnames(data_w)

#####

### ESTADÍSTICOS DESCRIPTIVOS

#antes que nada, eliminamos países que son colonias de otros (Gibraltar)
data_w<-data_w[, -10]
Y<-Y[, -10]

#analizamos los estadísticos descriptivos

data_w %>%
  select(-TIME) %>% # Excluir la columna TIME
  summary()

#####

#### ANÁLISIS DE LA MEDIA Y LA VARIANZA

library(zoo)

library(ggplot2)

# Crear un dataframe vacío para almacenar los resultados
resultados_media <- data.frame(matrix(ncol=3, nrow=0))
colnames(resultados_media) <- c("Variable", "Tiempo", "Media")
resultados_varianza <- data.frame(matrix(ncol=3, nrow=0))
colnames(resultados_varianza) <- c("Variable", "Tiempo", "Varianza")

```

```

# Cambiar el nombre de las columnas del dataframe data_w
colnames(data_w)[-1] <- c("AUS", "CAN", "FRA", "ITA", "KOR", "NOR", "SWE",
"CHE", "USA", "OECD")

# Loop para calcular la media y la varianza móvil de cada variable y graficar
for (i in 2:11) { # Comenzar en la segunda columna porque la primera es la fecha
  variable <- names(data_w)[i] # Obtener el nombre de la variable
  media_movil <- rollapply(data_w[, i], width = 12, FUN = mean, align = "right", fill = NA) # Calcular la media móvil
  varianza_movil <- rollapply(data_w[, i], width = 12, FUN = var, align = "right", fill = NA) # Calcular la varianza móvil

  # Almacenar los resultados en un dataframe
  resultados_media_nuevos <- setNames(data.frame(Variable = variable, Tiempo = data_w$TIME, Media = media_movil), colnames(resultados_media))
  resultados_media <- rbind(resultados_media, resultados_media_nuevos)

  resultados_varianza_nuevos <- setNames(data.frame(Variable = variable, Tiempo = data_w$TIME, Varianza = varianza_movil), colnames(resultados_varianza))
  resultados_varianza <- rbind(resultados_varianza, resultados_varianza_nuevos)

  # Crear una gráfica para cada variable
  p <- ggplot() +
    geom_line(data = resultados_media_nuevos, aes(x = Tiempo, y = Media), color = "blue") +
    geom_line(data = resultados_varianza_nuevos, aes(x = Tiempo, y = Varianza), color = "red") +
    ggtitle(paste("Media y varianza móvil de", variable)) +
    xlab("Tiempo") +
    ylab("Valor") +
    scale_x_date(date_labels = "%Y", date_breaks = "1 year")
  print(p)
}

#####

### CÁLCULO Y REPRESENTACIÓN DEL ACF

# Calcula el ACF para todas las variables del dataframe data_w
lapply(data_w[, 2:11], acf)

#####

### TESTS DE RAÍZ UNITARIA

```

```

library(urca)
ur.df(data_w$AUS, type = "drift")
ur.df(data_w$CAN, type = "drift")
ur.df(data_w$FRA, type = "drift")
ur.df(data_w$ITA, type = "drift")
ur.df(data_w$KOR, type = "drift")
ur.df(data_w$NOR, type = "drift")
ur.df(data_w$SWE, type = "drift")
ur.df(data_w$CHE, type = "drift")
ur.df(data_w$USA, type = "drift")
ur.df(data_w$OECD, type = "drift")

```

```

#####

### AÑADIMOS VARIABLES

library(readxl)
GECON<-read_excel("GECON_indicator.xlsx")
head(GECON)

##modificamos La variable TIME

# Obtener La variable TIME del dataframe GECON
time <- GECON$TIME

# Reemplazar "M" con "-" utilizando La función gsub()
time <- gsub("M", "-", time)

# Utilizar La función as.Date() para convertir La cadena de caracteres a un
objeto de fecha
time <- as.Date(paste0(time, "-01"), format = "%Y-%m-%d")

GECON$TIME<-time
head(GECON)

# ahora eliminamos Las fechas que no utilizamos en data_w

# Convertir La variable TIME de data_w a formato de fecha
data_w$TIME <- as.Date(data_w$TIME, format = "%Y-%m-%d")

```

```

# Extraer solo las filas de GECON con fechas coincidentes con data_w
GECON_subset <- subset(GECON, TIME %in% data_w$TIME)

commodity_pf<-read_excel("BG_commodity_price_factor.xlsx")
head(commodity_pf)

# Transformar la columna TIME de IP al formato adecuado
commodity_pf$TIME <- as.Date(commodity_pf$TIME, "%Y-%m-%d")

## Warning in as.POSIXlt.POSIXct(x, tz = tz): unknown timezone '%Y-%m-%d'

# Convertir la columna TIME de data_w a fecha
data_w$TIME <- as.Date(data_w$TIME, "%Y-%m-%d")

# Seleccionar las filas de IP que corresponden a las fechas en data_w
commodity_pf_subset <- subset(commodity_pf, TIME %in% data_w$TIME)

head(commodity_pf_subset)

IP<-read_excel("OECD_plus6_industrial_production.xlsx")
head(IP)

# Transformar la columna TIME de IP al formato adecuado
IP$TIME <- as.Date(IP$TIME, "%Y-%m-%d")

# Convertir la columna TIME de data_w a fecha
data_w$TIME <- as.Date(data_w$TIME, "%Y-%m-%d")

# Seleccionar las filas de IP que corresponden a las fechas en data_w
IP_subset <- subset(IP, TIME %in% data_w$TIME)

head(IP_subset)

#juntamos Los dataframes

dd <- merge(data_w, GECON_subset, by = "TIME")
dd <- merge(dd, commodity_pf_subset, by = "TIME")
dd <- merge(dd, IP_subset, by = "TIME")
head(dd)

names(dd)

attach(dd)
str(dd)

#vemos que hay variables en formato character, para solucionarlo hacemos:
dd$`Smoothed GECON3` <-as.numeric(`Smoothed GECON3`)
dd$`Standardized GECON3`<-as.numeric(`Standardized GECON3`)
str(dd)

```

```
#####

### FORECASTING

# Matrices de info

table_ets_for_h1=c() #distinto para cada h
table_ets_for_h2=c() #distinto para cada h
table_ets_for_h3=c() #distinto para cada h
table_ets_for_h4=c() #distinto para cada h
table_ets_model=c() #mismo para todos Los h
table_ets_aicc=c() #mismo para todos Los h

# Convertir La variable TIME a un objeto de clase "ts"
ts_time <- ts(dd$TIME, start = c(1980, 2), frequency = 4)

# Crear una Lista para almacenar Los pronósticos
forecast_list <- list()

for (i in 1:length(window(ts_time, start = c(2001, 4), end = c(2019, 4))))
{
  # Iterar sobre todas Las columnas del conjunto de datos, excepto La columna TIME

  for (col in names(dd)[c(-1, -12:-17)]) {
    # Crear una serie de tiempo para La variable en Los últimos 20 años
    ts_col <- ts(dd[[col]], start = c(1980, 4), end = c(2019, 4), frequency = 4)

    ts_col_last_20_years <- ts_col[(i):(i+79)]

    # Ajustar un modelo ETS a La serie de tiempo en Los últimos 20 años
    model_col <- ets(ts_col_last_20_years)

    # Hacer un pronóstico para Los próximos 4 períodos
    forecast_col <- forecast(model_col, h = 4)

    # Agregar La columna al pronóstico y guardarlo en La Lista
    forecast_col$col <- col
    forecast_list[[col]] <- forecast_col
  }

  table_ets_for_h1=rbind(table_ets_for_h1,Reduce(cbind,lapply(forecast_list,function(x) x$mean[1])))
  table_ets_for_h2=rbind(table_ets_for_h2,Reduce(cbind,lapply(forecast_list,function(x) x$mean[2])))
  table_ets_for_h3=rbind(table_ets_for_h3,Reduce(cbind,lapply(forecast_list,function(x) x$mean[3])))
  table_ets_for_h4=rbind(table_ets_for_h4,Reduce(cbind,lapply(forecast_list,function(x) x$mean[4])))

  table_ets_model=rbind(table_ets_model,Reduce(cbind,lapply(forecast_list,fu
```



```

nction(x) x$model$method))
table_ets_aicc <- rbind(table_ets_aicc, Reduce(cbind, lapply(forecast_list, f
unction(x) x$model$aicc)))
}

