



INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO

Planeación y Control de la Producción

Forecast in Automotive Industry

Andrea de Anda Kuri 173347

Juliana Hernández Ottalengo 171405

Silvestre Leonardo González Abreu 170404

20 Septiembre 2020

La industria automotriz	3
Manipulación de los datos	6
Estructura de los documentos descargados	6
Código de R para manipular los datos	6
Metodología de pronóstico	8
DAUPSA antes del COVID-19 (hasta marzo 2020)	8
Análisis de la demanda	8
Búsqueda del mejor método para pronosticar	9
Pronóstico con el método seleccionado	10
DAUPSA incluyendo el COVID-19	12
Análisis de la demanda	12
Búsqueda del mejor método para pronosticar	13
Pronóstico con el método seleccionado	14
Ventas de autos híbridos en México (toda la serie de tiempo)	16
Análisis de la demanda	16
Búsqueda del mejor método para pronosticar	17
Pronóstico con el método seleccionado	19
Conclusiones	23
Pronósticos DAUPSA	23
Ventas de autos híbridos en México	25
Código documentado en R	27
DAUPSA antes del COVID (hasta marzo 2020)	27
DAUPSA incluyendo el COVID-19	30
Ventas de autos híbridos en México (toda la serie de tiempo)	32
Fuentes de consulta	37

1. La industria automotriz

Los automóviles se introdujeron en México en el año de 1908, inicialmente estos vehículos alcanzaban precios muy altos, pues la oferta era muy reducida debido a obstáculos en la importación y la técnica de ese tiempo. Durante los años de 1908 a 1916 entraron en promedio al país solamente 215 unidades, sin embargo en 1916 al declararse una huelga que paralizó los tranvías se comenzó a utilizar más el automóvil. Después durante la revolución, se observó un retroceso en la industria causado por los problemas en la importación. Al contrario, en Estados Unidos de manera simultánea la industria se estaba desarrollando de una manera impresionante dedicada al transporte de pasajeros y bienes.

Al terminar el movimiento armado de la Primera Guerra Mundial la demanda había aumentado en México, durante los años de 1917 a 1924 la demanda en el país era de 6,000 autos anuales y se pronosticó que para 1925 las unidades en circulación serían cercanas a 54,000 vehículos. En este año la primera compañía que estableció una ensambladora en el país fue la Ford Motor Company.

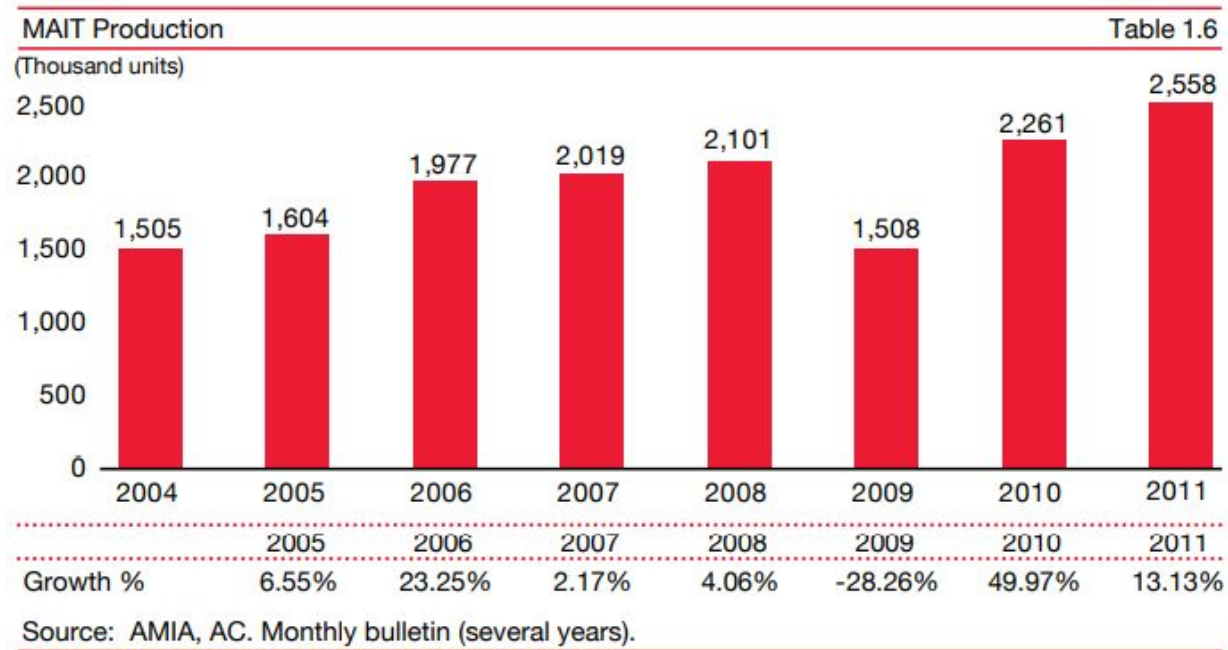
En los años 1937-1948 se lleva a cabo una etapa llamada “Desarrollo del Mercado”, en 1939 se establece en México la empresa “Fábricas Automex” que se encargaba de ensamblar vehículos de marcas como Chrysler, Dodge, Plymouth, DE Soto y Fargo, en el mismo año entró al mercado Armadora Automotriz también.

Durante la Segunda Guerra Mundial, la industria americana y de varios países cambiaron su configuración, en vez de fabricar autos de pasajeros y transporte se dedicó gran parte de la producción a tanques y vehículos blindados de combate.

A partir del año 1944 comienza a aparecer la industria dedicada a la construcción de camiones de carga y en 1951 Diesel se estableció en el país. En el año de 1970 operaban en México 17 compañías importadoras de vehículos. En la actualidad la industria está caracterizada por una globalización incremental, es decir hay una naturaleza multinacional no solo en los productores de autos, si no también en los mercados. En 1980 las compañías grandes produjeron entre el 10% y 50% de su demanda en otros países.

En 2008 la industria automotriz sufrió grandes cambios como reestructuraciones y presupuestos restringidos debido a la crisis económica, además en 2010 la hegemonía de Norte America llegó a su fin con el desarrollo de la industria en Japón y China.

En 2009 la industria automotriz en México tuvo una baja de 26% en el PIB y de 28% en el número total de unidades producidas debido a la misma crisis, como se muestra en la gráfica.

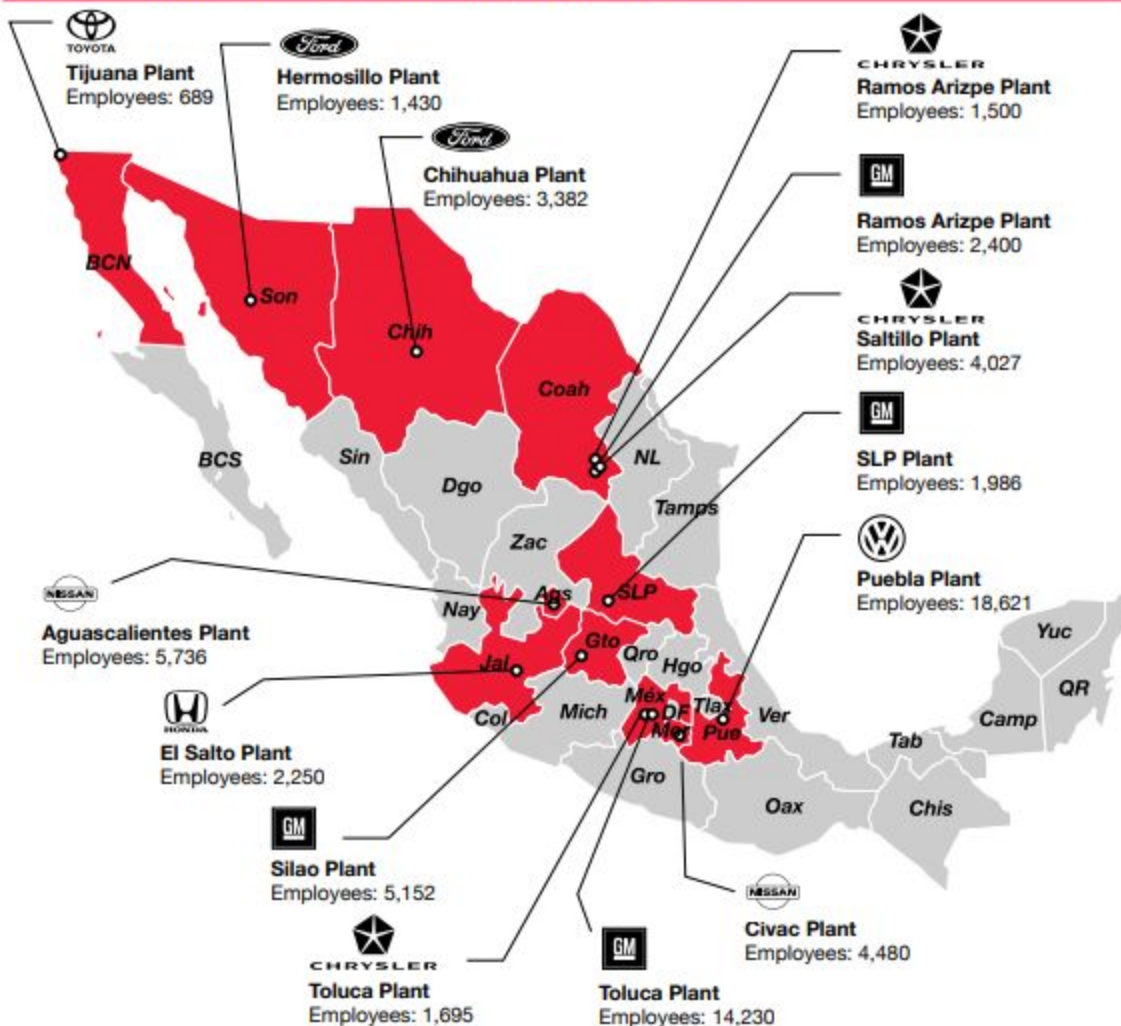


Tomado de PwC Mexico, The Automotive Industry

Pero la industria se recuperó poco a poco, en 2011 la exportación de vehículos a Canadá y Estados Unidos fue 6.7% mayor que en 2010.

Actualmente la industria automotriz es importante en el país no solo por su efecto multiplicador en la economía mexicana en general si no también por su capacidad de generar trabajos bien remunerados. En el 2012 dieciocho de las empresas manufactureras de vehículos tienen plantas de producción y ensamble en México.

Geographical distribution of plants and employees of light vehicles



Tomado de PwC Mexico, *The Automotive Industry*

El año 2020 con algunos desafíos para la industria automotriz en México, uno de los cuales eran los cambios debidos al nuevo acuerdo comercial T-MEC y el otro es revertir la tendencia negativa que ha habido en la venta interna de vehículos y a esto habrá que sumarle la no esperada pandemia mundial del COVID-19.

2. Manipulación de los datos

Estructura de los documentos descargados

El primer documento descargado fue “DAUPSA” utilizado para información sobre la producción doméstica de autos en Estados Unidos. Este era un archivo de excel, conformado por una sola hoja llamada “FRED Graph”, en esta las primeras 10 filas estaban sucias, pues en ellas había datos que no nos servían para el análisis en R, datos como links, títulos o fuentes. A partir de la onceava fila se encontraban los datos de producción del año 1993 a 2020, en la primera columna la fecha (formada por año-mes-día) y en la segunda columna el dato de producción de autos mensual en miles de unidades.

