**STEPS IN TIME SERIES DATA ANALYSIS**

Make sure that you have at least 50 data points in your dataset. Keep last 4 (for yearly and quarterly) or 12 (for monthly) observations as test data. Your data should be up to date.

1. Introduction covering data descripition, aim and the source of data.
2. Time series plot and interpretation (Visually determine the existence of a trend, seasonality, outliers).
3. Keep several observations out of the analysis to use them to measure the forecast accuracy of the models. (For yearly data and quarterly data 4 or 5, monthly data 12).
4. Box-Cox transformation analysis: If the series need any transformation, do it. If the information criterion values are too close to each other, don’t transform the data.
5. Make a anomaly detection and if necessary clean the series from anomalies (use *anomalize*, *forecast* (*tsclean* function) or *AnomalyDetection* packages).
6. ACF, PACF plots, KPSS and ADF or PP test results for zero mean, mean and trend cases and their interpretation. For seasonal unit root, HEGY and OCSB or Canova-Hansen tests are required.
7. If there is a trend, remove it either by detrending or differencing. You may need to apply unit root tests again.
8. Then, look at the time series plot of a stationary series, ACF and PACF plots, information table, ESACF.
9. Identify a proper ARMA or ARIMA model or SARIMA model.
10. After deciding the order of the possible model (s), run MLE or conditional or uncondinitional LSE and estimate the parameters. Compare the information criteria of several models. (Note: If there is a convergence problem, you can change your estimation method).
11. Diagnostic Checking:
12. On the residuals, perform portmanteau lack of fit test, look at the ACF-PACF plots of the resuduals (for all time points, ACF and PACF values should be in the white noise bands), look at the standardized residuals vs time plot to see any outliers or pattern.
13. Use histogram, QQ-plot and Shapiro-Wilk test (in ts analysis, economists prefer Jarque-Bera test) to check normality of residuals.
14. Perform Breusch-Godfrey test for possible autocorrelation in residual series. The result should be insignificant.
15. For the Heteroscedasticity, look at the ACF-PACF plots of the squared residuals (there should be no significant spikes); perform ARCH Engle's Test for Residual Heteroscedasticity under *aTSA* package. The result should be insignificant. If the result is significant, you can state that the error variance is not constant and it should be modelled, but don’t intend to model the variance. If there is a heteroscedasticity problem, most probably normality test on residuals will fail. The high values in the lower and upper extremes destroy the normality due to high variation. In your project, you can state these only. When solving a real life problem, you cannot just state and quit dealing this problem!
16. Forecasting: The number of forecasts should be same as the length of your test data.
17. Perform Minimum MSE Forecast for the stochastic models (like ARIMA or SARIMA)
18. Use *ets* code under the *forecast* package to choose the best exponential smoothing (simple, Holt’s, Holt-Winter’s) method that suits your series for deterministic forecasting.
19. Obtain forecasts using neural networks (*nnetar*)
20. If you transformed the series for SARIMA model, back transform the series to reach the estimates for the original units.
21. Calculate the forecast accuracy measures and state which model gives the highest performance for your dataset.
22. Provide plots of the original time series, predictions, forecasts and prediction intervals on the same plot drawing the forecast origin for ARIMA models, exponential smoothing method and neural networks. The plot for each model should look like the following plot.



1. Give your conclusion.