# Lista de pronósticos
# OJO: no contar pronóstico desde 2020q1-202q4 (osea los que se estiman en
2019)
table_ets_for=list("h1"=table_ets_for_h1[1:(length(table_ets_for_h1[,1])-1
),],
                  "h2"=table_ets_for_h2[1:(length(table_ets_for_h2[,1])-2
),],
                  "h3"=table_ets_for_h3[1:(length(table_ets_for_h3[,1])-3
),],
                  "h4"=table_ets_for_h4[1:(length(table_ets_for_h4[,1])-4
),])

# CÁLCULO DEL RMSE

# Crear una lista para almacenar los RMSE
mse_ets<- data.frame(matrix(nrow=10, ncol=4))
row.names(mse_ets)<-names(forecast_list)
colnames(mse_ets)<-c("h1", "h2", "h3", "h4")

# Para cada h
for (h in 1:4){
  # Iterar sobre los elementos de la lista forecast_list
  i<-0
  for (col in names(forecast_list)){
    i<-i+1
    ts_col <- ts(dd[[col]], start = c(1980, 4), end = c(2019, 4), frequenc
y = 4)
    actual_values <- ts_col[(length(ts_col)-length(table_ets_for[[h]][,i]
+1):length(ts_col)]
    mse_ets[i,h]<- round(sqrt(mean((table_ets_for[[h]][,i] - actual_values
)^2)),5)
  }
}
# A medida que nos alejamos en h, el error de predicción aumenta (correcto
!)
mse_ets

#check h=4 i=10
plot(table_ets_for[[h]][,i],t="l")
lines(actual_values,t="l",col="red") # Perfecto! ETS debería pronosticar e
n h=4 algo parecido que 4 periodos antes

#####

```

```

## ARIMA(p,d,q)

# Hacemos una prueba inicial para un ARIMA(1,0,1)

# Ajustar un modelo ARIMA (1,0,1) para cada variable en dd
arima101_list <- lapply(dd[,c(-1,-12:-17)], function(x) {
  model <- arima(x, order = c(1, 0, 1)) # Ajustar un modelo ARIMA(1, 0, 1)
})

# Matrices de info

table_arima_for_h1=c() #distinto para cada h
table_arima_for_h2=c() #distinto para cada h
table_arima_for_h3=c() #distinto para cada h
table_arima_for_h4=c() #distinto para cada h
table_arima_model=c() #mismo para todos Los h
table_arima_aicc=c() #mismo para todos Los h

# Convertir la variable TIME a un objeto de clase "ts"
ts_time <- ts(dd$TIME, start = c(1980, 2), frequency = 4)

# Crear una lista para almacenar los pronósticos
forecast_arima_list <- list()

for (i in 1:length(window(ts_time, start = c(2001, 4), end = c(2019, 4))))
{
  # Iterar sobre todas las columnas del conjunto de datos, excepto la columna TIME

  for (col in names(dd)[c(-1,-12:-17)]) {
    # Crear una serie de tiempo para la variable en los últimos 20 años
    ts_col <- ts(dd[[col]], start = c(1980, 4), end = c(2019, 4), frequency = 4)

    ts_col_last_20_years <- ts_col[(i):(i+79)]

    # Ajustar un modelo ETS a la serie de tiempo en los últimos 20 años
    model_col <- auto.arima(ts_col_last_20_years)

    # Hacer un pronóstico para los próximos 4 períodos
    forecast_col <- forecast(model_col, h = 4)

    # Agregar la columna al pronóstico y guardarlo en la lista
    forecast_col$col <- col
    forecast_arima_list[[col]] <- forecast_col
  }
}

table_arima_for_h1=rbind(table_arima_for_h1,Reduce(cbind,lapply(forecast_a

```

```

rima_list,function(x) x$mean[1]))
table_arima_for_h2=rbind(table_arima_for_h2,Reduce(cbind,lapply(forecast_a
rima_list,function(x) x$mean[2]))
table_arima_for_h3=rbind(table_arima_for_h3,Reduce(cbind,lapply(forecast_a
rima_list,function(x) x$mean[3]))
table_arima_for_h4=rbind(table_arima_for_h4,Reduce(cbind,lapply(forecast_a
rima_list,function(x) x$mean[4]))

table_arima_model=rbind(table_arima_model,Reduce(cbind,lapply(forecast_ari
ma_list,function(x) x$method))
table_arima_aicc <- rbind(table_arima_aicc,Reduce(cbind,lapply(forecast_ar
ima_list,function(x) x$model$aicc))

}

# Lista de pronósticos
# OJO: no contar pronóstico desde 2020q1-202q4 (osea Los que se estiman en
2019)
table_arima_for=list("h1"=table_arima_for_h1[1:(length(table_arima_for_h1[
,1])-1),],
                    "h2"=table_arima_for_h2[1:(length(table_arima_for_h2[,1
])-2),],
                    "h3"=table_arima_for_h3[1:(length(table_arima_for_h3[,1
])-3),],
                    "h4"=table_arima_for_h4[1:(length(table_arima_for_h4[,1
])-4),])

# Crear una lista para almacenar Los RMSE
mse_arima<- data.frame(matrix(nrow=10,ncol=4))
row.names(mse_arima)<-names(forecast_arima_list)
colnames(mse_arima)<-c("h1","h2","h3","h4")

# Para cada h
for (h in 1:4){
  # Iterar sobre Los elementos de La lista forecast_list
  i<-0
  for (col in names(forecast_arima_list)){
    i<-i+1
    ts_col <- ts(dd[[col]], start = c(1980, 4), end = c(2019, 4), frequenc
y = 4)
    actual_values <- ts_col[(length(ts_col)-length(table_arima_for[[h]][,i
])+1):length(ts_col)]
    mse_arima[i,h]<- round(sqrt(mean((table_arima_for[[h]][,i] - actual_va
lues)^2)),5)
  }
}
# A medida que nos alejamos en h, el error de predicción aumenta (correcto
!)
mse_arima

#check h=4 i=10
plot(table_arima_for[[h]][,i],t="l")

```

```

lines(actual_values,t="l",col="red") # Perfecto! ETS debería pronosticar e
n h=4 algo parecido que 4 periodos antes

#####

## Variables globales

# ¿Cuáles son estas variables globales?
names(dd[,12:17])

#####

## Global Economic Conditions Indicator (x1)

# Matrices de info

table_arima_x1_for_h1=c() #distinto para cada h
table_arima_x1_for_h2=c() #distinto para cada h
table_arima_x1_for_h3=c() #distinto para cada h
table_arima_x1_for_h4=c() #distinto para cada h
table_arima_x1_model=c() #mismo para todos los h
table_arima_x1_aicc=c() #mismo para todos los h

#(para testear porque había un problema con las length)

#modelo_arima_x1 <- auto.arima(ts_col, xreg = dd$'Global Economic Conditio
ns Indicator'[c(-1, -159)])
#forecast_arima_x1 <- forecast(modelo_arima_x1, xreg = tail(dd$'Global Eco
nomic Conditions Indicator', 4))

# Crear una lista para almacenar los pronósticos
forecast_arima_x1_list <- list()

for (i in 1:length(window(ts_time, start = c(2001, 4), end = c(2019, 4))))
{
  # Iterar sobre todas las columnas del conjunto de datos, excepto la colu
mna TIME

  for (col in names(dd)[c(-1,-12:-17)]) {
    # Crear una serie de tiempo para la variable en los últimos 20 años
    ts_col <- ts(dd[[col]], start = c(1980, 4), end = c(2019, 4), frequency
= 4)

    ts_col_last_20_years <- ts_col[(i):(i+79)]

    # Ajustar un modelo ETS a la serie de tiempo en los últimos 20 años
    model_col <- auto.arima(ts_col, xreg = dd$'Global Economic Conditions In
dicator'[c(-1, -159)])

    # Hacer un pronóstico para los próximos 4 períodos
    forecast_col <- forecast(model_col, xreg = tail(dd$'Global Economic Cond
itions Indicator', 4))

```

```

# Agregar la columna al pronóstico y guardarlo en la lista
forecast_col$col <- col
forecast_arima_x1_list[[col]] <- forecast_col
}

table_arima_x1_for_h1=rbind(table_arima_x1_for_h1,Reduce(cbind,lapply(fore
cast_arima_x1_list,function(x) x$mean[1])))
table_arima_x1_for_h2=rbind(table_arima_x1_for_h2,Reduce(cbind,lapply(fore
cast_arima_x1_list,function(x) x$mean[2])))
table_arima_x1_for_h3=rbind(table_arima_x1_for_h3,Reduce(cbind,lapply(fore
cast_arima_x1_list,function(x) x$mean[3])))
table_arima_x1_for_h4=rbind(table_arima_x1_for_h4,Reduce(cbind,lapply(fore
cast_arima_x1_list,function(x) x$mean[4])))

table_arima_x1_model=rbind(table_arima_x1_model,Reduce(cbind,lapply(foreca
st_arima_x1_list,function(x) x$method)))
table_arima_x1_aicc <- rbind(table_arima_x1_aicc,Reduce(cbind,lapply(forec
ast_arima_x1_list,function(x) x$model$aicc)))