Para leer los datos se utilizó la siguiente instrucción:

```
domesticAutoProduction <- read_excel("DAUPSA.xlsx", skip = 10, sheet =  
"FRED Graph", col_names = TRUE)
```

La segunda base de datos es sobre la venta de autos híbridos y eléctricos en México, este también es un documento de Excel llamado “RAIAVL”. En el documento original las ventas están clasificadas por año y por entidad federativa. A partir de la fila 4 se encuentra la información ordenada por columnas, la primera columna indica el año, la segunda el mes, en la tercera se encuentra la entidad federativa y en la última las unidades vehiculares híbridas vendidas. En este libro se observa el signo “-” cuando no hay datos de ese año, mes y entidad en específico.

Para leer los datos de esta base se utilizó la siguiente línea de código:

```
hybridCarsMX <- read_excel("RAIAVL_11.xlsx", skip = 3, sheet = "RAIAVL_11",  
col_names = FALSE)
```

Código de R para manipular los datos

La primera modificación importante de los datos que se tuvo que hacer en R fue en la base de datos llamada RAIAVL (venta de autos híbridos y eléctricos en México), y en ella se reemplazaron los guiones (-) por un cero (0), para poder utilizar los datos. Se utilizó el siguiente código:

```
for (i in 1:1782) {  
  if(hybridCarsMX[i,4] == '-'){  
    hybridCarsMX[i,4] <- '0'  
  }  
}
```

Una vez que todos los datos de la base se hicieron numéricos se utilizaron una pareja de fors anidados para poder hacer una suma de los datos y poder manejarlos en una serie de tiempo dividida en períodos y no en los diferentes estados del país. Dicha base de datos contiene un total de 1782 datos, divididos en 54 meses y 33 entidades (32 reales y 1 para no especificados), por lo que el for exterior recorre los 54 meses mientras que el interior realiza el agregado de las 33 entidades.

El código que se utilizó fue:

```
#Modificamos los datos para tomar el agregado cada mes
for (i in 0:53) { #1749=53*33
  suma<-0
  aux<-i*33
  for (j in 1:33) {
    #aux<-aux+j
    suma<-suma+as.double(hybridCarsMX[aux+j,4])
  }
  if (i==0){
    hybridCarsMXData<-c(suma)
  }else{
    hybridCarsMXData <- c(hybridCarsMXData, suma)
  }
}
```

3. Metodología de pronóstico

DAUPSA antes del COVID-19 (hasta marzo 2020)

Análisis de la demanda

Para poder realizar el pronóstico de la producción doméstica de automóviles antes del COVID-19 se tuvieron que acotar los datos en la serie de tiempo hasta marzo del 2020.

```
tsDomesticAutoProduction <- ts(domesticAutoProduction[,2], start =  
c(1993,1), end = c(2020,3), frequency = 12)
```

Una vez que se obtuvo una serie de tiempo de frecuencia mensual (12), se graficaron los datos para poder observarlos.

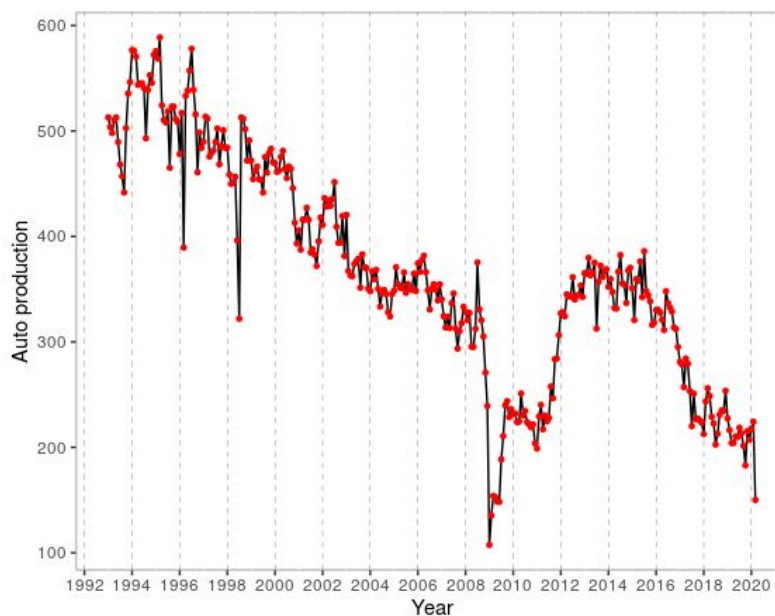


Figura 1: Producción doméstica de automóviles entre enero de 1993 y marzo de 2020

Aunque a simple vista parece que se puede hacer un pronóstico, fue necesario comprobar las condiciones de no aleatoriedad de la serie de tiempo por medio de las pruebas de autocorrelación y el valor-P de Ljung-box.

```
ggAcf(tsDomesticAutoProduction, lag.max = 342)
```

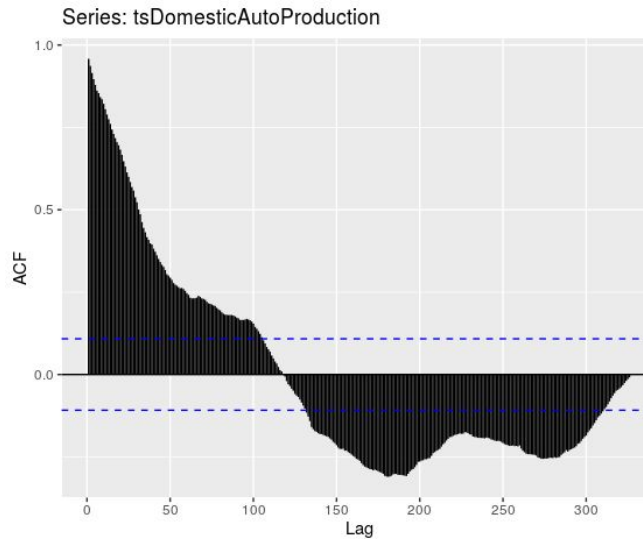



Figura 2: Correlograma de la serie de tiempo DAUPSA antes del COVID-19

En el correlograma se puede ver claramente que las correlaciones son significativamente diferentes de cero, por lo que se puede hacer un análisis de la producción.

Por otro lado, para comprobar el valor-P se utilizó el siguiente código:

```
Box.test(tsDomesticAutoProduction, lag = 1, type = "Ljung")
```

Que arroja el siguiente resultado: X-squared = 302.85, df = 1, p-value < 2.2e-16

En donde el valor-P es menor que nuestro nivel de significancia de 0.05 y nos indica que la probabilidad de que los datos sean aleatorios es prácticamente cero. De esta manera contamos con evidencia estadística para concluir que los datos **no** son aleatorios y podemos hacer un pronóstico.

Búsqueda del mejor método para pronosticar

Se dividieron los datos en dos conjuntos, el de entrenamiento (training data) y el de prueba (test data). El conjunto de prueba se eligió de seis meses (pues es el tiempo que se va a pronosticar) con las siguientes líneas de código:

```
trainingData <- window(tsDomesticAutoProduction, end = c(2019, 9))
testData <- window(tsDomesticAutoProduction, start = c(2019, 10))
```

Se probaron los siguientes métodos de pronóstico: promedio, naive, naive estacional, suavización exponencial simple, Holt, Holt-Winters aditivo y Holt-Winters multiplicativo. A continuación se muestra el código correspondiente y la gráfica de los distintos pronósticos sobre el conjunto de entrenamiento.

```
averageMethod <- meanf(trainingData, h=6)
naiveMethod <- naive(trainingData, h= 6)
seasonalNaiveMethod <- snaive(trainingData, h=6)
fcSimpleExpSmoo <- ses(trainingData, initial = c("optimal"), h= 6)
holtForecast <- holt(trainingData, h=6)
hwAdditive <- hw(trainingData, seasonal = "additive", h= 6)
hwMultiplicative <- hw(trainingData, seasonal = "multiplicative", h = 6)
```

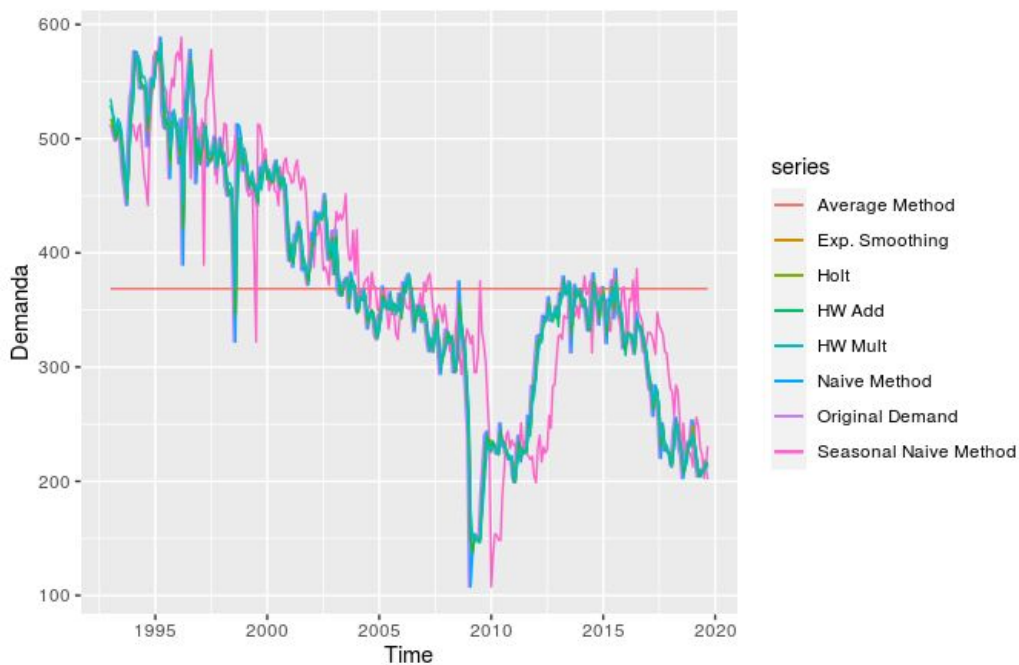


Figura 3: Series del conjunto de entrenamiento de DAUPSA antes del COVID-19

Y al correr la función *accuracy* se decidió que el método que mejor se ajusta a los datos es el de Holt, al tener la menor raíz del error cuadrático medio (RMSE) con respecto al set de entrenamiento y la menor U de Theil para el set de prueba..

```
accuracy(holtForecast, testData)
```

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	-0.007054991	25.58583	17.20231	-0.4680077	5.103026	0.4222483	0.01284708	NA
Test set	-1.742996193	25.41257	21.46345	-2.8047853	11.816589	0.5268421	-0.20204335	0.7020922

Pronóstico con el método seleccionado

Finalmente, se corrió el método seleccionado para toda la serie de tiempo y se pronosticó a seis meses:

```
theSelectedMethod <- holt(tsDomesticAutoProduction, h = 6)
```

Del pronóstico se obtuvo la siguiente gráfica:

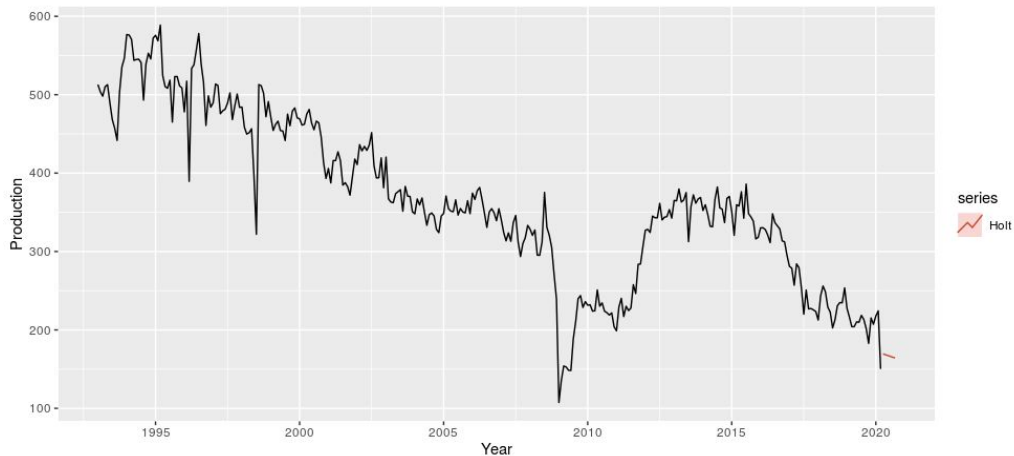


Figura 4: Pronóstico de la producción de autos doméstica previa al COVID-19 por medio del método de Holt

Una vez que se hizo el análisis de residuales se obtuvo el siguiente correlograma:

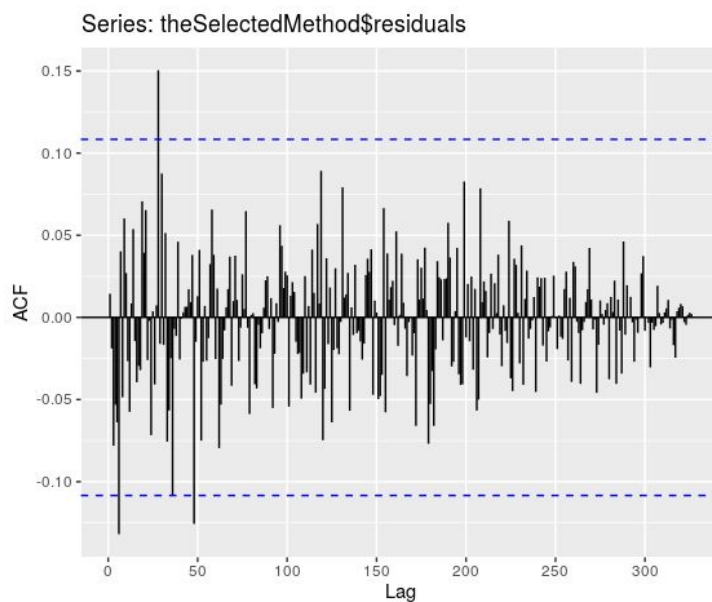


Figura 5: Correlograma de los residuales del pronóstico realizado por medio de Holt

Como las correlaciones de los residuales no son significativamente distintas de cero y al correr la prueba de Ljung-Box se obtuvo lo siguiente: $X\text{-squared} = 0.068605$, $df = 1$, $p\text{-value} = 0.7934$. Con un valor-P mayor a 0.05, poseemos evidencia estadística para concluir que los residuales **son** aleatorios y que el pronóstico probablemente no puede ser mejorado aunque se usara otro método.

DAUPSA incluyendo el COVID-19

Análisis de la demanda

Para realizar el pronóstico de la producción doméstica de automóviles se tomaron todos los datos de la base de datos.

```
tsDomesticAutoProduction <- ts(domesticAutoProduction[,2], start =  
c(1993,1), frequency = 12)
```

Una vez que se obtuvo una serie de tiempo de frecuencia mensual (12), se graficaron los datos para poder observarlos.

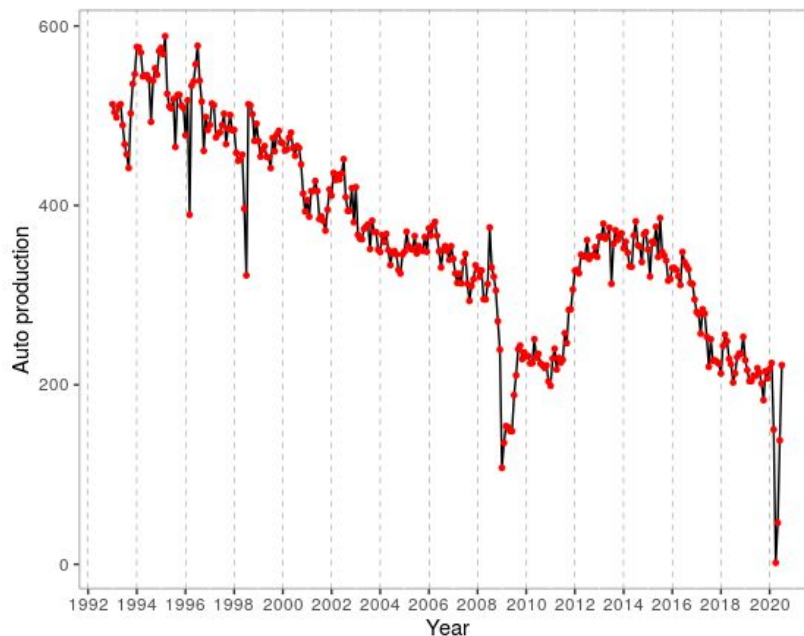


Figura 6: Producción doméstica de automóviles entre enero de 1993 y julio de 2020

Se comprobaron las condiciones de no aleatoriedad de la serie de tiempo por medio de las pruebas de autocorrelación y el valor-P de Ljung-box.

```
ggAcf(tsDomesticAutoProduction, lag.max = 342)
```

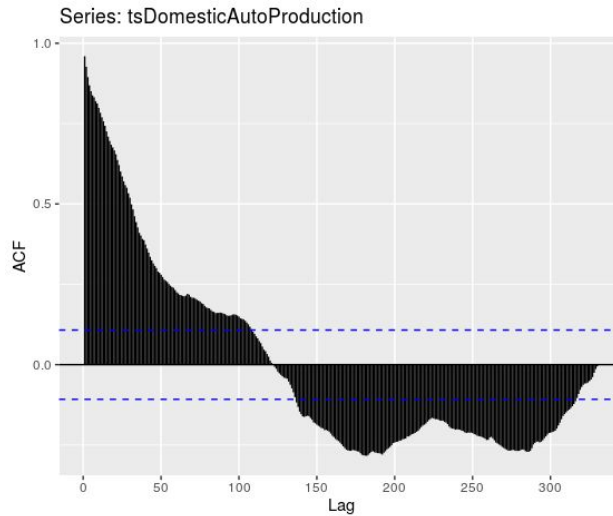


Figura 7: Correlograma de la serie de tiempo DAUPSA incluyendo el COVID-19

En el correlograma se puede ver claramente que las correlaciones son significativamente diferentes de cero y mantienen un patrón, por lo que probablemente se puede hacer un análisis de la producción.

Por otro lado, para comprobar el valor-P de la aleatoriedad se utilizó el siguiente código para implementar la prueba de Ljung-box:

```
Box.test(tsDomesticAutoProduction, lag = 1, type = "Ljung")
```

Que arroja el siguiente resultado: X-squared = 307.49, df = 1, p-value < 2.2e-16

En donde el valor-P es menor que 0.05, por lo que sabemos que los datos **no** son aleatorios y podemos hacer un pronóstico.

Búsqueda del mejor método para pronosticar

Se dividieron los datos en dos conjuntos, el de entrenamiento (training data) y el de prueba (test data). El conjunto de prueba se eligió de seis meses (pues es el tiempo que se va a pronosticar) con las siguientes líneas de código:

```
trainingData <- window(tsDomesticAutoProduction, end = c(2020, 1))
testData <- window(tsDomesticAutoProduction, start = c(2020, 2))
```

Se probaron los siguientes métodos de pronóstico: promedio, naive, naive estacional, suavización exponencial simple, Holt, Holt-Winters aditivo y Holt-Winters multiplicativo. A continuación se muestra el código correspondiente y la gráfica de los distintos pronósticos sobre el conjunto de entrenamiento.

```
averageMethod <- meanf(trainingData, h=6)
naiveMethod <- naive(trainingData, h= 6)
seasonalNaiveMethod <- snaive(trainingData, h=6)
```

```
fcSimpleExpSmoo <- ses(trainingData, initial = c("optimal"), h= 6)
holtForecast <- holt(trainingData, h=6)
hwAdditive <- hw(trainingData, seasonal = "additive", h= 6)
hwMultiplicative <- hw(trainingData, seasonal = "multiplicative", h = 6)
```

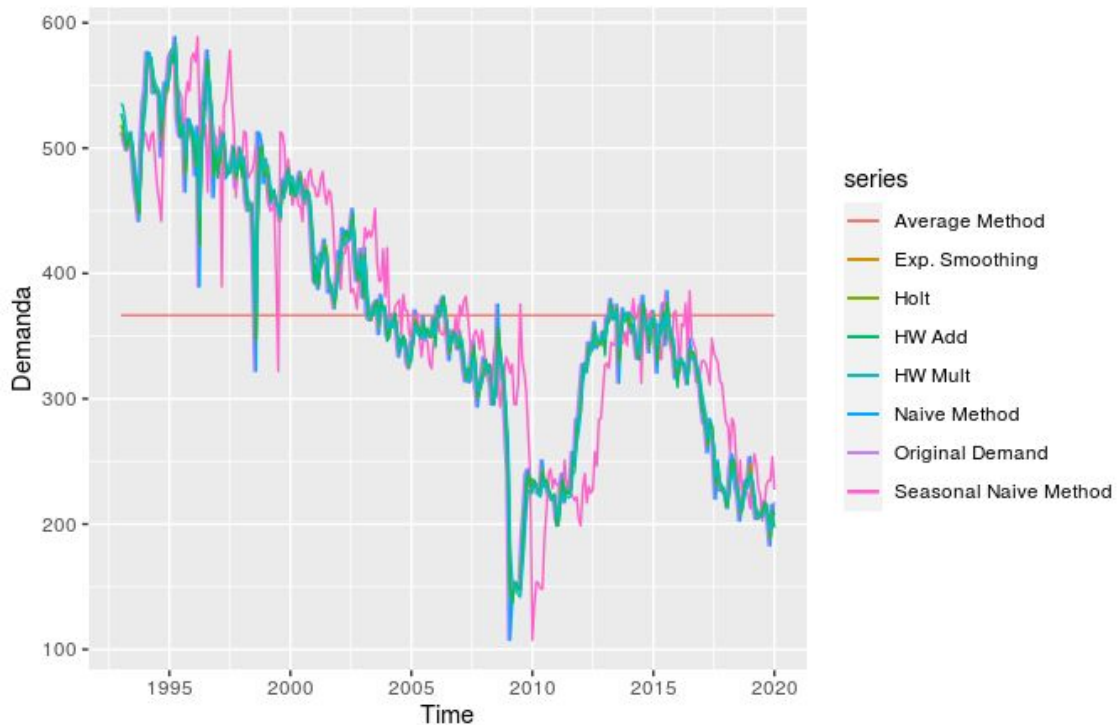


Figura 8: Series del conjunto de entrenamiento de DAUPSA incluyendo el COVID-19

Y al correr la función *accuracy* se decidió que el método que mejor se ajusta a los datos es el de Holt, al tener la menor raíz del error cuadrático medio (RMSE) con respecto al set de entrenamiento y la menor U de Theil para el set de prueba. Es notable que los efectos del COVID-19 elevan los errores del set de prueba, pues representan una disminución abrupta y repentina en la cantidad de autos producidos. Si se analizara de manera aislada podría parecer un cambio aleatorio en la serie de tiempo.

```
> accuracy(holtForecast, testData)
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1 Theil's U
Training set -0.007118058 25.50861 17.18793  -0.4585527  5.13581 0.4230751 0.01307773    NA
Test set     -80.703976416 115.76081 88.72355 -2130.8068946 2134.40262 2.1838998 0.24730132 3.681958
```

