

CP005 - Mapfre Seguros y Modelos ETS y ARIMA

Mayra Goicochea Neyra

13/11/2019

1. Objetivo

Estimar el número total de primas de la empresa Mapfre para los trimestres del 2018 y 2019. Se tiene la información histórica de los trimestres del 2007 al 2017.

1.1 Dataset Overview

El dataset se encuentra en un archivo csv. Incluye 40 casos con 2 variables

- **Primas Vida** : El número de primas vida.
- **Primas No Vida** : El número de otros seguros.

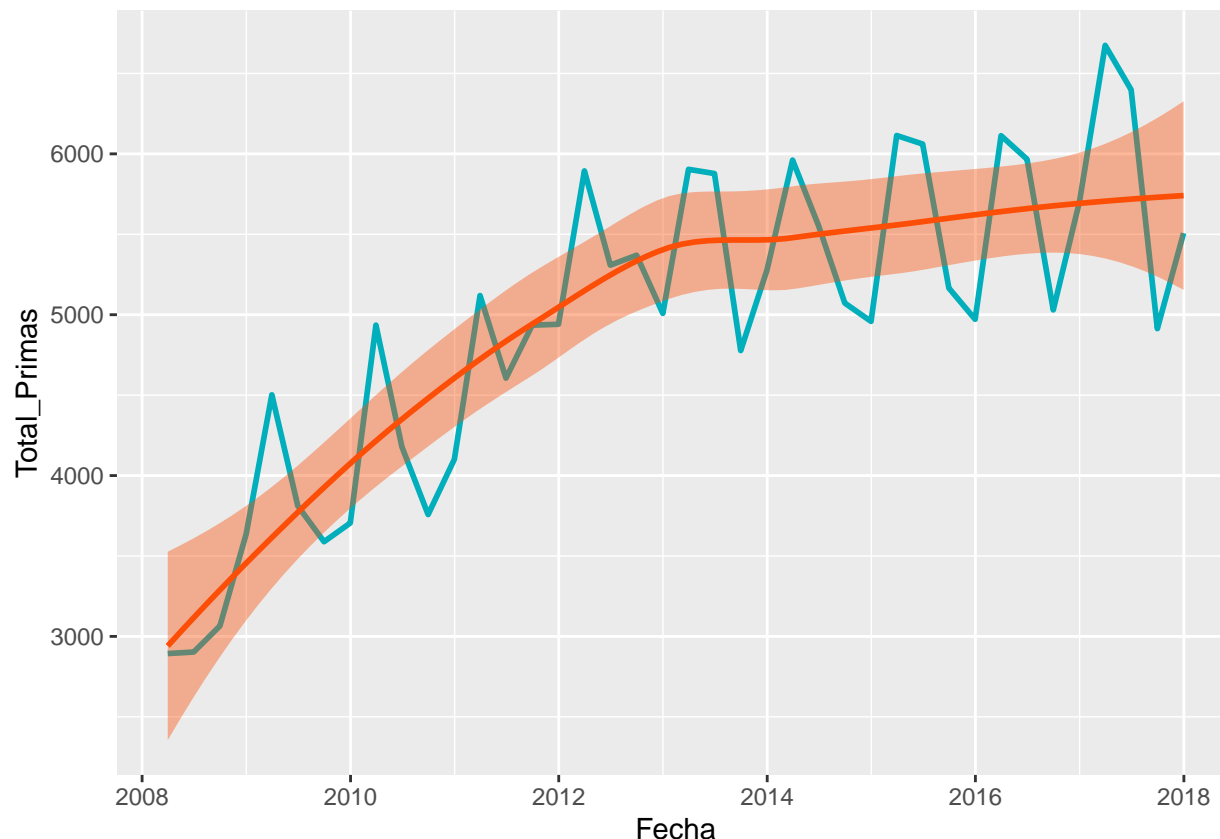
2. Trabajo Realizado

2.1. Extracción, Transformación y Carga de Datos

Se calcula el número total de primas a partir de la suma del número de primas vida y primas no vida. Ésta será la variable explicada del modelo. También se creará el objeto series de tiempo que se utilizará para el desarrollo del modelo.

2.2. Análisis Exploratorio de los datos

El primer grafico importante es el que muestra el comportamiento de las primas en los últimos años, donde los años están el eje “x” y el eje “y” es el número de primas.