}

# Lista de pronósticos
# OJO: no contar pronóstico desde 2020q1-2020q4 (osea los que se estiman en
2019)
table_arima_x1_for=list("h1"=table_arima_x1_for_h1[1:(length(table_arima_x
1_for_h1[,1])-1),],
                      "h2"=table_arima_x1_for_h2[1:(length(table_arima_x1_for
_h2[,1])-2),],
                      "h3"=table_arima_x1_for_h3[1:(length(table_arima_x1_for
_h3[,1])-3),],
                      "h4"=table_arima_x1_for_h4[1:(length(table_arima_x1_for
_h4[,1])-4),])

# Crear una lista para almacenar los RMSE
mse_arima_x1<- data.frame(matrix(nrow=10,ncol=4))
row.names(mse_arima_x1)<-names(forecast_arima_x1_list)
colnames(mse_arima_x1)<-c("h1","h2","h3","h4")

# Para cada h
for (h in 1:4){
  # Iterar sobre los elementos de la lista forecast_list
  i<-0
  for (col in names(forecast_arima_x1_list)){
    i<-i+1
    ts_col <- ts(dd[[col]], start = c(1980, 4), end = c(2019, 4), frequenc
y = 4)
    actual_values <- ts_col[(length(ts_col)-length(table_arima_x1_for[[h]]
[,i])+1):length(ts_col)]
    mse_arima_x1[i,h]<- round(sqrt(mean((table_arima_x1_for[[h]][,i] - act
ual_values)^2)),5)
  }
}

```

```

}
}
# MIRA EL RMSE! A medida que nos alejamos en h, el error de predicción aumenta (correcto!)
mse_arima_x1

#check h=4 i=10
#plot(table_arima_x1_for[[h]][,i],t="l")
#lines(actual_values,t="l",col="red") # Perfecto! ETS debería pronosticar en h=4 algo parecido que 4 periodos antes

## Real Commodity Price Factor (x2)

## Global Economic Conditions Indicator (x3)

#####

## Random Forest

# Matrices de info

table_rf_for_h1 <- c()
table_rf_for_h2 <- c()
table_rf_for_h3 <- c()
table_rf_for_h4 <- c()

set.seed(12345)
library(randomForest)

forecast_rf_list<-c()

for (i in 1:length(window(ts_time, start = c(2001, 4), end = c(2019, 4))))
{
  # Iterar sobre todas las columnas del conjunto de datos, excepto la columna TIME
  for (col in names(dd)[c(-1,-12:-17)]) {
    ts_col <- ts(dd[[col]], start = c(1980, 4), end = c(2019, 4), frequency = 4)

    ar1 <- ts_col[4:(length(ts_col)-1)] # AR1
    ar2 <- ts_col[3:(length(ts_col)-2)] # AR2
    ar3 <- ts_col[2:(length(ts_col)-3)] # AR3
    ar4 <- ts_col[1:(length(ts_col)-4)] # AR4
    Ytrain <- ts_col[5:(length(ts_col))] # Variable dependiente

    tail(cbind(Ytrain, ar1, ar2, ar3, ar4)) # Comprobación

    Xtrain <- cbind(ar1, ar2, ar3, ar4)

    # h1
    Xtest <- c(ts_col[length(ts_col)], ts_col[(length(ts_col)-1)], ts_col[(length(ts_col)-2)], ts_col[(length(ts_col)-3)])
    rf_model <- randomForest(y = Ytrain, x = Xtrain, mtry = ncol(Xtrain)/3

```

```

, ntree = 500)
  rf_forecast_h1 <- predict(rf_model, Xtest)

  # h2
  Xtest <- c(rf_forecast_h1, ts_col[length(ts_col)], ts_col[(length(ts_col)-1)], ts_col[(length(ts_col)-2)])
  rf_forecast_h2 <- predict(rf_model, Xtest)

  # h3
  Xtest <- c(rf_forecast_h2, rf_forecast_h1, ts_col[length(ts_col)], ts_col[(length(ts_col)-1)])
  rf_forecast_h3 <- predict(rf_model, Xtest)

  # h4
  Xtest <- c(rf_forecast_h3, rf_forecast_h2, rf_forecast_h1, ts_col[length(ts_col)])
  rf_forecast_h4 <- predict(rf_model, Xtest)

  # Hacer un pronóstico para Los próximos 4 períodos
  forecast_col <- c(rf_forecast_h1, rf_forecast_h2, rf_forecast_h3, rf_forecast_h4)

  # Agregar la columna al pronóstico y guardarlo en la lista
  forecast_col$col <- col
  forecast_rf_list[[col]] <- forecast_col
}

# Actualizar los resultados en las tablas
table_rf_for_h1 = rbind(table_rf_for_h1, Reduce(cbind, lapply(forecast_rf_list, function(x) x[1])))
table_rf_for_h2 = rbind(table_rf_for_h2, Reduce(cbind, lapply(forecast_rf_list, function(x) x[2])))
table_rf_for_h3 = rbind(table_rf_for_h3, Reduce(cbind, lapply(forecast_rf_list, function(x) x[3])))
table_rf_for_h4 = rbind(table_rf_for_h4, Reduce(cbind, lapply(forecast_rf_list, function(x) x[4])))
}

#plot(c(rf_forecast_h1, rf_forecast_h2, rf_forecast_h3, rf_forecast_h4), t="L")

#CON BASE EN ESTA RUTINA CREAR UN CIRCUITO DE PRONOSTICO IGUAL QUE AL RESTO

# DE MODELOS. IMPORTANTE: PARA ESTE MODELO SOLO GUARDAR PRONÓSTICOS!

table_rf_for_h1 <- as.data.frame(table_rf_for_h1)
table_rf_for_h1 <- matrix(unlist(table_rf_for_h1), nrow = nrow(table_rf_for_h1), byrow = TRUE)

table_rf_for_h2 <- as.data.frame(table_rf_for_h2)

```

```

table_rf_for_h2 <- matrix(unlist(table_rf_for_h2), nrow = nrow(table_rf_for_h2), byrow = TRUE)

table_rf_for_h3<-as.data.frame( table_rf_for_h3)
table_rf_for_h3 <- matrix(unlist(table_rf_for_h3), nrow = nrow(table_rf_for_h3), byrow = TRUE)

table_rf_for_h4<-as.data.frame( table_rf_for_h4)
table_rf_for_h4 <- matrix(unlist(table_rf_for_h4), nrow = nrow(table_rf_for_h4), byrow = TRUE)

# Lista de pronósticos
# OJO: no contar pronóstico desde 2020q1-202q4 (osea los que se estiman en 2019)
table_rf_for=list("h1"=table_rf_for_h1[1:(length(table_rf_for_h1[,1])-1),],
,
                "h2"=table_rf_for_h2[1:(length(table_rf_for_h2[,1])-2),],
],
                "h3"=table_rf_for_h3[1:(length(table_rf_for_h3[,1])-3),],
],
                "h4"=table_rf_for_h4[1:(length(table_rf_for_h4[,1])-4),],
])

# Crear una lista para almacenar Los RMSE
mse_rf<- data.frame(matrix(nrow=10,ncol=4))
row.names(mse_rf)<-names(forecast_List)
colnames(mse_rf)<-c("h1", "h2", "h3", "h4")

# Para cada h
for (h in 1:4){
  # Iterar sobre Los elementos de la lista forecast_List
  i<-0
  for (col in names(forecast_rf_List)){
    i<-i+1
    ts_col <- ts(dd[[col]], start = c(2001, 4), end = c(2019, 4), frequency = 4)
    actual_values <- ts_col[(length(ts_col)-length(table_rf_for[[h]][,i])+1):length(ts_col)]
    mse_rf[i,h]<- round(sqrt(mean((table_rf_for[[h]][,i] - actual_values)^2)),5)
  }
}
# A medida que nos alejamos en h, el error de predicción aumenta (correcto !)
mse_rf