Pronóstico con el método seleccionado

Finalmente, se corrió el método seleccionado para toda la serie de tiempo:

```
theSelectedMethod <- holt(tsDomesticAutoProduction, h = 6)
```

Del pronóstico se obtuvo la siguiente gráfica:

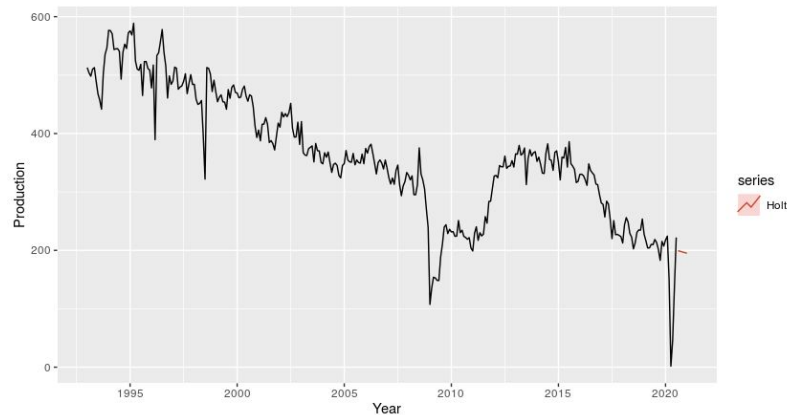


Figura 9: Pronóstico de la producción de autos doméstica incluyendo al COVID-19 por medio del método de Holt

Una vez que se hizo el análisis de residuales se obtuvo el siguiente correlograma:

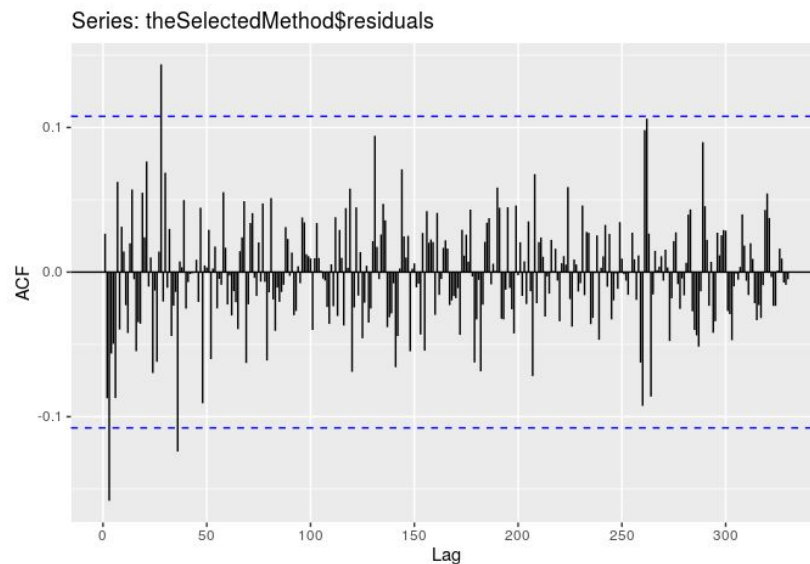


Figura 10: Correlograma de los residuales del pronóstico realizado por medio de Holt

Como las correlaciones de los residuales no son significativamente distintas de cero y al correr la prueba de Ljung-Box se obtuvo lo siguiente: : $X\text{-squared} = 0.23451$, $df = 1$, $p\text{-value} = 0.6282$ Por lo que podemos concluir que los residuales son aleatorios, pues el valor-P es considerablemente mayor a 0.05.

Ventas de autos híbridos en México (toda la serie de tiempo)

Análisis de la demanda

Para poder realizar el pronóstico de la producción doméstica de automóviles se tuvo que manipular los datos originales, como se mencionó previamente, para obtener las ventas mensuales.

```
tsHybridCarsMX <- ts(hybridCarsMXData, start = c(2016,1), frequency = 12 )
```

Una vez que se obtuvo una serie de tiempo de frecuencia mensual (12), se graficaron los datos para poder observarlos.

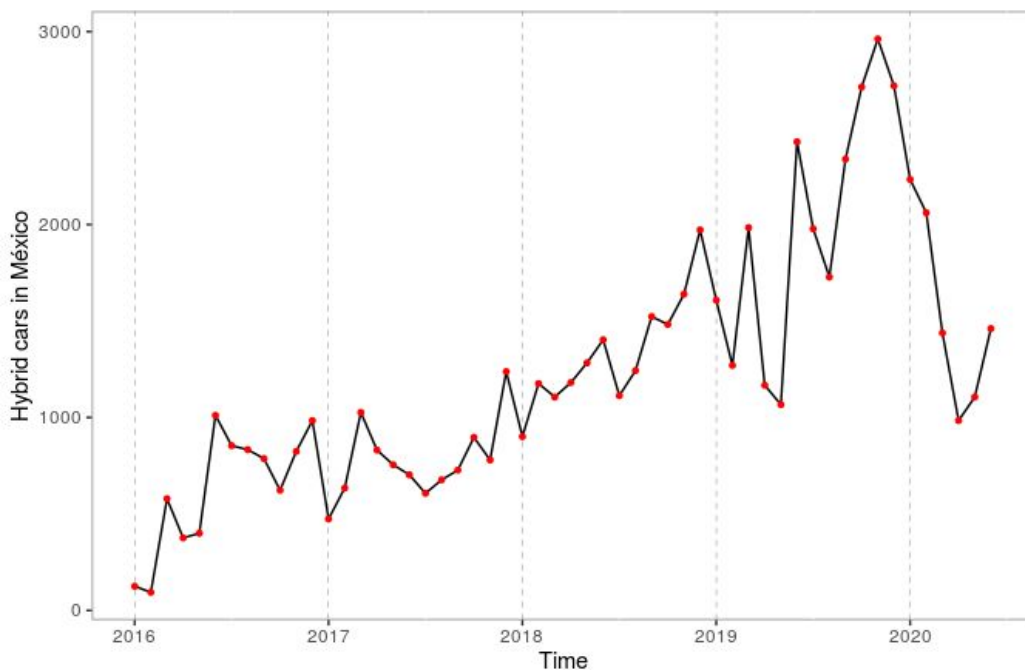


Figura 11: Ventas de autos híbridos en México entre enero de 2016 y junio de 2020

Aunque a simple vista parece que se puede hacer un pronóstico, fue necesario comprobar las condiciones de no aleatoriedad de la serie de tiempo por medio de las pruebas de autocorrelación y el valor-P de Ljung-box.

```
ggAcf(tsHybridCarsMX, lag.max = 54)
```

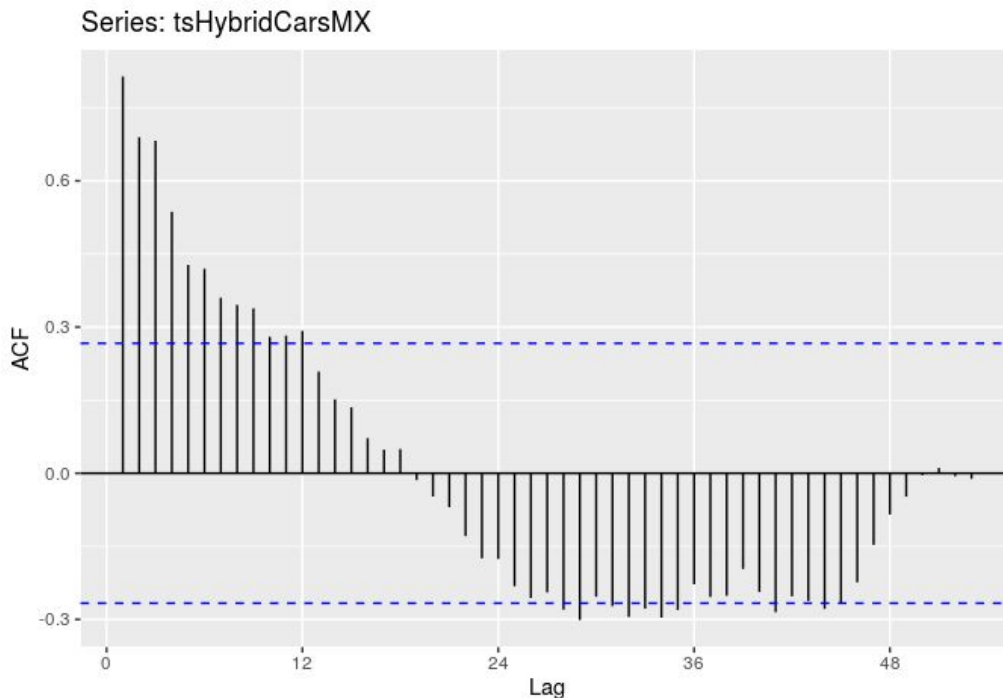



Figura 12: Correlograma de la serie de tiempo de ventas de autos híbridos en México

En el correlograma se puede ver que gran parte de las correlaciones son significativamente diferentes de cero, por lo que probablemente se puede hacer un análisis de la serie.

Por otro lado, para comprobar el valor-P de la aleatoriedad se utilizó el siguiente código para implementar la prueba de Ljung-box:

```
Box.test(tsHybridCarsMX, lag = 1, type = "Ljung")
```

Que arroja el siguiente resultado: X-squared = 37.822, df = 1, p-value = 7.75e-10

En donde el valor-P es menor que nuestro nivel de significancia de 0.05 y nos indica que la probabilidad de que los datos sean aleatorios es prácticamente cero. De esta manera contamos con evidencia estadística para concluir que los datos **no** son aleatorios y podemos hacer un pronóstico.