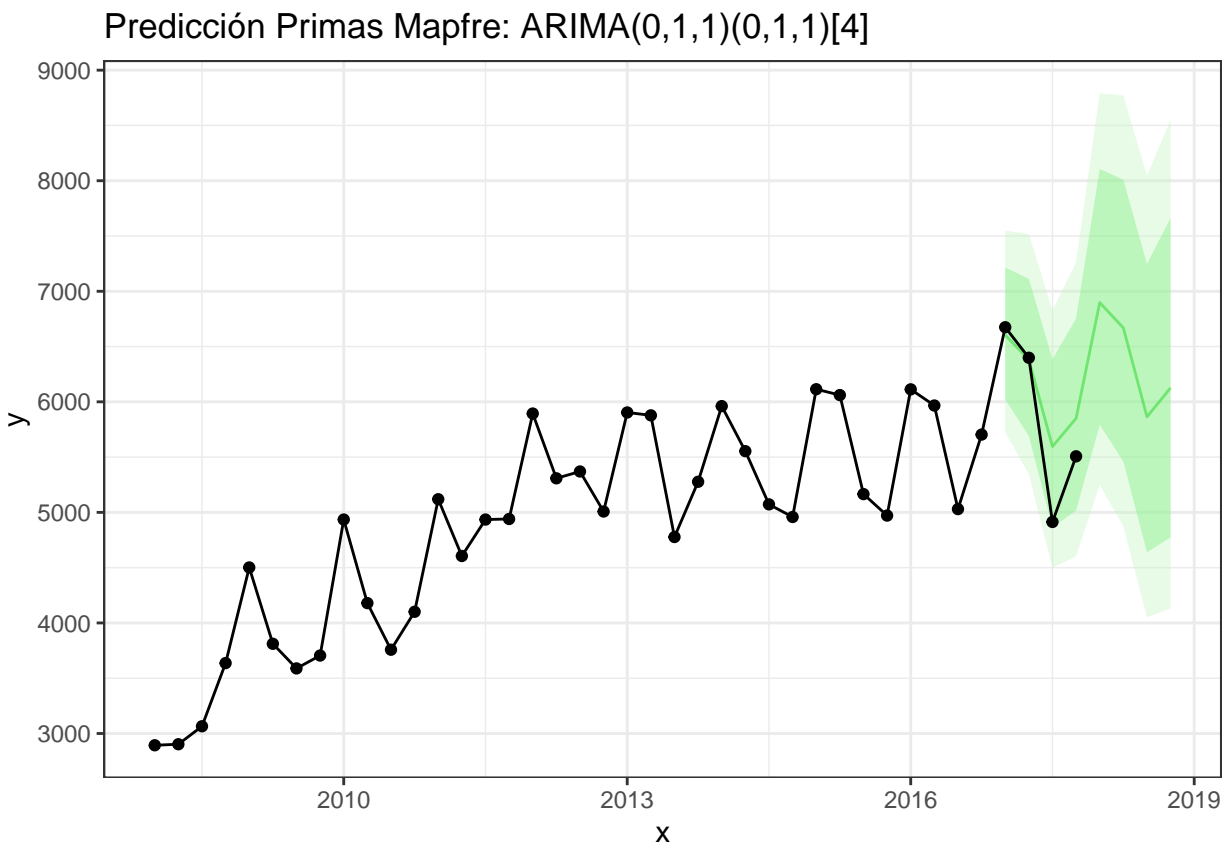
```
## function (base_size = 11, base_family = "", base_line_size = base_size/22,
##   base_rect_size = base_size/22)
## {
##   theme_grey(base_size = base_size, base_family = base_family,
##     base_line_size = base_line_size, base_rect_size = base_rect_size) %+replace%
##     theme(panel.background = element_rect(fill = "white",
##       colour = NA), panel.border = element_rect(fill = NA,
##       colour = "grey20"), panel.grid = element_line(colour = "grey92"),
##     panel.grid.minor = element_line(size = rel(0.5)),
##     strip.background = element_rect(fill = "grey85",
##       colour = "grey20"), legend.key = element_rect(fill = "white",
##       colour = NA), complete = TRUE)
## }
## <bytecode: 0x00000000188c7848>
## <environment: namespace:ggplot2>
```

Entre las principales características que se encuentran: es que tiene tendencia creciente amortiguada a partir del 2013. No es estacional y hay variaciones entre los años. Se agregan más gráficos en el anexo sobre la estacionalidad y la descomposición de la curva.