#check h=4 i=10
#plot(table_arima_x3_for[[h]][,i],t="L")
#lines(actual_values,t="L",col="red") # Perfecto! ETS debería pronosticar en h=4 algo parecido que 4 periodos antes

```



```
#####

## RF con variable regresora x

# Matrices de info

table_rf_x_for_h1 <- c()
table_rf_x_for_h2 <- c()
table_rf_x_for_h3 <- c()
table_rf_x_for_h4 <- c()

set.seed(12345)
library(randomForest)

forecast_rf_x_list<-c()

for (i in 1:length(window(ts_time, start = c(2001, 4), end = c(2019, 4))))
{
  # Iterar sobre todas las columnas del conjunto de datos, excepto la columna TIME
  for (col in names(dd)[c(-1,-12:-17)]) {
    ts_col <- ts(dd[[col]], start = c(1980, 4), end = c(2019, 4), frequency = 4)

    ar1 <- ts_col[4:(length(ts_col)-1)] # AR1
    ar2 <- ts_col[3:(length(ts_col)-2)] # AR2
    ar3 <- ts_col[2:(length(ts_col)-3)] # AR3
    ar4 <- ts_col[1:(length(ts_col)-4)] # AR4
    Ytrain <- ts_col[5:(length(ts_col))] # Variable dependiente

    tail(cbind(Ytrain, ar1, ar2, ar3, ar4)) # Comprobación

    Xtrain <- cbind(ar1, ar2, ar3, ar4)

    # h1
    Xtest <- c(ts_col[length(ts_col)], ts_col[(length(ts_col)-1)], ts_col[(length(ts_col)-2)], ts_col[(length(ts_col)-3)])
    rf_model2 <- randomForest(y = Ytrain, x = Xtrain, mtry = ncol(Xtrain)/3, ntree = 500, xreg = dd$'Global Economic Conditions Indicator'[c(-1, -15 9)])
    rf_x_forecast_h1 <- predict(rf_model2, Xtest)

    # h2
    Xtest <- c(rf_forecast_h1, ts_col[length(ts_col)], ts_col[(length(ts_col)-1)], ts_col[(length(ts_col)-2)])
    rf_x_forecast_h2 <- predict(rf_model2, Xtest)

    # h3
    Xtest <- c(rf_forecast_h2, rf_forecast_h1, ts_col[length(ts_col)], ts_col[(length(ts_col)-1)])
    rf_x_forecast_h3 <- predict(rf_model2, Xtest)
  }
}

```

```

# h4
Xtest <- c(rf_forecast_h3, rf_forecast_h2, rf_forecast_h1, ts_col[leng
th(ts_col)])
rf_x_forecast_h4 <- predict(rf_model2, Xtest)

# Hacer un pronóstico para Los próximos 4 períodos
forecast_col <- c(rf_x_forecast_h1,rf_x_forecast_h2,rf_x_forecast_h3,rf_
x_forecast_h4)

# Agregar La columna al pronóstico y guardarlo en La Lista
forecast_col$col <- col
forecast_rf_x_list[[col]] <- forecast_col
}

#rbind(table_arima_x3_for_h1,Reduce(cbind,lapply(forecast_arima_x3_list,
function(x) x$mean[1])))

# Actualizar Los resultados en Las tablas
table_rf_x_for_h1 =rbind(table_rf_x_for_h1,Reduce(cbind,lapply(forecas
t_rf_x_list,function(x) x[1])))
table_rf_x_for_h2 = rbind(table_rf_x_for_h2,Reduce(cbind,lapply(foreca
st_rf_x_list,function(x) x[2])))
table_rf_x_for_h3 =rbind(table_rf_x_for_h3,Reduce(cbind,lapply(forecas
t_rf_x_list,function(x) x[3])))
table_rf_x_for_h4 = rbind(table_rf_x_for_h4,Reduce(cbind,lapply(foreca
st_rf_x_list,function(x) x[4])))
}

#plot(c(rf_forecast_h1,rf_forecast_h2,rf_forecast_h3,rf_forecast_h4),t="L
")

#CON BASE EN ESTA RUTINA CREAR UN CIRCUITO DE PRONOSTICO IGUAL QUE AL REST
O

# DE MODELOS. IMPORTANTE: PARA ESTE MODELO SOLO GUARDAR PRONÓSTICOS!

table_rf_x_for_h1 <- as.data.frame(table_rf_x_for_h1)
table_rf_x_for_h1 <- matrix(unlist(table_rf_x_for_h1), nrow = nrow(table_r
f_x_for_h1), byrow = TRUE)

table_rf_x_for_h2<-as.data.frame( table_rf_x_for_h2)
table_rf_x_for_h2 <- matrix(unlist(table_rf_x_for_h2), nrow = nrow(table_r
f_x_for_h2), byrow = TRUE)

table_rf_x_for_h3<-as.data.frame( table_rf_x_for_h3)
table_rf_x_for_h3 <- matrix(unlist(table_rf_x_for_h3), nrow = nrow(table_r
f_x_for_h3), byrow = TRUE)

table_rf_x_for_h4<-as.data.frame( table_rf_x_for_h4)
table_rf_x_for_h4 <- matrix(unlist(table_rf_x_for_h4), nrow = nrow(table_r
f_x_for_h4), byrow = TRUE)

```

```

# Lista de pronósticos
# OJO: no contar pronóstico desde 2020q1-202q4 (osea los que se estiman en
2019)
table_rf_x_for=list("h1"=table_rf_x_for_h1[1:(length(table_rf_x_for_h1[,1])
-1),],
                   "h2"=table_rf_x_for_h2[1:(length(table_rf_x_for_h2[,1])
-2),],
                   "h3"=table_rf_x_for_h3[1:(length(table_rf_x_for_h3[,1])
-3),],
                   "h4"=table_rf_x_for_h4[1:(length(table_rf_x_for_h4[,1])
-4),])

# Crear una lista para almacenar los RMSE
mse_rf_x<- data.frame(matrix(nrow=10,ncol=4))
row.names(mse_rf_x)<-names(forecast_list)
colnames(mse_rf_x)<-c("h1", "h2", "h3", "h4")

# Para cada h
for (h in 1:4){
  # Iterar sobre los elementos de la lista forecast_list
  i<-0
  for (col in names(forecast_rf_x_list)){
    i<-i+1
    ts_col <- ts(dd[[col]], start = c(2001, 4), end = c(2019, 4), frequenc
y = 4)
    actual_values <- ts_col[(length(ts_col)-length(table_rf_x_for[[h]][,i]
)+1):length(ts_col)]
    mse_rf_x[i,h]<- round(sqrt(mean((table_rf_x_for[[h]][,i] - actual_valu
es)^2)),5)
  }
}

mse_rf_x

#check h=4 i=10
#plot(table_arima_x3_for[[h]][,i],t="L")
#lines(actual_values,t="L",col="red") # Perfecto! ETS debería pronosticar
en h=4 algo parecido que 4 periodos antes

#####

## Neural network

# Matrices de info
table_nnt_for_h1=c() #distinto para cada h
table_nnt_for_h2=c() #distinto para cada h
table_nnt_for_h3=c() #distinto para cada h
table_nnt_for_h4=c() #distinto para cada h

```

```

table_nnt_model=c() #mismo para todos Los h
table_nnt_aicc=c() #mismo para todos Los h

# Convertir La variable TIME a un objeto de clase "ts"
ts_time <- ts(dd$TIME, start = c(1980, 2), frequency = 4)

# Crear una Lista para almacenar Los pronósticos
forecast_nnt_list <- list()

for (i in 1:length(window(ts_time, start = c(2001, 4), end = c(2019, 4))))
{
  # Iterar sobre todas Las columnas del conjunto de datos, excepto La columna TIME

for (col in names(dd)[c(-1,-12:-17)]) {
  # Crear una serie de tiempo para La variable en Los últimos 20 años
  ts_col <- ts(dd[[col]], start = c(1980, 4), end = c(2019, 4), frequency = 4)

  ts_col_last_20_years <- ts_col[(i):(i+79)]