Búsqueda del mejor método para pronosticar

Se dividieron los datos en training data y test data, con un test data de seis meses (pues es el tiempo que se va a pronosticar) con las siguientes líneas de código:

```
trainingData <- window(tsHybridCarsMX, end = c(2019, 12))
testData <- window(tsHybridCarsMX, start = c(2020, 1))
```

Se probaron los siguientes métodos de pronóstico: promedio, naive, naive estacional, suavización exponencial simple, Holt, Holt-Winters aditivo y Holt-Winters

multiplicativo; todos con un horizonte de 6 periodos. A continuación se muestra el código correspondiente y la gráfica de los distintos pronósticos sobre el conjunto de entrenamiento.

```
averageMethod <- meanf(trainingData, h=6)
naiveMethod <- naive(trainingData, h= 6)
seasonalNaiveMethod <- snaive(trainingData, h=6)
fcSimpleExpSmoo <- ses(trainingData, initial = c("optimal"), h= 6)
holtForecast <- holt(trainingData, h=6)
hwAdditive <- hw(trainingData, seasonal = "additive", h= 6)
hwMultiplicative <- hw(trainingData, seasonal = "multiplicative", h = 6)
```

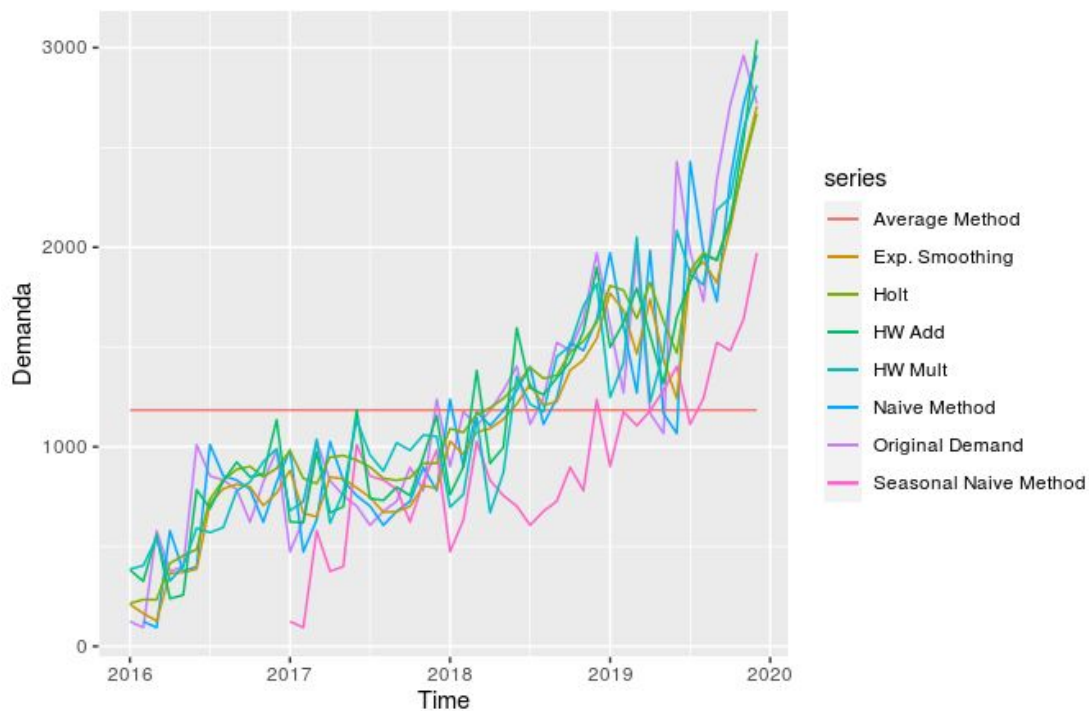


Figura 13: Series del conjunto de entrenamiento de ventas de autos híbridos en México

Y al correr la función *accuracy* se decidió inicialmente que el mejor método sería el de Holt Winters Multiplicativo, pues presentó la menor RMSE con el conjunto de entrenamiento y la menor U de Theil para el conjunto de prueba.

```
> accuracy(hwMultiplicative, testData)
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1 Theil's U
Training set  24.41537 238.3946 188.7895 -11.88491 28.41312 0.3661076 0.3565352      NA
Test set     -812.44707 1107.0074 922.4648 -65.54324 70.46794 1.7888781 0.1538340  3.766436
```

Pronóstico con el método seleccionado

Se corrió el método seleccionado de Holt-Winters Multiplicativo para toda la serie de tiempo para realizar el pronóstico a seis meses.

```
theSelectedMethod <- hw(tsHybridCarsMX, seasonal = "multiplicative", h = 6)
```

Del pronóstico se obtuvo la siguiente gráfica:

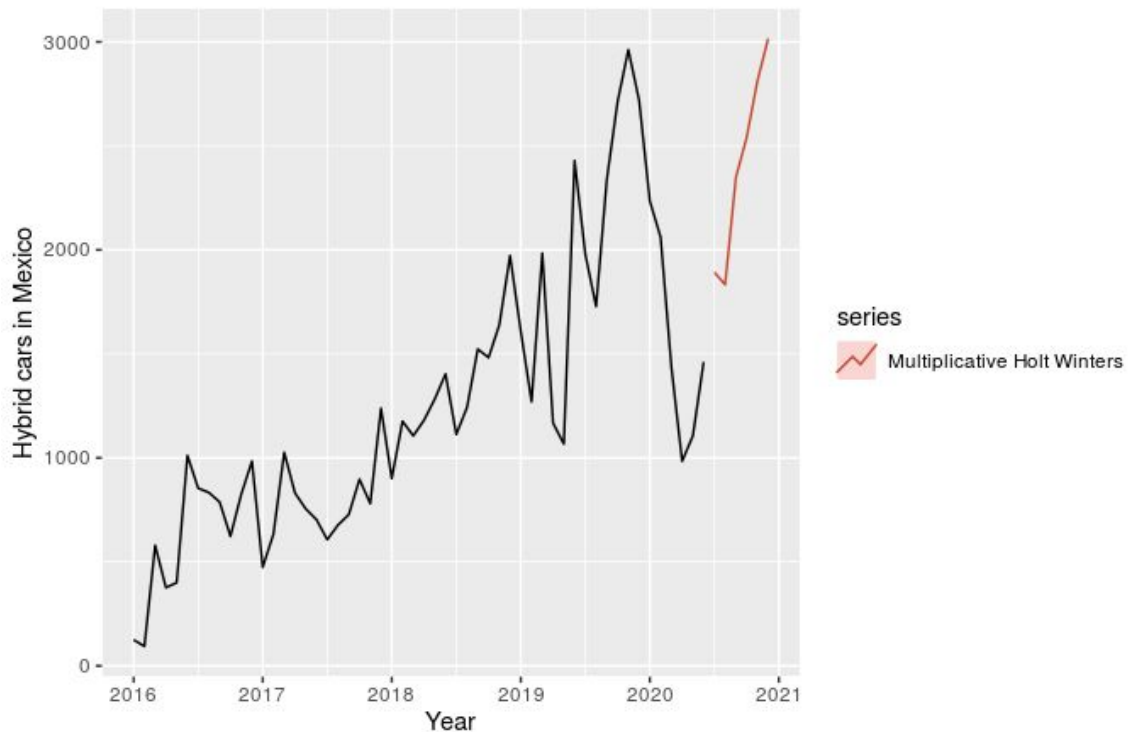


Figura 14: Pronóstico de las ventas de autos híbridos en México por medio del método de Holt-Winters multiplicativo

Una vez que se hizo el análisis de residuales se obtuvo el siguiente correlograma:

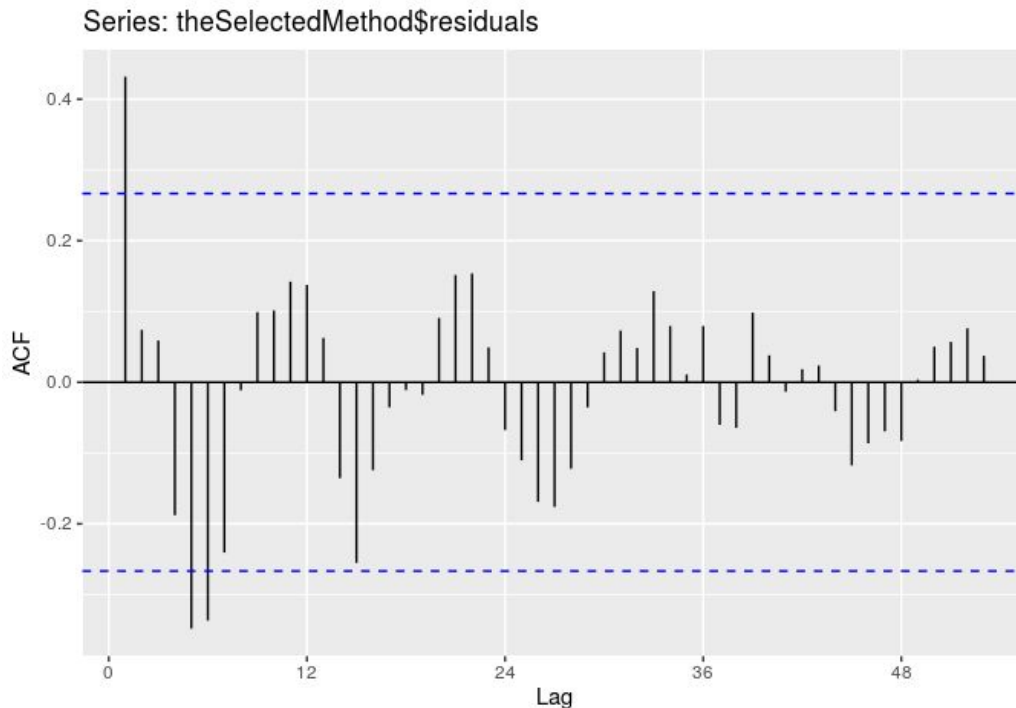


Figura 15: Correlograma de los residuales del pronóstico realizado con el método Holt-Winters multiplicativo

Se observa en el correlograma que sólo uno de los datos es significativamente distinto de cero, pero se observa un patrón en las correlaciones. Además, se obtuvo lo siguiente al realizar la prueba de Ljung-box sobre los residuales: $X\text{-squared} = 10.635$, $df = 1$, $p\text{-value} = 0.00111$. El valor-P menor a 0.05 nos indica, junto con lo observado en el correlograma, que los residuales probablemente no son aleatorios y el pronóstico puede ser mejorado con otro método.

Se selecciona el método con la segunda mejor RMSE y U de Theil, que en este caso es de Holt-Winters aditivo y se repite el procedimiento.