2.3. Selección del Modelo Predictivo

Cuando se trata de información relacionada al tiempo, como es el número de primas, se pueden utilizar los modelos ETS y ARIMA. Primero se dividió la muestra en dos grupos: * **Train** : La información del 2007 al 2016 se utilizó como entrenamiento de los modelos ETS y ARIMA. * **Test** : La información del 2017 fue utilizada para probar los modelos ETS y ARIMA, y escoger el modelo más eficiente para el caso de MAPFRE.

Se generó el modelo ETS (A,A,A) y el modelo ARIMA(0,1,1)(0,1,1)[4]. El modelo ARIMA resultó ser el más adecuado para esta información por tener un menor margen de error (con un porcentaje de error medio absoluto de 5.64% sobre la muestra). Además las estimaciones de la muestra test encaja con los datos reales en los primeros dos trimestres del 2017:



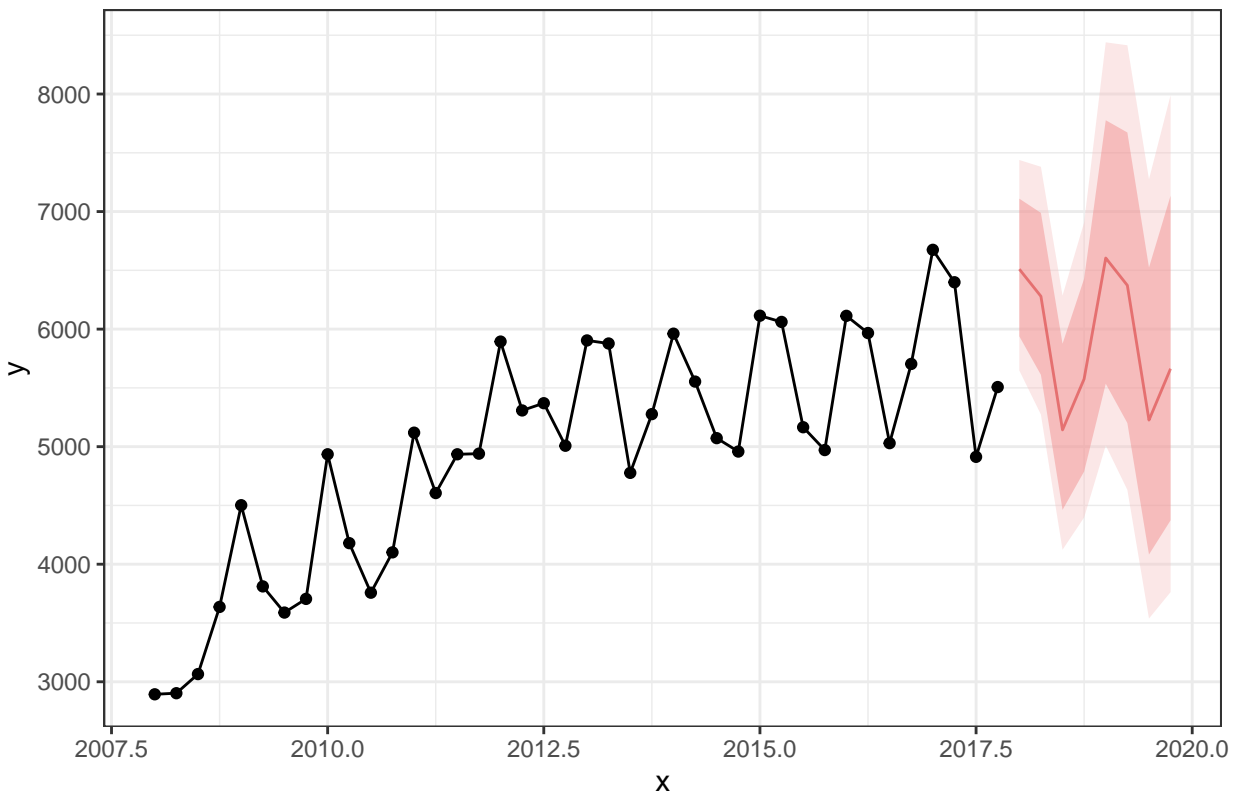
En el anexo, se encuentran los datos adicionales del modelo ARIMA como del modelo ETS.