  # Ajustar un modelo ETS a La serie de tiempo en Los últimos 20 años
  model_col <- nnetar(ts_col_last_20_years)

  # Hacer un pronóstico para Los próximos 4 períodos
  forecast_col <- forecast(model_col, h = 4)

  # Agregar La columna al pronóstico y guardarlo en La lista
  forecast_col$col <- col
  forecast_nnt_list[[col]] <- forecast_col
}

table_nnt_for_h1=rbind(table_nnt_for_h1,Reduce(cbind,lapply(forecast_nnt_list,function(x) x$mean[1])))
table_nnt_for_h2=rbind(table_nnt_for_h2,Reduce(cbind,lapply(forecast_nnt_list,function(x) x$mean[2])))
table_nnt_for_h3=rbind(table_nnt_for_h3,Reduce(cbind,lapply(forecast_nnt_list,function(x) x$mean[3])))
table_nnt_for_h4=rbind(table_nnt_for_h4,Reduce(cbind,lapply(forecast_nnt_list,function(x) x$mean[4])))

table_nnt_model=rbind(table_nnt_model,Reduce(cbind,lapply(forecast_nnt_list,function(x) x$method)))
table_nnt_aicc <- rbind(table_nnt_aicc,Reduce(cbind,lapply(forecast_nnt_list,function(x) x$model$aicc)))

}

# Lista de pronósticos
# OJO: no contar pronóstico desde 2020q1-202q4 (osea Los que se estiman en 2019)
table_nnt_for=list("h1"=table_nnt_for_h1[1:(length(table_nnt_for_h1[,1])-1),],

```

```

),],
      "h2"=table_nnt_for_h2[1:(length(table_nnt_for_h2[,1])-2
),],
      "h3"=table_nnt_for_h3[1:(length(table_nnt_for_h3[,1])-3
),],
      "h4"=table_nnt_for_h4[1:(length(table_nnt_for_h4[,1])-4
),])

# Crear una Lista para almacenar Los RMSE
mse_nnt<- data.frame(matrix(nrow=10,ncol=4))
row.names(mse_nnt)<-names(forecast_nnt_list)
colnames(mse_nnt)<-c("h1", "h2", "h3", "h4")

# Para cada h
for (h in 1:4){
  # Iterar sobre Los elementos de La Lista forecast_list
  i<-0
  for (col in names(forecast_nnt_list)){
    i<-i+1
    ts_col <- ts(dd[[col]], start = c(1980, 4), end = c(2019, 4), frequenc
y = 4)
    actual_values <- ts_col[(length(ts_col)-length(table_nnt_for[[h]][,i]
+1):length(ts_col)]
    mse_nnt[i,h]<- round(sqrt(mean((table_nnt_for[[h]][,i] - actual_values
)^2)),5)
  }
}
# MIRA EL RMSE! A medida que nos alejamos en h, el error de predicci3n aum
enta (correcto!)
mse_nnt

#check h=4 i=10
plot(table_nnt_for[[h]][,i],t="l")
lines(actual_values,t="l",col="red") # Perfecto! ETS deber3a pronosticar e
n h=4 algo parecido que 4 periodos antes

#####

## Neural Network con variable regresora x

# Matrices de info

table_nnt_x_for_h1=c() #distinto para cada h
table_nnt_x_for_h2=c() #distinto para cada h
table_nnt_x_for_h3=c() #distinto para cada h
table_nnt_x_for_h4=c() #distinto para cada h
table_nnt_x_model=c() #mismo para todos Los h
table_nnt_x_aicc=c() #mismo para todos Los h

# Convertir La variable TIME a un objeto de clase "ts"
ts_time <- ts(dd$TIME, start = c(1980, 2), frequency = 4)

```

```

# Crear una lista para almacenar Los pronósticos
forecast_nnt_x_list <- list()

for (i in 1:length(window(ts_time, start = c(2001, 4), end = c(2019, 4))))
{
  # Iterar sobre todas las columnas del conjunto de datos, excepto la columna TIME

  for (col in names(dd)[c(-1,-12:-17)]) {
    # Crear una serie de tiempo para la variable en los últimos 20 años
    ts_col <- ts(dd[[col]], start = c(1980, 4), end = c(2019, 4), frequency = 4)

    ts_col_last_20_years <- ts_col[(i):(i+79)]

    # Ajustar un modelo ETS a la serie de tiempo en los últimos 20 años
    model_col <- nnetar(ts_col, xreg = dd$'Global Economic Conditions Indicator'[c(-1, -159)])

    # Hacer un pronóstico para los próximos 4 períodos
    forecast_col <- forecast(model_col, xreg = tail(dd$'Global Economic Conditions Indicator', 4))

    # Agregar la columna al pronóstico y guardarlo en la lista
    forecast_col$col <- col
    forecast_nnt_x_list[[col]] <- forecast_col
  }

  table_nnt_x_for_h1=rbind(table_nnt_x_for_h1,Reduce(cbind,lapply(forecast_nnt_x_list,function(x) x$mean[1])))
  table_nnt_x_for_h2=rbind(table_nnt_x_for_h2,Reduce(cbind,lapply(forecast_nnt_x_list,function(x) x$mean[2])))
  table_nnt_x_for_h3=rbind(table_nnt_x_for_h3,Reduce(cbind,lapply(forecast_nnt_x_list,function(x) x$mean[3])))
  table_nnt_x_for_h4=rbind(table_nnt_x_for_h4,Reduce(cbind,lapply(forecast_nnt_x_list,function(x) x$mean[4])))

  table_nnt_x_model=rbind(table_nnt_x_model,Reduce(cbind,lapply(forecast_nnt_list,function(x) x$method)))
  table_nnt_x_aicc <- rbind(table_nnt_x_aicc,Reduce(cbind,lapply(forecast_nnt_list,function(x) x$model$aicc)))

}

# Lista de pronósticos
# OJO: no contar pronóstico desde 2020q1-2020q4 (osea los que se estiman en 2019)
table_nnt_x_for=list("h1"=table_nnt_x_for_h1[1:(length(table_nnt_x_for_h1[,1])-1),],
                    "h2"=table_nnt_x_for_h2[1:(length(table_nnt_x_for_h2[,1])-2),],
                    "h3"=table_nnt_x_for_h3[1:(length(table_nnt_x_for_h3[,1]

```

```

])-3),],
                                "h4"=table_nnt_x_for_h4[1:(length(table_nnt_x_for_h4[,1
])-4),])

# Crear una Lista para almacenar Los RMSE
mse_nnt_x<- data.frame(matrix(nrow=10,ncol=4))
row.names(mse_nnt_x)<-names(forecast_nnt_x_List)
colnames(mse_nnt_x)<-c("h1", "h2", "h3", "h4")

# Para cada h
for (h in 1:4){
  # Iterar sobre los elementos de la lista forecast_List
  i<-0
  for (col in names(forecast_nnt_x_List)){
    i<-i+1
    ts_col <- ts(dd[[col]], start = c(1980, 4), end = c(2019, 4), frequenc
y = 4)
    actual_values <- ts_col[(length(ts_col)-length(table_nnt_x_for[[h]][,i
])+1):length(ts_col)]
    mse_nnt_x[i,h]<- round(sqrt(mean((table_nnt_x_for[[h]][,i] - actual_va
lues)^2)),5)
  }
}
# MIRA EL RMSE! A medida que nos alejamos en h, el error de predicción aum
enta (correcto!)
mse_nnt_x

#check h=4 i=10
#plot(table_nnt_x_for[[h]][,i],t="l")
#lines(actual_values,t="l",col="red") # Perfecto! ETS debería pronosticar
en h=4 algo parecido que 4 periodos antes

### EXPORTAR TABLAS

###RMSE
write.table(mse_ets, file = paste0("tablas/rmse_ets",".txt"), sep = ",", q
uote = FALSE, row.names = F)
write.table(mse_arma, file = paste0("tablas/rmse_arma",".txt"), sep = ",
", quote = FALSE, row.names = F)
write.table(mse_arma_x1, file = paste0("tablas/rmse_arma_x1",".txt"), se
p = ",", quote = FALSE, row.names = F)
write.table(mse_arma_x2, file = paste0("tablas/rmse_arma_x2",".txt"), se
p = ",", quote = FALSE, row.names = F)
write.table(mse_arma_x3, file = paste0("tablas/rmse_arma_x3",".txt"), se
p = ",", quote = FALSE, row.names = F)
write.table(mse_nnt, file = paste0("tablas/rmse_nnetar",".txt"), sep = ",",
, quote = FALSE, row.names = F)
write.table(mse_nnt_x, file = paste0("tablas/rmse_nnetar_x",".txt"), sep =
",", quote = FALSE, row.names = F)
write.table(mse_rf, file = paste0("tablas/rmse_rf",".txt"), sep = ",", quo
te = FALSE, row.names = F)

```

```

write.table(mse_rf_x, file = paste0("tablas/rmse_rf_x", ".txt"), sep = ",",
quote = FALSE, row.names = F)