Se corrió el método seleccionado de Holt-Winters aditivo para toda la serie de tiempo para realizar el pronóstico a seis meses.

```
theSelectedMethod <- hw(tsHybridCarsMX, seasonal = "additive", h = 6)
```

Del pronóstico se obtuvo la siguiente gráfica:

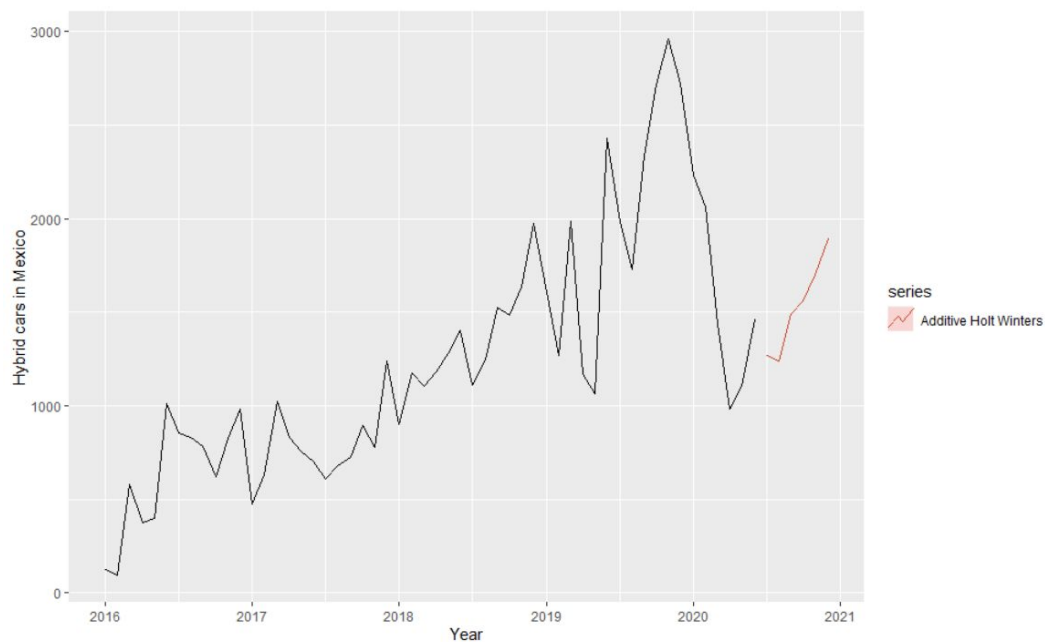


Figura 16: Pronóstico de las ventas de autos híbridos en México por medio del método de Holt-Winters aditivo

Una vez que se hizo el análisis de residuales se obtuvo el siguiente correlograma:

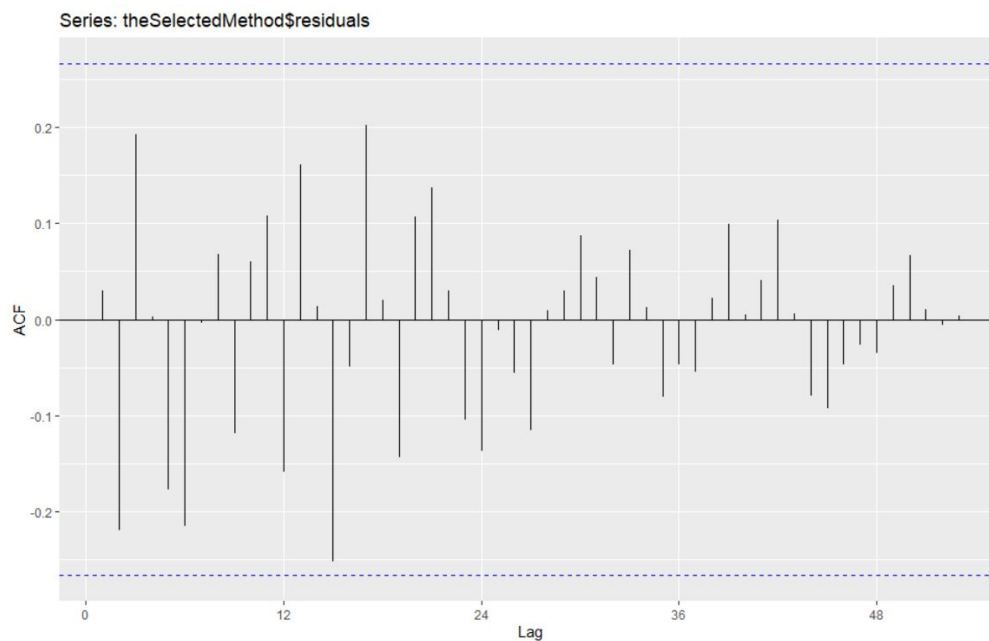


Figura 17: Correlograma de los residuales del pronóstico realizado con el método Holt-Winters aditivo

Esta vez se observa que ninguno de los datos del correlograma es significativamente distinto de cero, y no se observa patrón alguno en él, por lo que es posible que los

residuales sean aleatorios en esta ocasión. Para comprobar se realizó la prueba de Ljung-box sobre los residuales del método seleccionado, de la que se obtuvo lo siguiente: $X^2 = 0.0517$, $df = 1$, $p\text{-value} = 0.8201$. Con un valor-P mayor a 0.05, poseemos evidencia estadística para concluir que los residuales **son** aleatorios y que el pronóstico con Holt-Winters aditivo es el adecuado en este caso.

4. Conclusiones

Pronósticos DAUPSA

Al correr los pronósticos de producción de autos domésticos en Estados Unidos con toda la serie de tiempo la demanda para los siguientes 6 meses es la siguiente:

```
> theSelectedMethod
      Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
Aug 2020    199.4651 163.0166 235.9135 143.72196 255.2082
Sep 2020    198.5595 152.0375 245.0814 127.41032 269.7086
Oct 2020    197.6539 142.8789 252.4289 113.88275 281.4250
Nov 2020    196.7483 134.8086 258.6880 102.01975 291.4768
Dec 2020    195.8427 127.4837 264.2017  91.29661 300.3888
Jan 2021    194.9371 120.7104 269.1638  81.41722 308.4570
```

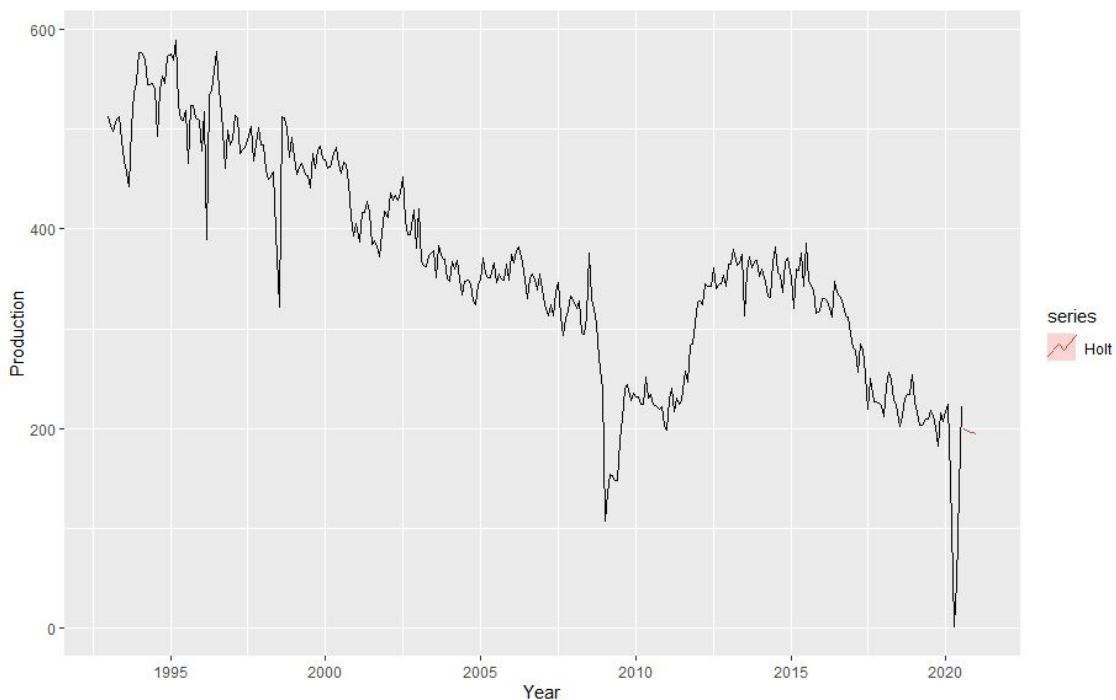


Figura 18: Pronóstico de la producción de autos posterior al COVID-19

Sin embargo, los pronósticos para los mismos 6 meses (Agosto 2020 - Enero 2021) tomando en cuenta sólo los meses previos a la pandemia del COVID-19 son:

```
> theSelectedMethod2
```

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Aug 2020	165.2644	107.70715	222.8216	77.23820	253.2905
Sep 2020	164.2067	102.02779	226.3856	69.11228	259.3011
Oct 2020	163.1490	96.66779	229.6302	61.47476	264.8232
Nov 2020	162.0913	91.56866	232.6139	54.23623	269.9464
Dec 2020	161.0336	86.68786	235.3794	47.33159	274.7357
Jan 2021	159.9759	81.99326	237.9586	40.71172	279.2401

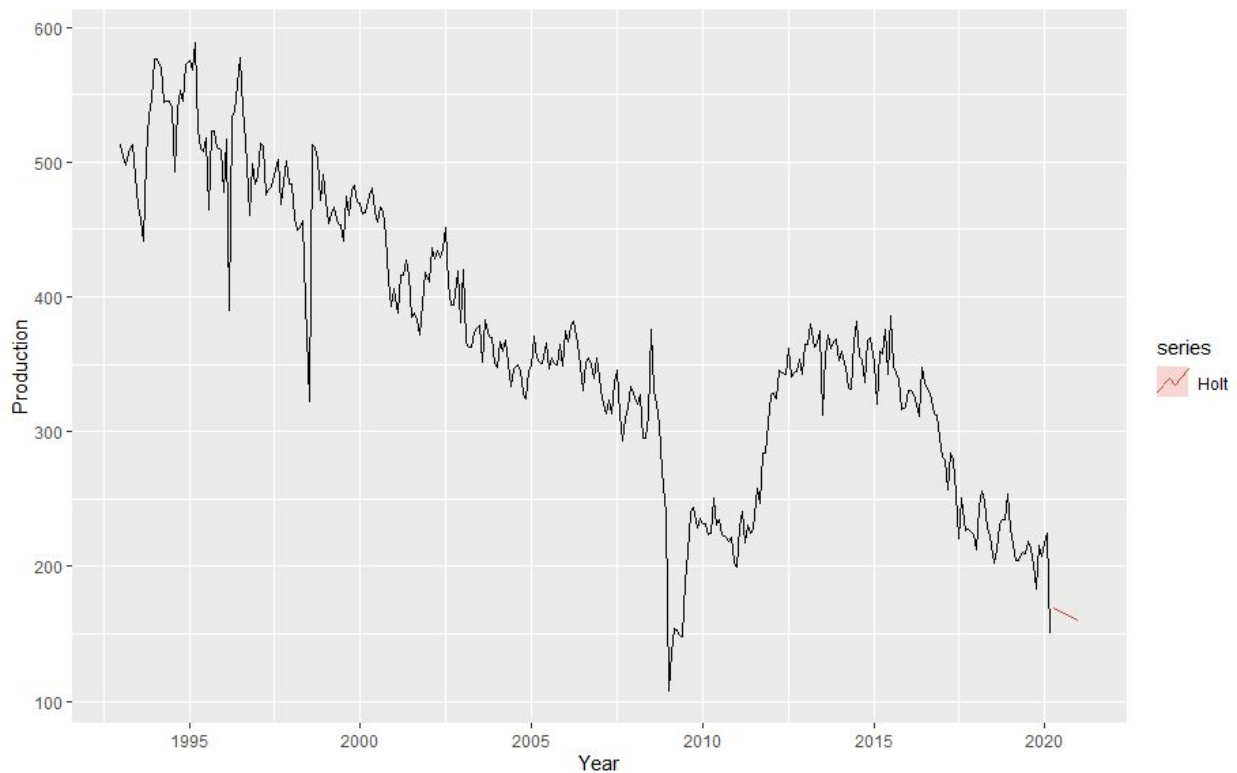


Figura 19: Pronóstico de la producción de autos previo al COVID-19

Sabemos que la producción tuvo una baja significativa a partir de marzo del 2020 debido a la pandemia, los pronósticos para abril, mayo, junio y julio de este año eran:

```
> theSelectedMethod2
```

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Apr 2020	169.4951	136.31573	202.6745	118.75163	220.2386
May 2020	168.4374	127.77260	209.1022	106.24595	230.6289
Jun 2020	167.3797	120.40591	214.3535	95.53947	239.2200
Jul 2020	166.3220	113.79007	218.8540	85.98132	246.6628

En cambio los números reales fueron mucho menores de lo esperado:

	Feb	Mar	Apr	May	Jun	Jul
2020	224.3	150.2	1.7	46.2	138.1	222.0

Comparando ambos pronósticos para los siguientes 6 meses comenzando por Agosto de 2020, nos damos cuenta de que la producción esperada es mayor cuando tomamos en cuenta toda la serie de tiempo, aunque en principio pensemos que será lo contrario. Esto puede ser debido a que el peso que se le otorga a los datos más recientes no es tan diferente del peso otorgado a los datos un poco menos cercanos al presente.