2.4. Predicción del 2018 y 2019

Con el modelo ARIMA, se procede a estimar los valores que MAPFRE obtendrá para cada trimestre de los años 2018 y 2019. El grafico obtenido de las estimaciones es el siguiente:

```
## Warning in geom_forecast(pred.2018, alpha = 0.4, col = "red"): Use
## autolayer instead of geom_forecast to add a forecast layer to your ggplot
## object.
```

Predicción Primas Mapfre 2018: Modelo ETS(A,A,A)



3. Conclusión

El número de primas es una variable que se ha comportado de forma creciente en los últimos 12 años. Tiene mucha varianza entre los trimestres. El modelo ARIMA se ajusta eficientemente a este comportamiento, lo que permite tener una estimación de los próximos trimestres. Cabe observar también que la muestra es pequeña, lo recomendable es una muestra de 60 casos, dado que permite realizar métodos de validación como Cross Validation. Es preferible que la información hubiera sido mensual para analizar los estados de estacionalidad.

Anexos

Gráfico de Estacionalidad:

```
ggseasonplot(ts_data, year.labels = TRUE, year.labels.left = TRUE) +  
  ylab("Numero de Primas") +  
  ggtitle("Seasonal plot: Primas Vida Mapfre")
```

Seasonal plot: Primas Vida Mapfre

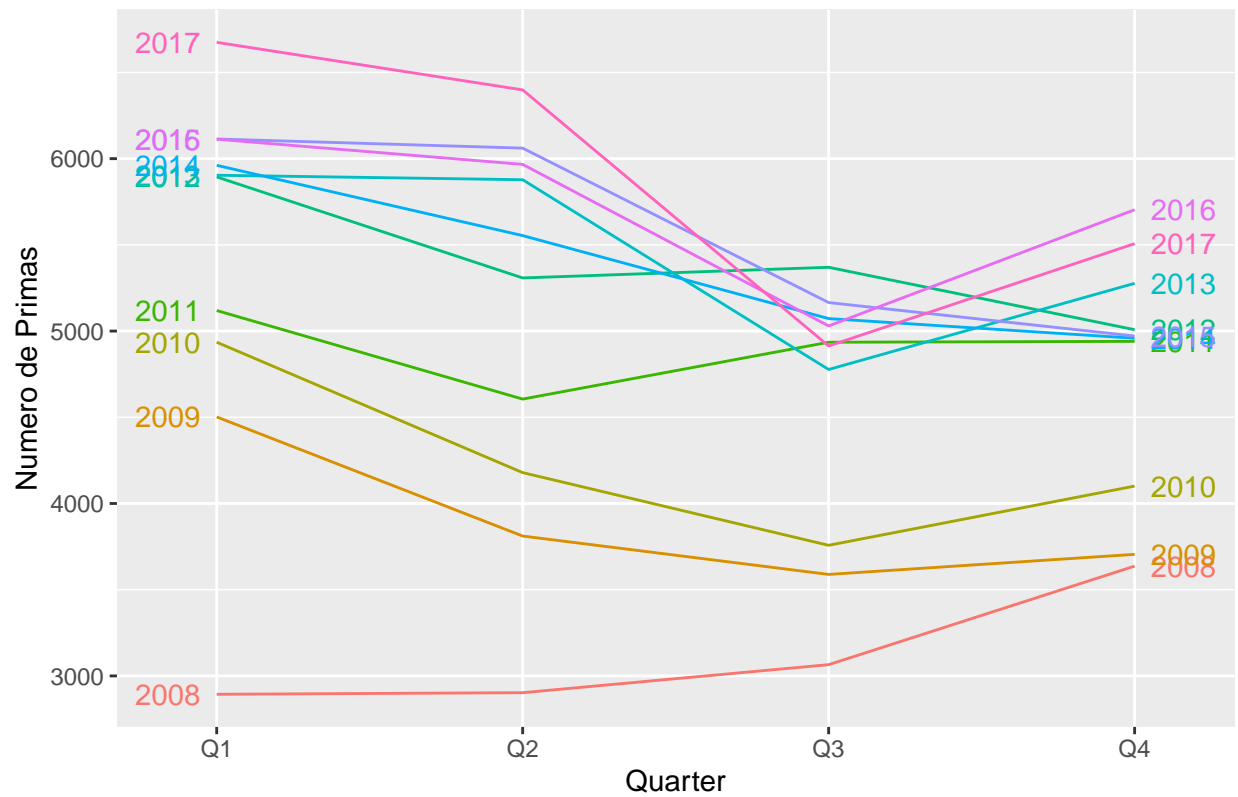
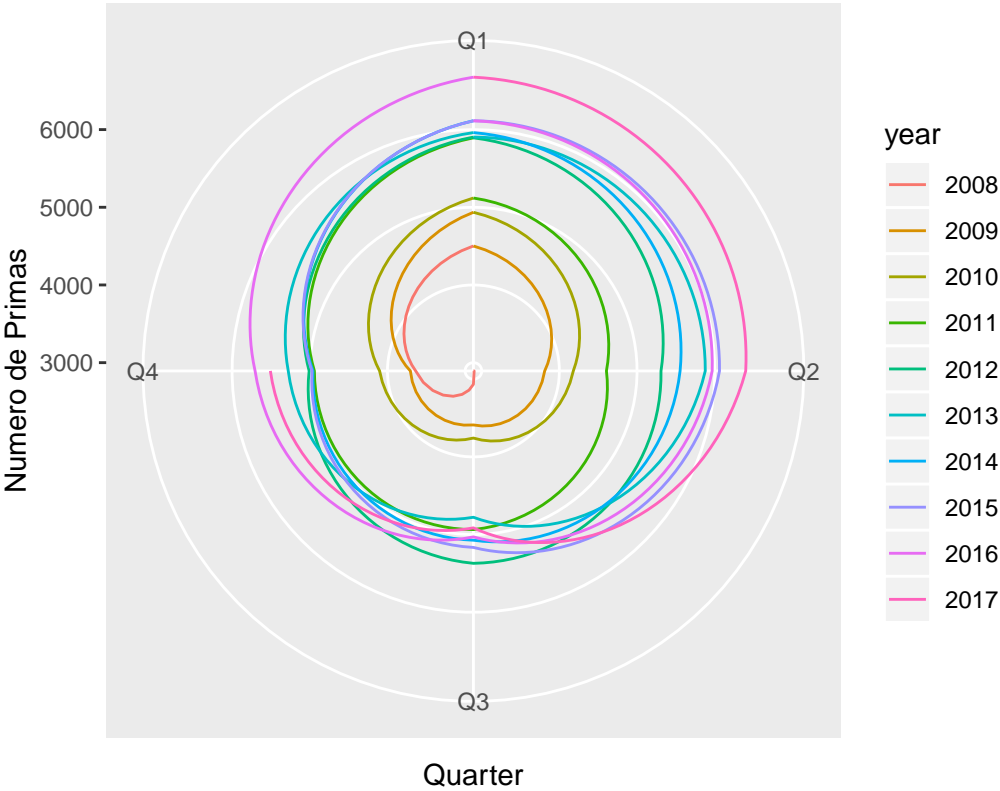


Gráfico Polar de Estacionalidad:

```
ggseasonplot(ts_data, polar = TRUE) +
  ylab("Numero de Primas") +
  ggtitle("Polar seasonal plot: Primas Vida Mapfre")
```

Polar seasonal plot: Primas Vida Mapfre



Descomposición de la Variable Primas:

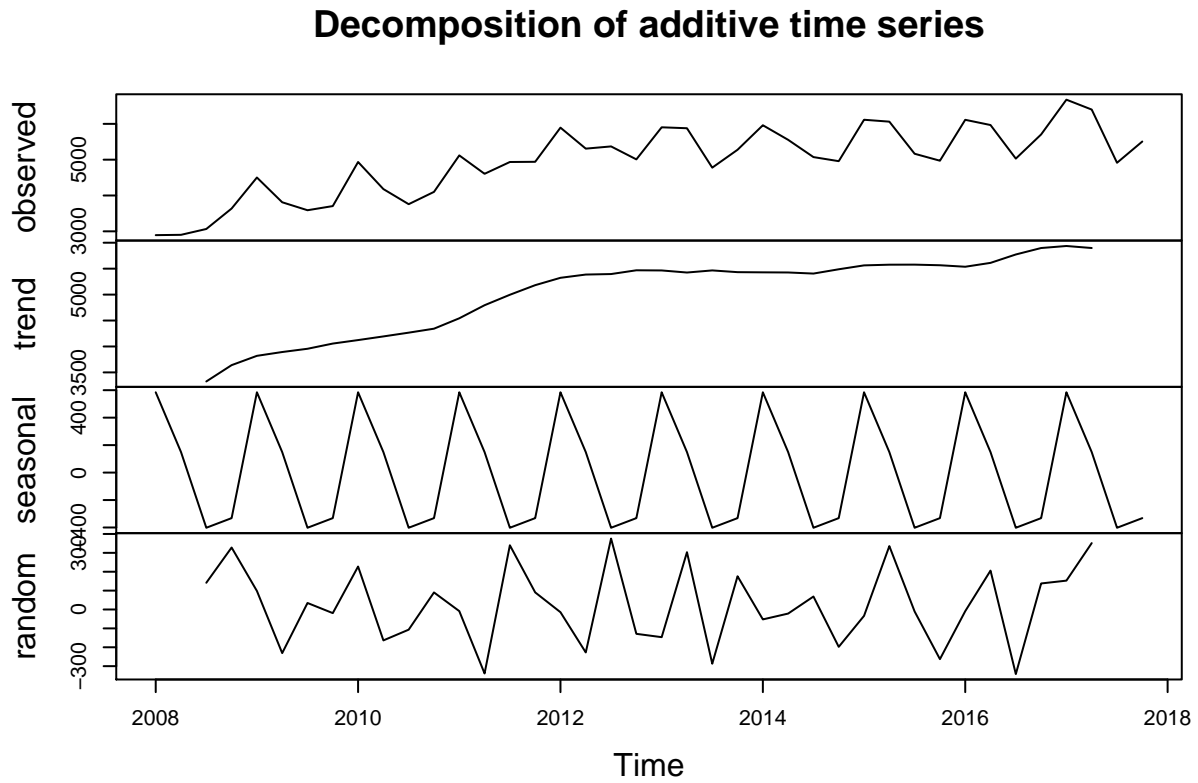
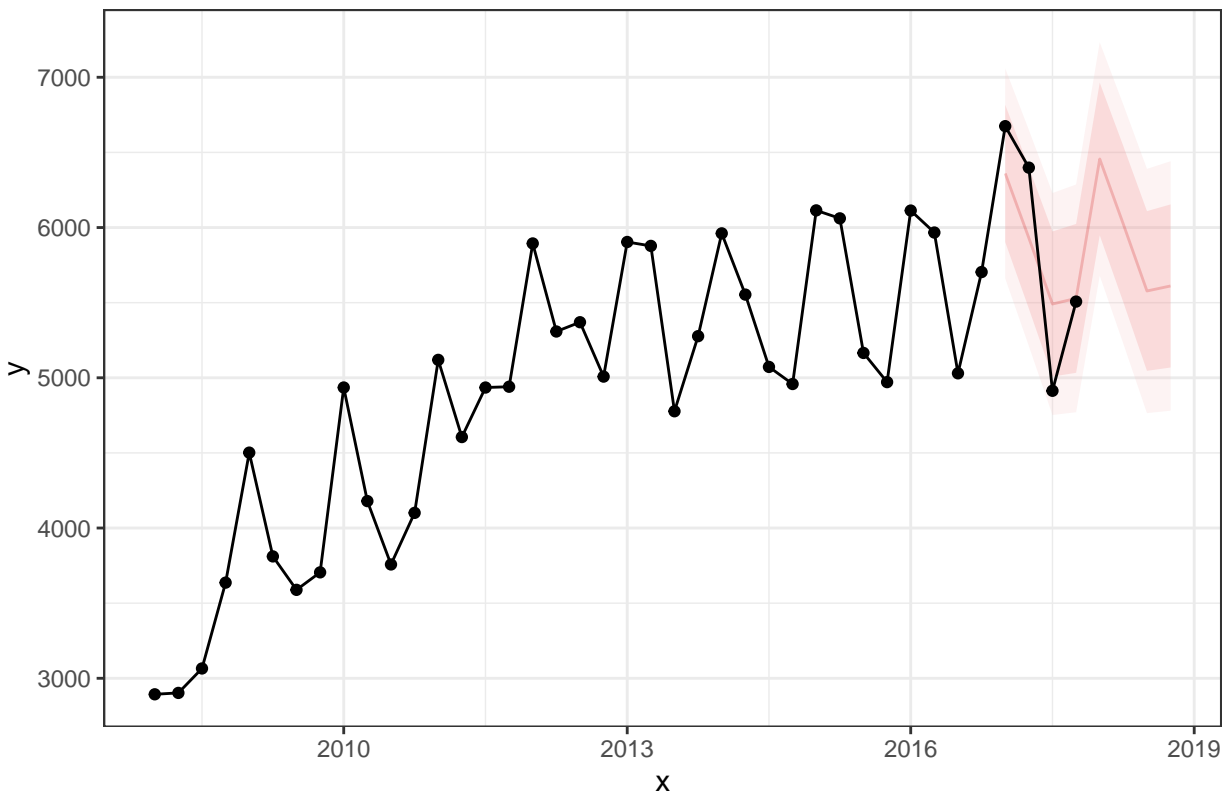