##FORECAST
#ARIMA
for (i in 1:4) {
  table <- table_arima_for[[i]]
  file_name <- paste0("tablas/table_arima_for_h", i, ".txt")
  write.table(table, file = file_name, sep = ",", quote = FALSE, row.names
= FALSE)
}

##modelo
write.table(table_arima_model, file = paste0("tablas/table_arima_model", ".
txt"), sep = ";", quote = FALSE, row.names = F)

##aicc
write.table(table_arima_aicc, file = paste0("tablas/table_arima_aicc", ".tx
t"), sep = ",", quote = FALSE, row.names = F)

#ARIMA_X1
for (i in 1:4) {
  table <- table_arima_x1_for[[i]]
  file_name <- paste0("tablas/table_arima_x1_for_h", i, ".txt")
  write.table(table, file = file_name, sep = ",", quote = FALSE, row.names
= FALSE)
}

##modelo
write.table(table_arima_x1_model, file = paste0("tablas/table_arima_x1_mod
el", ".txt"), sep = ";", quote = FALSE, row.names = F)

##aicc
write.table(table_arima_x1_aicc, file = paste0("tablas/table_arima_x1_aicc
", ".txt"), sep = ",", quote = FALSE, row.names = F)

#ARIMA_X2
for (i in 1:4) {
  table <- table_arima_x2_for[[i]]
  file_name <- paste0("tablas/table_arima_x2_for_h", i, ".txt")
  write.table(table, file = file_name, sep = ",", quote = FALSE, row.names
= FALSE)
}

##modelo
write.table(table_arima_x2_model, file = paste0("tablas/table_arima_x2_mod
el", ".txt"), sep = ";", quote = FALSE, row.names = F)

##aicc
write.table(table_arima_x2_aicc, file = paste0("tablas/table_arima_x2_aicc
", ".txt"), sep = ",", quote = FALSE, row.names = F)

#ARIMA_X3

```



```

for (i in 1:4) {
  table <- table_arima_x3_for[[i]]
  file_name <- paste0("tablas/table_arima_x3_for_h", i, ".txt")
  write.table(table, file = file_name, sep = ",", quote = FALSE, row.names
= FALSE)
}

###modelo
write.table(table_arima_x3_model, file = paste0("tablas/table_arima_x3_mod
el", ".txt"), sep = ";", quote = FALSE, row.names = F)

###aicc
write.table(table_arima_x3_aicc, file = paste0("tablas/table_arima_x3_aicc
", ".txt"), sep = ",", quote = FALSE, row.names = F)

#ETS
for (i in 1:4) {
  table <- table_ets_for[[i]]
  file_name <- paste0("tablas/table_ets_for_h", i, ".txt")
  write.table(table, file = file_name, sep = ",", quote = FALSE, row.names
= FALSE)
}

###modelo
write.table(table_ets_model, file = paste0("tablas/table_ets_model", ".txt"
), sep = ";", quote = FALSE, row.names = F)

###aicc
write.table(table_ets_aicc, file = paste0("tablas/table_ets_aicc", ".txt"),
sep = ",", quote = FALSE, row.names = F)

#NNETAR
for (i in 1:4) {
  table <- table_nnt_for[[i]]
  file_name <- paste0("tablas/table_nnetar_for_h", i, ".txt")
  write.table(table, file = file_name, sep = ",", quote = FALSE, row.names
= FALSE)
}

###modelo
write.table(table_nnt_model, file = paste0("tablas/table_nnetar_model", ".t
xt"), sep = ";", quote = FALSE, row.names = F)

###aicc
write.table(table_nnt_aicc, file = paste0("tablas/table_nnetar_aicc", ".txt
"), sep = ",", quote = FALSE, row.names = F)

#NNETAR_X
for (i in 1:4) {
  table <- table_nnt_x_for[[i]]
  file_name <- paste0("tablas/table_nnetar_x_for_h", i, ".txt")

```

```

write.table(table, file = file_name, sep = ",", quote = FALSE, row.names
= FALSE)
}

###modelo
write.table(table_nnt_model, file = paste0("tablas/table_nnetar_x_model",
.txt"), sep = ";", quote = FALSE, row.names = F)

###aicc
write.table(table_nnt_aicc, file = paste0("tablas/table_nnetar_x_aicc", ".t
xt"), sep = ",", quote = FALSE, row.names = F)

#RF
for (i in 1:4) {
  table <- table_rf_for[[i]]
  file_name <- paste0("tablas/table_rf_for_h", i, ".txt")
  write.table(table, file = file_name, sep = ",", quote = FALSE, row.names
= FALSE)
}

#RF_X
for (i in 1:4) {
  table <- table_rf_x_for[[i]]
  file_name <- paste0("tablas/table_rf_x_for_h", i, ".txt")
  write.table(table, file = file_name, sep = ",", quote = FALSE, row.names
= FALSE)
}

### RESULTADOS

#Redondear resultados

#mse_ets<-round(mse_ets,3)
#mse_arima<-round(mse_arima,3)
#mse_arima_x1<-round(mse_arima_x1,3)
#mse_nnt<-round(mse_nnt,3)
#mse_nnt_x<-round(mse_nnt_x,3)
#mse_rf<-round(mse_rf,3)
#mse_rf_x<-round(mse_rf_x,3)

## Mejor RMSE para cada país y h

# Crear un data frame para almacenar los resultados
best_mse <- data.frame(Fila = character(),
  Columna = character(),
  mse_ets = numeric(),
  mse_arima = numeric(),
  mse_arima_x1 = numeric(),
  mse_nnt = numeric(),
  mse_nnt_x = numeric(),
  mse_rf = numeric(),

```

```

        mse_rf_x = numeric(),
        Mejor_MSE = numeric(),
        stringsAsFactors = FALSE)

# Obtener Los nombres de Las filas y columnas de mse_ets
filas <- rownames(mse_ets)
columnas <- colnames(mse_ets)

# Iterar sobre Las filas y columnas
for (i in 1:nrow(mse_ets)) {
  for (j in 1:ncol(mse_ets)) {
    mse_valores <- c(mse_ets[i, j], mse_arima[i, j], mse_arima_x1[i, j], mse_nnt[i, j], mse_nnt_x[i, j], mse_rf[i, j], mse_rf_x[i, j])

    # Obtener el valor mínimo entre todas Las tablas
    mejor_mse <- min(mse_valores)

    # Agregar Los resultados al data frame
    best_mse <- rbind(best_mse, data.frame(Fila = filas[i],
                                          Columna = columnas[j],
                                          mse_ets = mse_ets[i, j],
                                          mse_arima = mse_arima[i, j],
                                          mse_arima_x1 = mse_arima_x1[i,
j],
                                          mse_nnt = mse_nnt[i, j],
                                          mse_nnt_x = mse_nnt_x[i, j],
                                          mse_rf = mse_rf[i, j],
                                          mse_rf_x = mse_rf_x[i, j],
                                          Mejor_MSE = mejor_mse))
  }
}

best_mse

# Crear un vector para almacenar Los nombres de Las columnas coincidentes
nombres_coincidentes <- character(nrow(best_mse))

# Realizar un bucle for para buscar La columna coincidente para cada fila
for (i in 1:nrow(best_mse)) {
  fila <- best_mse[i, -c(1, ncol(best_mse))] # Excluir La primera columna
y La columna Mejor_MSE

  # Buscar La columna coincidente para La fila actual
  columna_coincidente <- colnames(best_mse)[-c(1, ncol(best_mse))][fila ==
best_mse$Mejor_MSE[i]]

  # Guardar el nombre de La columna coincidente en el vector
  nombres_coincidentes[i] <- columna_coincidente
}

# Crear el dataframe con Los nombres de Las columnas coincidentes
df_nombres_coincidentes <- data.frame(Mejor_Tabla = nombres_coincidentes)

```

```

# Añadir la columna "modelo" al dataframe best_mse
best_mse <- cbind(best_mse, modelo = df_nombres_coincidentes$Mejor_Tabla)

# Mostrar el dataframe resultante
best_mse

write.table(best_mse, file = paste0("tablas/best_rmse", ".txt"), sep = ",",
quote = FALSE, row.names = F)