Tomando el pronóstico más cercano a la realidad, es decir el que toma en cuenta los meses de marzo a julio también, podemos pensar que aunque la industria tuvo una baja en los meses en que comenzó la pandemia para julio había recuperado sus valores previos, y por lo tanto es más preciso hacer el pronóstico posterior a la recuperación.

Ventas de autos híbridos en México

El caso de las ventas de autos híbridos nos proporcionó de un buen ejemplo para la selección de métodos de pronóstico. Al realizar las pruebas sobre el conjunto de entrenamiento y prueba respectivamente se determinó que el mejor método era el de Holt-Winters multiplicativo al tener una RMSE de 238.3946, el menor de todos los métodos probados. De la misma manera, el método Holt-Winters multiplicativo obtuvo el menor U de Theil de los métodos no naive, de 3.766436. Con esto en mente se seleccionó este método para realizar el pronóstico (Figura 14), y se realizaron las pruebas de aleatoriedad de los residuales. Sin embargo, tanto el correlograma (Figura 15) y la prueba de Ljung-box mostraban que los residuales no eran en efecto aleatorios, por lo que debía existir otro método que pudiera ajustarse mejor a los datos de las ventas de autos.

Se seleccionó como candidato el método de Holt-Winters aditivo, pues tuvo la segunda menor RMSE y la segunda mejor U de Theil, con un valores de 249.5303 y 4.148773, respectivamente. Al realizar el pronóstico y el análisis de los residuales se encontró que a pesar de ser el segundo mejor en las primeras pruebas, el método de Holt-Winters aditivo presentó residuales efectivamente aleatorios. En otras palabras, los errores del método seleccionado finalmente son ruido blanco y no es posible extraer más información de ellos.

A continuación se presentan los resultados obtenidos con el método de Holt-Winters aditivo, los cuales son considerablemente menores a los pronosticados por la variante del mismo método multiplicativa, y parecen ser más conservadores

en cuanto a la recuperación esperada de las ventas posterior a la pandemia. Por otro lado, la variante multiplicativa (como se muestra en la Figura 14) nos promete una recuperación más acelerada en donde las ventas recuperan sus niveles previos a la pandemia tan pronto como enero de 2021. Como se muestra en la Figura 20, el pronóstico de ventas con el método seleccionado finalmente (HW aditivo) no recupera los niveles previos a la pandemia.

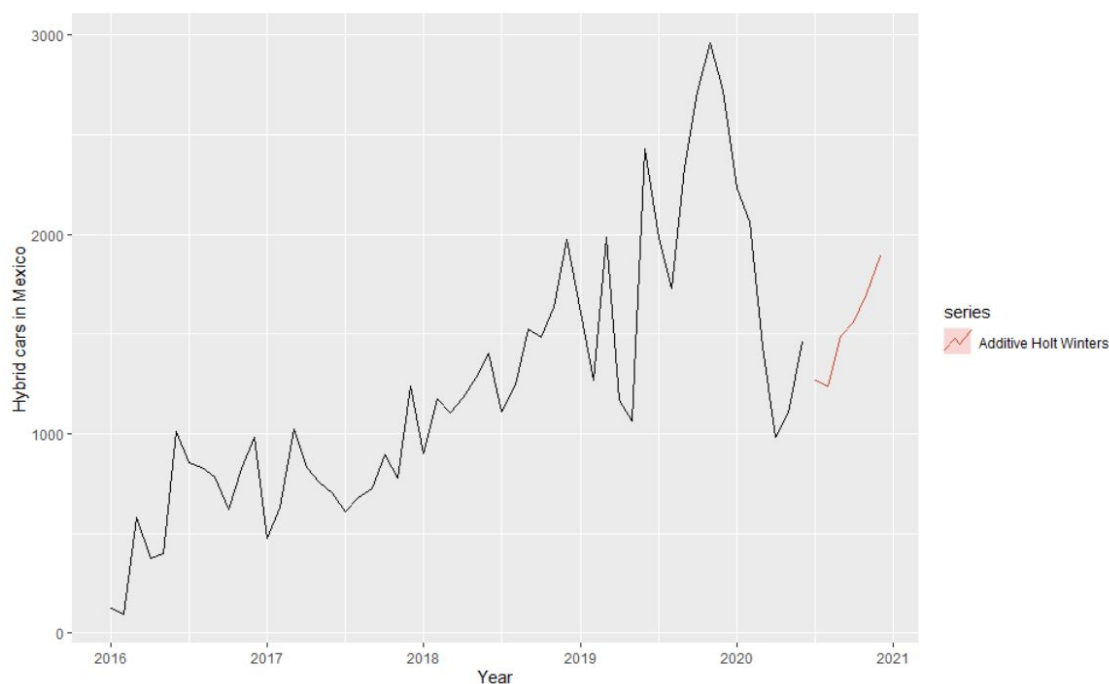


Figura 20: Pronóstico de las ventas de autos híbridos en México.

```
> theSelectedMethod
      Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
Jul 2020      1270.644  839.8571 1701.432  611.8121 1929.477
Aug 2020      1239.957  682.5011 1797.412  387.4020 2092.511
Sep 2020      1490.927  830.6501 2151.204  481.1205 2500.734
Oct 2020      1562.758  813.6212 2311.895  417.0522 2708.464
Nov 2020      1701.134  872.5957 2529.673  433.9939 2968.275
Dec 2020      1895.235  994.2480 2796.221  517.2945 3273.175
```

5. Código documentado en R

DAUPSA antes del COVID (hasta marzo 2020)

```
#Carga de paquetes
library(readxl)
library(fpp2)
library(ggplot2)

#-----Carga de datos
#Lee los datos del archivo de excel, saltando las primeras 10 filas pues
estaban sucias:
domesticAutoProduction <- read_excel("DAUPSA.xlsx", skip = 10, sheet =
"FRED Graph", col_names = TRUE)

#-----Hacemos serie de tiempo:
tsDomesticAutoProduction <- ts(domesticAutoProduction[,2], start =
c(1993,1), end = c(2020,3), frequency = 12)

#-----
#Análisis de la producción
#-----
#-----Grafica la producción de autos domésticos:
autoplot(tsDomesticAutoProduction) + geom_point(color = "red", size = 1) +
ylab("Auto production") +
  xlab("Year") + theme(panel.background = element_rect(fill = "white",
colour = "grey50"),
                      panel.grid.major.y = element_blank(),
                      panel.grid.minor.y = element_blank(),
                      panel.grid.major = element_line(colour = "grey20",
linetype = "dashed", size = 0.1)) +
  scale_x_continuous(breaks = scales::extended_breaks(20))

#----- Verificar condiciones-----
ggAcf(tsDomesticAutoProduction, lag.max = 342)
#Las autocorrelaciones son significativamente distintas de cero, por lo
que podemos hacer un análisis de la producción
Box.test(tsDomesticAutoProduction, lag = 1, type = "Ljung")
#Como p-value es menor a 0.5 sabemos que los datos en general no son
aleatorios por lo que podemos hacer pronósticos
```

```

#-----
#Encontrar el mejor método para pronosticar
#-----

trainingData <- window(tsDomesticAutoProduction, end = c(2019, 9))
testData <- window(tsDomesticAutoProduction, start = c(2019,10))

#-----Hacer pronósticos-----
averageMethod <- meanf(trainingData, h=6)
naiveMethod <- naive(trainingData, h= 6)
seasonalNaiveMethod <- snaive(trainingData, h=6)
fcSimpleExpSmoo <- ses(trainingData, initial = c("optimal"), h= 6)
holtForecast <- holt(trainingData, h=6)
hwAdditive <- hw(trainingData, seasonal = "additive", h= 6)
hwMultiplicative <- hw(trainingData, seasonal = "multiplicative", h = 6)

#Graficamos los sets de entrenamiento
autoplot(trainingData,series = "Original Demand") + ylab("Demanda") +
  autolayer(fitted(averageMethod),series = "Average Method") +
  autolayer(fitted(naiveMethod),series = "Naive Method") +
  autolayer(fitted(seasonalNaiveMethod),series = "Seasonal Naive Method") +
  autolayer(fitted(fcSimpleExpSmoo),series = "Exp. Smoothing") +
  autolayer(fitted(holtForecast),series = "Holt") +
  autolayer(fitted(hwAdditive),series = "HW Add") +
  autolayer(fitted(hwMultiplicative),series = "HW Mult")

#-----Checar la exactitud de los métodos
accuracy(averageMethod)
accuracy(naiveMethod)
accuracy(seasonalNaiveMethod)
accuracy(fcSimpleExpSmoo)
accuracy(holtForecast)
accuracy(hwAdditive)
accuracy(hwMultiplicative)
#El método con menor RMSE es el de Holt

#-----Checar exactitud con respecto a los datos de prueba

accuracy(averageMethod, testData)
accuracy(naiveMethod, testData)
accuracy(seasonalNaiveMethod, testData)
accuracy(fcSimpleExpSmoo, testData)

```

```

accuracy(holtForecast, testData)
accuracy(hwAdditive, testData)
accuracy(hwMultiplicative, testData)
#El método con mejor U de theil es el de Holt Winters Multiplicativo, sin
embargo la de Holt es cercana

#-----
#Pronosticamos con el método seleccionado
#-----

theSelectedMethod <- holt(tsDomesticAutoProduction, h = 6)
autoplot(tsDomesticAutoProduction) + xlab("Year") + ylab("Production") +
  autolayer(theSelectedMethod, PI=FALSE, series = "Holt")

#---Análisis de residuales

ggAcf(theSelectedMethod$residuals, lag.max = 342)
#los residuales parecen ser aleatorios pues sus correlaciones no son
significativamente distintas de cero.
Box.test(theSelectedMethod$residuals, lag = 1, type = "Ljung")
#el p-value es de 0.7932 por lo que podemos inferir que los errores son
aleatorios.

#-----
#---Pronosticar los siguientes 12 meses (hasta Enero 2021)
#-----

theSelectedMethod2 <- holt(tsDomesticAutoProduction, h = 12)
autoplot(tsDomesticAutoProduction) + xlab("Year") + ylab("Production") +
  autolayer(theSelectedMethod2, PI=FALSE, series = "Holt")

#---Análisis de residuales

ggAcf(theSelectedMethod2$residuals, lag.max = 342)
#los residuales parecen ser aleatorios pues sus correlaciones no son
significativamente distintas de cero.
Box.test(theSelectedMethod2$residuals, lag = 1, type = "Ljung")
#el p-value es de 0.6282 por lo que podemos inferir que los errores son
aleatorios.

