MADS – "Statec Workshop" (Vasja S.)

Homework IV – NN model

In this homework, you will forecast the time series in a NN model. You can use alternative packages, commands or software. As before, return the homework in a html, word, pdf or markdown file. Please also attach the script&data which performed the analysis.

Deadline for homework IV is 15th of January.

Name all the submitted files as follows:

HWIII_analysis_[yourLastName].ipnyb/doc/pdf – In this file you describe the analysis

HWIII_data_[yourLastName].* - this file should contain the data

HWIII_script_[yourLastName].* - this file should contain the code which produces results

If you are using markdown, HWII_script_[yourLastName].* is not needed.

Please perform the following steps:

1. Forecast the series that you selected in HWI with a Neural Network (use the 80-20 sample split as in HWII and HWIII). Use the series from HWII (VAR), and their lags. The Neural Network model does not have to be perfect. It is sufficient to estimate a small network, with one hidden layer and a small number of nodes (say 2-5). I would advise the use of python packages (Keras, TensorFlow, PyTorch…) since the code which we used in class was highly fine-tuned to make it work. Most Python packages take care of initialization and fine-tuning automatically. This will make your task a lot easier.
2. Compare forecasting performance of ARIMA (HWII), VAR (HWIII) and the NN model (this HW). For each of the 3 models calculate RMSFE (Root Mean Squared Forecast error; see below) on the last 20% of the sample (using the 1-step-ahead expanding-window forecasts – do not use multi-step ahead forecasts!). Describe which model performs best. Comment on what could be the reasons for that model to perform the best?

BONUS: Estimate a more advanced NN such as LSTM.

RMSFE is a measure of the spread of the forecast error distribution (=y_true – y_forecasted). It is calculated as the square root of the average (squared) forecast error:

$$RMSFE = \sqrt{\frac{\sum_i \left(y_i^{true} - \widehat{y_i}^{forecast}\right)^2}{N}}$$

A perfect model would produce RMSFE = 0. The larger the RMSFE the worse the model performs. RMSFE is "scale dependent". Its value depends on the scale of y. This implies that you can only compare models (ARIMA/VAR/NN) in which variable y is on the same scale (for example, for all the 3 models y needs to be in growth rates; it cannot be that for NN y is in levels and for VAR in growth rates…).