Gráfico del Modelo ETS (A,A,A)

```
plot + pets +  
  ggtitle("Predicción Primas Mapfre: Modelo ETS(A,A,A)") +  
  geom_point(aes(x = time,y = value)) +  
  geom_line(aes(x = time,y = value)) +  
  theme_bw()
```

Predicción Primas Mapfre: Modelo ETS(A,A,A)



Resumen del modelo ETS (A,A,A)

```
summary(fets)
```

```
##
## Forecast method: ETS(A,Ad,A)
##
## Model Information:
## ETS(A,Ad,A)
##
## Call:
## ets(y = oPrimas)
##
## Smoothing parameters:
##   alpha = 0.2424
##   beta  = 1e-04
##   gamma = 1e-04
##   phi   = 0.9505
##
## Initial states:
##   l = 2931.1472
##   b = 178.7185
##   s = -335.3013 -349.5291 114.6065 570.2239
##
## sigma: 356.3793
##
```



```
##      AIC      AICc      BIC
## 561.7218 570.5218 577.5570
##
## Error measures:
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -5.403611 308.6335 249.7589 -0.4948873 5.595178 0.6033052
##              ACF1
## Training set 0.1276507
##
## Forecasts:
##      Point Forecast      Lo 80      Hi 80      Lo 95      Hi 95
## 2017 Q1      6360.067 5903.349 6816.786 5661.577 7058.558
## 2017 Q2      5930.371 5460.419 6400.324 5211.641 6649.102
## 2017 Q3      5490.845 5008.011 5973.679 4752.414 6229.275
## 2017 Q4      5528.540 5033.151 6023.929 4770.908 6286.172
## 2018 Q1      6456.184 5948.533 6963.836 5679.798 7232.570
## 2018 Q2      6021.727 5502.104 6541.349 5227.033 6816.421
## 2018 Q3      5577.675 5046.343 6109.006 4765.074 6390.276
## 2018 Q4      5611.068 5068.274 6153.862 4780.937 6441.200
```

Resumen del modelo ARIMA(0,1,1)(0,1,1)[4]