```

DAUPSA incluyendo el COVID-19

```
#Carga de paquetes
library(readxl)
library(fpp2)
library(ggplot2)

#-----Carga de datos
#Lee los datos del archivo de excel, saltando las primeras 10 filas pues
estaban sucias:
domesticAutoProduction <- read_excel("DAUPSA.xlsx", skip = 10, sheet =
"FRED Graph", col_names = TRUE)

#-----Hacemos serie de tiempo:
tsDomesticAutoProduction <- ts(domesticAutoProduction[,2], start =
c(1993,1), frequency = 12)

#-----
#Análisis de la producción
#-----
#-----Grafica la producción de autos domésticos:
autoplot(tsDomesticAutoProduction) + geom_point(color = "red", size = 1) +
ylab("Auto production") +
  xlab("Year") + theme(panel.background = element_rect(fill = "white",
colour = "grey50"),
                      panel.grid.major.y = element_blank(),
                      panel.grid.minor.y = element_blank(),
                      panel.grid.major = element_line(colour = "grey20",
linetype = "dashed", size = 0.1)) +
  scale_x_continuous(breaks = scales::extended_breaks(20))

#----- Verificar condiciones-----
ggAcf(tsDomesticAutoProduction, lag.max = 342)
#Las autocorrelaciones son significativamente distintas de cero, por lo
que podemos hacer un análisis de la producción
Box.test(tsDomesticAutoProduction, lag = 1, type = "Ljung")
#Como p-value es menor a 0.5 sabemos que los datos en general no son
aleatorios por lo que podemos hacer pronósticos

#-----
#Encontrar el mejor método para pronosticar
```

```

#-----

trainingData <- window(tsDomesticAutoProduction, end = c(2020, 1))
testData <- window(tsDomesticAutoProduction, start = c(2020,2))

#-----Hacer pronósticos-----
averageMethod <- meanf(trainingData, h=6)
naiveMethod <- naive(trainingData, h= 6)
seasonalNaiveMethod <- snaive(trainingData, h=6)
fcSimpleExpSmoo <- ses(trainingData, initial = c("optimal"), h= 6)
holtForecast <- holt(trainingData, h=6)
hwAditive <- hw(trainingData, seasonal = "additive", h= 6)
hwMultiplicative <- hw(trainingData, seasonal = "multiplicative", h = 6)

#-----Checar la exactitud de los métodos
accuracy(averageMethod)
accuracy(naiveMethod)
accuracy(seasonalNaiveMethod)
accuracy(fcSimpleExpSmoo)
accuracy(holtForecast)
accuracy(hwAditive)
accuracy(hwMultiplicative)
#El método con menor RMSE es el de Holt

#-----Checar exactitud con respecto a los datos de prueba

accuracy(averageMethod, testData)
accuracy(naiveMethod, testData)
accuracy(seasonalNaiveMethod, testData)
accuracy(fcSimpleExpSmoo, testData)
accuracy(holtForecast, testData)
accuracy(hwAditive, testData)
accuracy(hwMultiplicative, testData)

#Graficamos los sets de entrenamiento
autoplot(trainingData,series = "Original Demand") + ylab("Demanda") +
  autolayer(fitted(averageMethod),series = "Average Method") +
  autolayer(fitted(naiveMethod),series = "Naive Method") +
  autolayer(fitted(seasonalNaiveMethod),series = "Seasonal Naive Method") +
  autolayer(fitted(fcSimpleExpSmoo),series = "Exp. Smoothing") +
  autolayer(fitted(holtForecast),series = "Holt") +
  autolayer(fitted(hwAditive),series = "HW Add") +

```

```

autolayer(fitted(hwMultiplicative),series = "HW Mult")

#El método con mejor U de theil es el de Holt Winters Multiplicativo, sin
embargo la de Holt es cercana

#-----
#Pronosticamos con el método seleccionado
#-----

theSelectedMethod <- holt(tsDomesticAutoProduction, h = 6)
autoplot(tsDomesticAutoProduction) + xlab("Year") + ylab("Production") +
  autolayer(theSelectedMethod, PI=FALSE, series = "Holt")

#----Análisis de residuales

ggAcf(theSelectedMethod$residuals, lag.max = 342)
#los residuales parecen ser aleatorios pues sus correlaciones no son
significativamente distintas de cero.
Box.test(theSelectedMethod$residuals, lag = 1, type = "Ljung")
#el p-value es de 0.6282 por lo que podemos inferir que los errores son
aleatorios.

```

Ventas de autos híbridos en México (toda la serie de tiempo)

```

#Carga de paquetes
library(readxl)
library(fpp2)
library(ggplot2)

#-----
#-----Carga de datos
#-----

#Lee los datos del archivo de excel, saltando las primeras 3 filas pues
estaban sucias:
hybridCarsMX <- read_excel("RAIAVL_11.xlsx", skip = 3, sheet = "RAIAVL_11",
col_names = FALSE)

#Hacemos modificaciones a los datos para poder tratarlos adecuadamente
#Cambiamos los valores '-' por 0 para poder hacer operaciones

for (i in 1:1782) {

```



```

    if(hybridCarsMX[i,4] == '-'){
      hybridCarsMX[i,4] <- '0'
    }
  }

#Modificamos los datos para tomar el agregado cada mes
for (i in 0:53) { #1749=53*33
  suma<-0
  aux<-i*33
  for (j in 1:33) {
    #aux<-aux+j
    suma<-suma+as.double(hybridCarsMX[aux+j,4])
  }
  if (i==0){
    hybridCarsMXData<-c(suma)
  }else{
    hybridCarsMXData <- c(hybridCarsMXData, suma)
  }
}

#----Hacemos serie de tiempo
tsHybridCarsMX <- ts(hybridCarsMXData, start = c(2016,1), frequency = 12 )

#-----
#Análisis de la demanda de autos híbridos en México
#-----

#-----Gráfica la demanda de autos híbridos:
autoplot(tsHybridCarsMX) + geom_point(color = "red", size =1 ) +
  ylab("Hybrid cars in México") +
  xlab("Time") + theme(panel.background = element_rect(fill = "white",
  colour = "grey50"),
                      panel.grid.major.y = element_blank(),
                      panel.grid.minor.y = element_blank(),
                      panel.grid.major = element_line(colour = "grey20",
linetype = "dashed", size = 0.1)) +
  scale_x_continuous(breaks = scales::extended_breaks(5))

#----- Verificar condiciones-----
ggAcf(tsHybridCarsMX, lag.max = 54)
#Las autocorrelaciones son significativamente distintas de cero, por lo que
podemos hacer un análisis de la producción

```

```

Box.test(tsHybridCarsMX, lag = 1, type = "Ljung")
#Como p-value = 7.507e-10 es menor a 0.5 sabemos que los datos en general
no son aleatorios por lo que podemos hacer pronósticos

#-----
#Encontrar el mejor método para pronosticar
#-----

trainingData <- window(tsHybridCarsMX, end = c(2019, 12))
testData <- window(tsHybridCarsMX, start = c(2020,1))

#-----Hacer pronósticos-----
averageMethod <- meanf(trainingData, h=6)
naiveMethod <- naive(trainingData, h= 6)
seasonalNaiveMethod <- snaive(trainingData, h=6)
fcSimpleExpSmoo <- ses(trainingData, initial = c("optimal"), h= 6)
holtForecast <- holt(trainingData, h=6)
hwAdditive <- hw(trainingData, seasonal = "additive", h= 6)
hwMultiplicative <- hw(trainingData, seasonal = "multiplicative", h = 6)

#-----Checar la exactitud de los métodos
accuracy(averageMethod)
accuracy(naiveMethod)
accuracy(seasonalNaiveMethod)
accuracy(fcSimpleExpSmoo)
accuracy(holtForecast)
accuracy(hwAdditive)
accuracy(hwMultiplicative)
#El método con menor RMSE es el de Holt Winters Multiplicativo

#-----Checar exactitud con respecto a los datos de prueba
accuracy(averageMethod, testData)
accuracy(naiveMethod, testData)
accuracy(seasonalNaiveMethod, testData)
accuracy(fcSimpleExpSmoo, testData)
accuracy(holtForecast, testData)
accuracy(hwAdditive, testData)
accuracy(hwMultiplicative, testData)
#El método con mejor U de theil es el método de Holt Winters Multiplicativo

#Graficamos los sets de entrenamiento
autoplot(trainingData, series = "Original Demand") + ylab("Demanda") +

```

```

autolayer(fitted(averageMethod),series = "Average Method") +
autolayer(fitted(naiveMethod),series = "Naive Method") +
autolayer(fitted(seasonalNaiveMethod),series = "Seasonal Naive Method") +
autolayer(fitted(fcSimpleExpSmoo),series = "Exp. Smoothing") +
autolayer(fitted(holtForecast),series = "Holt") +
autolayer(fitted(hwAditive),series = "HW Add") +
autolayer(fitted(hwMultiplicative),series = "HW Mult")

#-----
#Pronosticamos con el método seleccionado (Holt Winters Multiplicativo)
#-----

theSelectedMethod <- hw(tsHybridCarsMX, seasonal = "multiplicative", h = 6)
autoplot(tsHybridCarsMX) + xlab("Year") + ylab("Hybrid cars in Mexico") +
  autolayer(theSelectedMethod, PI=FALSE, series = "Multiplicative Holt
Winters")

#----Analizamos los residuales
ggAcf(theSelectedMethod$residuals, lag.max = 53)
#no se puede inferir la aleatoriedad pues el primer elemento es
significativamente distinto de cero
Box.test(theSelectedMethod$residuals, lag = 1, type = "Ljung")
#el p-value es de 0.00111 por lo que no podemos inferir que los residuales
sean aleatorios

#debido a que no podemos inferir residuales aleatorios se prueba con el
segundo mejor metodo de las pruebas
#es decir, se elige el metodo de HW aditivo, el que tiene la segunda mejor
RMSE y U de Theil

#-----
#Pronosticamos con el método seleccionado (Holt Winters Aditivo)
#-----

theSelectedMethod <- hw(tsHybridCarsMX, seasonal = "additive", h = 6)
autoplot(tsHybridCarsMX) + xlab("Year") + ylab("Hybrid cars in Mexico") +
  autolayer(theSelectedMethod, PI=FALSE, series = "Additive Holt Winters")

#----Analizamos los residuales
ggAcf(theSelectedMethod$residuals, lag.max = 53)
#los residuales parecen ser aleatorios pues sus correlaciones no son
significativamente distintas de cero.

```

```
Box.test(theSelectedMethod$residuals, lag = 1, type = "Ljung")  
#el p-value es de 0.8201 por lo que podemos inferir que los errores son  
aleatorios.
```

Fuentes de consulta

- Piña, F. (2019). El nuevo rumbo de la industria automotriz en México. Retrieved 20 September 2020, from <https://www.milenio.com/especiales/el-nuevo-rumbo-de-la-industria-automotriz-en-mexico>
- (Junio 2012). Doing Business in Mexico Automotive Industry. Septiembre 20, 2020, de PwC Mexico Sitio web: <https://biblioteca.itam.mx>
- Berry, S., Grilli V. & López-de-Silanes, F.. (Agosto 1992). The automobile industry and the Mexico - US free trade agreement. NBER Working paper series, 75. 2020, septiembre 20, De Biblioteca ITAM Base de datos
- Buendia I.. (1970). Introducción. En Diversos Aspectos Administrativos de una Empresa Industrial del Ramo Automotriz(241). México, D.F.