1 Validation and Forecasting

In this practice, the validation of models adjusted to series will be worked on, as well as the selection of the best model among different alternatives. Predictions will be calculated using the selected model. As has been done in the previous practices, each group of students must perform the script of the practice below, performing the necessary calculations and graphics in the R environment.

2 Questions

In this quiz we continue to work on some of the time series from the previous session that you can find in the series repository in the subject area at ATENEA. We are adding two new series:
`ATUR`: https://expinterweb.mites.gob.es/series/) Individuals registered as unemployed at INEM/SEPE offices
We propose, for this series the following models for the logarithms:

a) $ARIMA(1,1,1)(0,1,1)_{12}$ without constant

b) $ARIMA(8,1,0)(0,1,1)_{12}$ Without constant (considering all parameters significant)

For each model, you must obtain validation charts for residual analysis and interpret the results:

1. Plot the residuals and the square root of the absolute value of the residuals with a smooth line. Can we consider the variance to be constant?
2. Make the normality plot and the histogram with the normal curve superimposed. Apply the Shapiro-Wilks test to the residuals. Can we consider the residuals to come from a normal distribution?
3. Do the ACF and PACF of the residuals and plot the p-values from the Ljung-Box test. Can we consider the residuals to be independent?
4. For the characteristic polynomials of part AR and MA, calculate the modulus of their roots. Is the estimated model causal? Is it invertible?
5. Calculate measures of data adequacy (AIC and BIC)
6. Fit the model to all data and to data without the last 12 observations (TRAIN set). Can we consider the model stable?
7. For the fitted model without the last 12 observations, obtain the point predictions and the corresponding 95% confidence interval for the last year. Plot the original series (last 5 years) with the predictions and intervals overlaid.
8. Calculate measures of forecasting ability (RMSPE and MAPE) from the above point predictions. This gives an idea of the accuracy of the predictions.
9. Calculate the mean of the widths of the prediction bounding intervals. This is a measure of the precision of the predictions.
10. With all the above information, which of the two proposed models do you select as the best? For the chosen model fitted to all data, calculate forecasts with confidence intervals for the next year.

2.1 Hints:

R code for validating ARIMA models

Let `model` be the R object resulting from fitting and `resi` the extraction of residuals (`resi<-resid(model)`). The following functions are useful for validation:

- Plot of the residuals:

  ```
  plot(resi); abline(h=0); abline(h=c(-3*sd(resi),3*sd(resi)),lty=3,col=4)
  ```

- Plot of the square root of the absolute value of the residuals with smooth fitting:

  ```
  scatter.smooth(sqrt(abs(resi)), lpars=list(col=2))
  ```

- Normality plot:

  ```
  qqnorm(resi); qqline(resi,col=2,lwd=2)
  ```

- Histogram with normal curve superimposed:

  ```
  hist(resid,breaks=20, freq=FALSE); curve(dnorm(x, mean=mean(resid), sd=sd(resid)),
  col=2, add=T)
  ```

- Shapiro-Wilks Normality Test (Null Hypothesis of Normality):

  ```
  shapiro.test(resi)
  ```

- P-values for Ljung-Box tests:

  ```
  tsdiag(model,gof.lag=72)
  ```

- Modules of the roots of the characteristic polynomials:

  ```
  Mod(polyroot(c(1,-model$model$phi))); Mod(polyroot(c(1,model$model$theta)))
  ```

- Extraction of the AIC and the BIC of the model:

  ```
  AIC(model); BIC(model)
  ```

- Series without the last 12 observations:

  ```
  ultim=c(2017,12); serie2=window(serie, end=ultim)
  ```

- Long-term predictions (pred: point predictions, se: standard deviation):

  ```
  predict(model, n.ahead=12)
  ```

- Plot of the original series and the point and interval predictions (pr: point prediction, tl: lower limit of the prediction, tu: upper limit of the prediction):

  ```
  ts.plot(serie,tl,tu,pr,lty=c(1,2,2,1),col=c(1,4,4,2),xlim=c(2013,2018),
  type="o"); abline(v=2013:2018,lty=3,col=4)
  ```

- Calculation of predictive ability measures (obs: original series values reserved, pr: point predictions):

  ```
  RMSPE=sqrt(mean(((obs-pr)/obs)^2)); MAPE=mean(abs(obs-pr)/obs)
  ```

- Average length of prediction intervals (tl: lower prediction limit, tu: upper prediction limit)

  ```
  mean(tu-tl)
  ```