#####

## ELECCIÓN DEL MODELO

# ETS

modelo_ets <- data.frame(matrix(NA, ncol = 10, nrow = 1))

for (i in 1:10) {
  column_values <- table_ets_model[, i]
  frequencies <- table(column_values)
  most_frequent_value <- names(frequencies)[which.max(frequencies)]
  modelo_ets[1, i] <- most_frequent_value
}

names(modelo_ets) <- rownames(mse_ets)
modelo_ets

# ARIMA

modelo_arima <- data.frame(matrix(NA, ncol = 10, nrow = 1))

for (i in 1:10) {
  column_values <- table_arima_model[, i]
  frequencies <- table(column_values)
  most_frequent_value <- names(frequencies)[which.max(frequencies)]
  modelo_arima[1, i] <- most_frequent_value
}

names(modelo_arima) <- rownames(mse_arima)
modelo_arima

# ARIMA_X1

modelo_arima_x1 <- data.frame(matrix(NA, ncol = 10, nrow = 1))

for (i in 1:10) {
  column_values <- table_arima_x1_model[, i]
  frequencies <- table(column_values)
  most_frequent_value <- names(frequencies)[which.max(frequencies)]
  modelo_arima_x1[1, i] <- most_frequent_value
}

```

```

names(modelo_arima_x1) <- rownames(mse_arima_x1)
modelo_arima_x1

# NNETAR

modelo_nnt <- data.frame(matrix(NA, ncol = 10, nrow = 1))

for (i in 1:10) {
  column_values <- table_nnt_model[, i]
  frequencies <- table(column_values)
  most_frequent_value <- names(frequencies)[which.max(frequencies)]
  modelo_nnt[1, i] <- most_frequent_value
}

names(modelo_nnt) <- rownames(mse_nnt)
modelo_nnt

# NNETAR_X

modelo_nnt_x <- data.frame(matrix(NA, ncol = 10, nrow = 1))

for (i in 1:10) {
  column_values <- table_nnt_x_model[, i]
  frequencies <- table(column_values)
  most_frequent_value <- names(frequencies)[which.max(frequencies)]
  modelo_nnt_x[1, i] <- most_frequent_value
}

names(modelo_nnt_x) <- rownames(mse_nnt_x)
modelo_nnt_x

write.table(modelo_ets, file = paste0("tablas/modelo_ets", ".txt"), sep = ";" , quote = FALSE, row.names = F)
write.table(modelo_arima, file = paste0("tablas/modelo_arima", ".txt"), sep = ";" , quote = FALSE, row.names = F)
write.table(modelo_arima_x1, file = paste0("tablas/modelo_arima_x1", ".txt"), sep = ";" , quote = FALSE, row.names = F)
write.table(modelo_nnt, file = paste0("tablas/modelo_nnt", ".txt"), sep = ";" , quote = FALSE, row.names = F)
write.table(modelo_nnt_x, file = paste0("tablas/modelo_nnt_x", ".txt"), sep = ";" , quote = FALSE, row.names = F)

#####

# APLICACIÓN IMPACTO COVID

# Lectura de datos

```

```

data2020<-read.csv2("gdp2020.csv",header=T,dec=".",sep=",")
head(data2020)

data2020<-read.csv2("gdp2020.csv",sep=";",dec=".")
names(data2020)<-c("LOCATION","INDICATOR","SUBJECT","MEASURE","FREQUENCY",
"TIME","Value","Flag.Codes")
data20_w<-data2020[,c(1,6,7)] %>% pivot_wider(names_from ="LOCATION",value
s_from = "Value" )

ddi<-data20_w[1:5,]
ddi<-ddi[,c(1,2,5,9,15,17,22,27,28,31,40)]
head(ddi)

ddi<-ddi[1:5,]%>%
  select_if(~ !any(is.na(.))) %>%
  mutate_at(vars(-"TIME"), # rest I(1)
            .funs = function(x) log(x/lag(x,1))) %>%
  mutate (TIME=seq(as.Date("2019-10-1"), by = "quarter",
                  length.out = 5-1+1, n=seq(1,by=1,5-1+1)))

ddi %>% pivot_longer(cols = -"TIME",names_to = "Country",values_to = "Value") %>%
  group_by(Country) %>%
  summarise(n=n(),mean=mean(Value,na.rm = T))

ddi<-ddi[-1,]

Y20<-data20_w %>%
  mutate_at(vars(-"TIME"), # rest I(1)
            .funs = function(x) log(x/lag(x,12)))

# urca::ur.ers
colnames(data20_w)

head(ddi)

#Añadimos x1

# Extraer solo las filas de GECON con fechas coincidentes con data_w
GECON_subset <- subset(GECON, TIME %in% ddi$TIME)

ddi[,12]<-GECON_subset[,2]
head(ddi)

#AUS

#Matrices de info

table_au_for_h1=c() #distinto para cada h
table_au_for_h2=c() #distinto para cada h

```

```

table_au_for_h3=c() #distinto para cada h
table_au_for_h4=c() #distinto para cada h
table_au_model=c() #mismo para todos Los h
table_au_aicc=c() #mismo para todos Los h

# Convertir La variable TIME a un objeto de clase "ts"
ts_time <- ts(ddi$TIME, start = c(2020, 1), frequency = 4)

# Crear una lista para almacenar Los pronósticos
forecast_au_list <- list()

for (i in 1:length(window(ts_time, start = c(2020, 1), end = c(2020, 12))))
{
  # Iterar sobre todas Las columnas del conjunto de datos, excepto La columna TIME

for (col in names(ddi)[c(-1,-3:-12)]) {
  # Crear una serie de tiempo para La variable en Los últimos 20 años
  ts_col <- ts(ddi[[col]], start = c(2020, 4), end = c(2020, 12), frequency = 4)

  ts_col_last_20_years <- ts_col[(i):(i+3)]

  # Ajustar un modelo ETS a La serie de tiempo en Los últimos 20 años
  model_col <- nnetar(ts_col_last_20_years)

  # Hacer un pronóstico para Los próximos 4 períodos
  forecast_col <- forecast(model_col, h = 4)

  # Agregar La columna al pronóstico y guardarlo en La lista
  forecast_col$col <- col
  forecast_au_list[[col]] <- forecast_col
}

table_au_for_h1=rbind(table_au_for_h1,Reduce(cbind,lapply(forecast_au_list,function(x) x$mean[1])))
table_au_for_h2=rbind(table_au_for_h2,Reduce(cbind,lapply(forecast_au_list,function(x) x$mean[2])))
table_au_for_h3=rbind(table_au_for_h3,Reduce(cbind,lapply(forecast_au_list,function(x) x$mean[3])))
table_au_for_h4=rbind(table_au_for_h4,Reduce(cbind,lapply(forecast_au_list,function(x) x$mean[4])))

table_au_model=rbind(table_au_model,Reduce(cbind,lapply(forecast_au_list,function(x) x$method)))
table_au_aicc <- rbind(table_au_aicc,Reduce(cbind,lapply(forecast_au_list,function(x) x$model$aicc)))

}

# Lista de pronósticos
# OJO: no contar pronóstico desde 2020q1-2020q4 (osea Los que se estiman en 2019)

```

```

table_au_for=list("h1"=table_au_for_h1[1:(length(table_au_for_h1[,1])-1
),],
                 "h2"=table_au_for_h2[1:(length(table_au_for_h2[,1])-2
),],
                 "h3"=table_au_for_h3[1:(length(table_au_for_h3[,1])-3
),],
                 "h4"=table_au_for_h4[1:(length(table_au_for_h4[,1])-4
),])

#CAN USA OECD

#Matrices de info

table_arima_im_for_h1=c() #distinto para cada h
table_arima_im_for_h2=c() #distinto para cada h
table_arima_im_for_h3=c() #distinto para cada h
table_arima_im_for_h4=c() #distinto para cada h
table_arima_im_model=c() #mismo para todos los h
table_arima_im_aicc=c() #mismo para todos los h

# Convertir la variable TIME a un objeto de clase "ts"
ts_time <- ts(ddi$TIME, start = c(2020, 1), frequency = 4)

# Crear una lista para almacenar los pronósticos
forecast_arima_im_list <- list()

for (i in 1:length(window(ts_time, start = c(2020, 1), end = c(2020, 12))))
{
  # Iterar sobre todas las columnas del conjunto de datos, excepto la columna TIME

  for (col in names(ddi)[c(3,10:11)]) {
    # Crear una serie de tiempo para la variable en los últimos 20 años
    ts_col <- ts(ddi[[col]], start = c(2020, 1), end = c(2020, 12), frequency = 4)

    ts_col_last_20_years <- ts_col[(i):(i+3)]