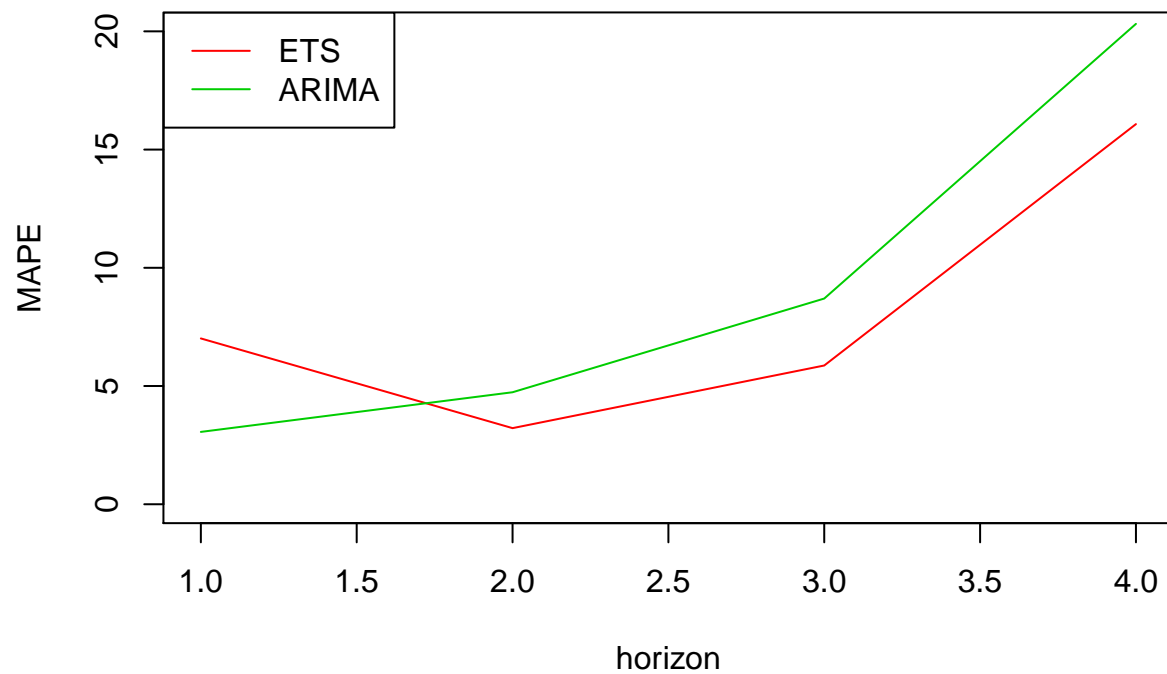
```
summary(farima)
```

```
##
## Forecast method: ARIMA(0,1,1)(0,1,1)[4]
##
## Model Information:
## Series: oPrimas
## ARIMA(0,1,1)(0,1,1)[4]
## Box Cox transformation: lambda= 0.4588303
##
## Coefficients:
##      ma1      sma1
##      -0.3121 -0.6042
## s.e.   0.1935   0.1815
##
## sigma^2 estimated as 15.87: log likelihood=-86.77
## AIC=179.54 AICc=180.43 BIC=183.84
##
## Error measures:
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -109.4965 353.9768 270.3513 -2.730747 5.647908 0.653047
##              ACF1
## Training set -0.1828349
##
## Forecasts:
##      Point Forecast      Lo 80      Hi 80      Lo 95      Hi 95
## 2017 Q1      6604.021 6022.471 7214.687 5726.328 7549.812
## 2017 Q2      6380.374 5691.575 7111.944 5344.123 7516.658
## 2017 Q3      5595.920 4863.469 6384.270 4498.143 6824.435
## 2017 Q4      5851.347 5017.660 6754.779 4604.267 7261.544
## 2018 Q1      6897.593 5792.877 8107.326 5250.078 8790.715
## 2018 Q2      6668.577 5460.605 8008.134 4873.680 8771.206
```

```
## 2018 Q3      5864.570 4638.970 7246.943 4052.624 8043.139
## 2018 Q4      6126.495 4774.497 7662.500 4132.016 8551.279
```

Cross Validation de los modelos ETS y ARIMA

```
k <- 10 # 60 is minimum data length for fitting a model, but dataset length is 40
n <- length(ts_data)
mape1 <- mape2 <- matrix(NA,4,4)
st <- tsp(ts_data)[1] + (k - 1)/4
for(i in 1:4)
{
  xshort <- window(ts_data, end = st + (i-1))
  xnext <- window(ts_data, start = st + (i-1) + 1/4, end=st + i)
  #Automatic ETS Model
  fit1 <- ets(xshort, model = "AAA")
  fcast1 <- forecast(fit1, h = 4)
  #ARIMA
  fit2 <- Arima(xshort, order = c(0,1,1), seasonal = list(order = c(0,1,1), period = 4), lambda = 0.458)
  fcast2 <- forecast(fit2, h = 4)
  #MAPE
  mape1[i,] <- abs((xnext- fcast1[['mean']])/xnext)*100
  mape2[i,] <- abs((xnext- fcast2[['mean']])/xnext)*100
}
plot(1:4, rowMeans(mape1), type = "l", col = 2, xlab = "horizon", ylab = "MAPE",
     ylim = c(0,20))
lines(1:4, rowMeans(mape2), type = "l",col = 3)
legend("topleft",legend = c("ETS","ARIMA"),col = 2:3,lty = 1)
```



Bibliografia

- Forecasting: Principles and Practice, Rob J Hyndman and George Athanasopoulos.
- Time series cross validation, en sitio web: <https://robjhyndman.com/hyndsight/tscvexample/>