    # Ajustar un modelo ETS a la serie de tiempo en los últimos 20 años
    model_col <- auto.arima(ts_col_last_20_years)

    # Hacer un pronóstico para los próximos 4 períodos
    forecast_col <- forecast(model_col, h = 4)

    # Agregar la columna al pronóstico y guardarlo en la lista
    forecast_col$col <- col
    forecast_arima_im_list[[col]] <- forecast_col
  }
}

table_arima_im_for_h1=rbind(table_arima_im_for_h1,Reduce(cbind,lapply(fore

```



```

cast_arima_im_list,function(x) x$mean[1]))
table_arima_im_for_h2=rbind(table_arima_im_for_h2,Reduce(cbind,lapply(fore
cast_arima_im_list,function(x) x$mean[2]))))
table_arima_im_for_h3=rbind(table_arima_im_for_h3,Reduce(cbind,lapply(fore
cast_arima_im_list,function(x) x$mean[3]))))
table_arima_im_for_h4=rbind(table_arima_im_for_h4,Reduce(cbind,lapply(fore
cast_arima_im_list,function(x) x$mean[4]))))

table_arima_im_model=rbind(table_arima_im_model,Reduce(cbind,lapply(foreca
st_arima_im_list,function(x) x$method)))
table_arima_im_aicc <- rbind(table_arima_im_aicc,Reduce(cbind,lapply(forec
ast_arima_im_list,function(x) x$model$aicc)))

}

#KOR NOR SWE CHE

#Matrices de info

table_arima_x_for_h1=c() #distinto para cada h
table_arima_x_for_h2=c() #distinto para cada h
table_arima_x_for_h3=c() #distinto para cada h
table_arima_x_for_h4=c() #distinto para cada h
table_arima_x_model=c() #mismo para todos los h
table_arima_x_aicc=c() #mismo para todos los h

# Crear una lista para almacenar los pronósticos
forecast_arima_x_list <- list()

for (i in 1:length(window(ts_time, start = c(2020, 1), end = c(2023, 1))))
{
  # Iterar sobre todas las columnas del conjunto de datos, excepto la colu
mna TIME

for (col in names(ddi)[c(6:9)]) {
  # Crear una serie de tiempo para la variable en los últimos 20 años
  ts_col <- ts(ddi[[col]], start = c(2020, 1), end = c(2020, 12), frequenc
y = 4)

  ts_col_last_20_years <- ts_col[(i):(i+3)]

  # Ajustar un modelo ETS a la serie de tiempo en los últimos 20 años
  model_col <- auto.arima(ts_col_last_20_years, xreg = ddi$'Global Economi
c Conditions Indicator')

  # Hacer un pronóstico para los próximos 4 períodos
  forecast_col <- forecast(model_col, xreg = tail(ddi$'Global Economic Con
ditions Indicator', 4))

  # Agregar la columna al pronóstico y guardarlo en la lista
  forecast_col$col <- col
  forecast_arima_x_list[[col]] <- forecast_col
}

```

```

table_arma_x_for_h1=rbind(table_arma_x_for_h1,Reduce(cbind,lapply(forecast_
st_arma_x_list,function(x) x$mean[1])))
table_arma_x_for_h2=rbind(table_arma_x_for_h2,Reduce(cbind,lapply(foreca
st_arma_x_list,function(x) x$mean[2])))
table_arma_x_for_h3=rbind(table_arma_x_for_h3,Reduce(cbind,lapply(foreca
st_arma_x_list,function(x) x$mean[3])))
table_arma_x_for_h4=rbind(table_arma_x_for_h4,Reduce(cbind,lapply(foreca
st_arma_x_list,function(x) x$mean[4])))

table_arma_x_model=rbind(table_arma_x_model,Reduce(cbind,lapply(forecast
_arma_x_list,function(x) x$method)))
table_arma_x_aicc <- rbind(table_arma_x_aicc,Reduce(cbind,lapply(forecas
t_arma_x_list,function(x) x$model$aicc)))

}

# Lista de pronósticos
# OJO: no contar pronóstico desde 2020q1-2020q4 (osea los que se estiman en
2019)
table_arma_x_for=list("h1"=table_arma_x_for_h1[1:(length(table_arma_x_f
or_h1[,1])-1)],,
                    "h2"=table_arma_x_for_h2[1:(length(table_arma_x_for_h
2[,1])-2)],,
                    "h3"=table_arma_x_for_h3[1:(length(table_arma_x_for_h
3[,1])-3)],,
                    "h4"=table_arma_x_for_h4[1:(length(table_arma_x_for_h
4[,1])-4)],)

#FRA ITA
set.seed(12345)
library(randomForest)

ts_col <- ts(ddi[[4]], start = c(2020, 1), end = c(2020, 12), frequenc
y = 4)

ar1 <- ts_col[4:(length(ts_col)-1)] # AR1
ar2 <- ts_col[3:(length(ts_col)-2)] # AR2
ar3 <- ts_col[2:(length(ts_col)-3)] # AR3
ar4 <- ts_col[1:(length(ts_col)-4)] # AR4
Ytrain <- ts_col[5:(length(ts_col))] # Variable dependiente

tail(cbind(Ytrain, ar1, ar2, ar3, ar4)) # Comprobación

Xtrain <- cbind(ar1, ar2, ar3, ar4)

# h1
Xtest <- c(ts_col[length(ts_col)], ts_col[(length(ts_col)-1)], ts_col[
(length(ts_col)-2)], ts_col[(length(ts_col)-3)])

```

```

rf_model2 <- randomForest(y = Ytrain, x = Xtrain, mtry = ncol(Xtrain)/
3, ntree = 500, xreg = ddi$'Global Economic Conditions Indicator')

fra_forecast_h1 <- predict(rf_model2, Xtest)

# h2
Xtest <- c(rf_forecast_h1, ts_col[length(ts_col)], ts_col[(length(ts_c
ol)-1)], ts_col[(length(ts_col)-2)])
fra_forecast_h2 <- predict(rf_model2, Xtest)

# h3
Xtest <- c(rf_forecast_h2, rf_forecast_h1, ts_col[length(ts_col)], ts_
col[(length(ts_col)-1)])
fra_forecast_h3 <- predict(rf_model2, Xtest)

# h4
Xtest <- c(rf_forecast_h3, rf_forecast_h2, rf_forecast_h1, ts_col[leng
th(ts_col)])
fra_forecast_h4 <- predict(rf_model2, Xtest)

#plot(c(fra_forecast_h1,fra_forecast_h2,fra_forecast_h3,fra_forecast_h4),t
="L")

set.seed(12345)
library(randomForest)

ts_col <- ts(ddi[[5]], start = c(1980, 4), end = c(2019, 4), frequency
= 4)

ar1 <- ts_col[4:(length(ts_col)-1)] # AR1
ar2 <- ts_col[3:(length(ts_col)-2)] # AR2
ar3 <- ts_col[2:(length(ts_col)-3)] # AR3
ar4 <- ts_col[1:(length(ts_col)-4)] # AR4
Ytrain <- ts_col[5:(length(ts_col))] # Variable dependiente

tail(cbind(Ytrain, ar1, ar2, ar3, ar4)) # Comprobación

Xtrain <- cbind(ar1, ar2, ar3, ar4)

# h1
Xtest <- c(ts_col[length(ts_col)], ts_col[(length(ts_col)-1)], ts_col[
(length(ts_col)-2)], ts_col[(length(ts_col)-3)])
rf_model2 <- randomForest(y = Ytrain, x = Xtrain, mtry = ncol(Xtrain)/
3, ntree = 500, xreg = ddi$'Global Economic Conditions Indicator')

ita_forecast_h1 <- predict(rf_model2, Xtest)

# h2
Xtest <- c(rf_forecast_h1, ts_col[length(ts_col)], ts_col[(length(ts_c
ol)-1)], ts_col[(length(ts_col)-2)])
ita_forecast_h2 <- predict(rf_model2, Xtest)

```

```

# h3
Xtest <- c(rf_forecast_h2, rf_forecast_h1, ts_col[length(ts_col)], ts_
col[(length(ts_col)-1)])
ita_forecast_h3 <- predict(rf_model2, Xtest)

# h4
Xtest <- c(rf_forecast_h3, rf_forecast_h2, rf_forecast_h1, ts_col[leng
th(ts_col)])
ita_forecast_h4 <- predict(rf_model2, Xtest)

#plot(c(ita_forecast_h1,ita_forecast_h2,ita_forecast_h3,ita_forecast_h4),
t="L")

write.table(dd, file = paste0("tablas/datos",".txt"), sep = ";", quote = F
ALSE, row.names = F)
write.table(ddi, file = paste0("tablas/datos20",".txt"), sep = ";", quote
= FALSE, row.names = F)

```