Time series project, review of article : ARIMA vs. ARIMAX, by Durka Peter

Alexandre ANDREINI
MVA
alexandreandreini@hotmail.fr

Antoine DOIZE
MVA
antoinedoize@hotmail.fr

Abstract

This is a report for the Time Series project of the MVA course. The objective of the project is to provide the key concepts in order to follow our presentation during the student seminar of the Times Series course. We will explain the key concepts of the article [1], the different steps of our implementations and explain the way we tried to contest the conclusions of the article. We will show that the method Durka used to prove that ARIMA is better than ARIMAX is flawed, and provide another comparison of the two methods. The code can be found on github at this link [2]

1 Presentation of the article

The goal of the article is to compare two models of time series forecasting : ARIMA and ARIMAX. Those are two autoregressive models that forecast values but by two different ways.

Let us call B the operator on time series such that :

$$B(x_t)_{t \in R} = (x_{t-1})_{t \in R_*^+}$$

ARIMA is forecasting the next value of a sequence y_t with the previous values of the same sequence : (combining the values y_t and their seasonality and the values of the noise a_t and their seasonality).

$$\begin{aligned} y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \Phi_1 y_{t-s} + \Phi_2 y_{t-2s} + \cdots + \Phi_p y_{t-P*s} \\ + a_t - \theta_1 a_{t-1} - \cdots - \theta_p a_{t-q} + \Theta_1 a_{t-s} - \Theta_2 a_{t-2s} - \cdots - \Theta_Q a_{t-Q*s} \end{aligned} \tag{1}$$

Note that for a given a sequence x_t, we denote $ARIMA(p, d, q)(P, D, Q)$ if we apply ARIMA to $(1 - B)^d (1 - B^s)^D$. This is the ARIMA model extended.

ARIMAX is forecasting the next value of a sequence y_t with the previous values of another sequence x_t : (combining the values x_t and their seasonality and the values of the noise a_t and their seasonality).

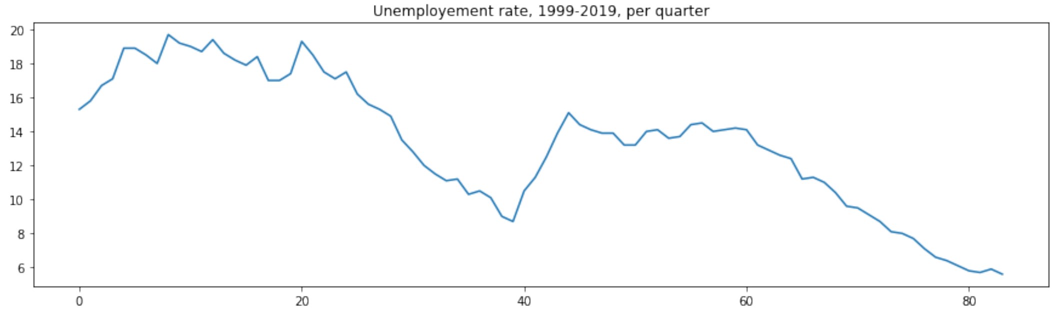
$$\begin{aligned} y_t = C + \nu_0 x_t + \nu_1 x_{t-1} + \nu_2 x_{t-2} + \cdots + \nu_K x_{t-K} \\ + a_t - \theta_1 a_{t-1} - \cdots - \theta_p a_{t-q} + \Theta_1 a_{t-s} - \Theta_2 a_{t-2s} - \cdots - \Theta_Q a_{t-Q*s} \end{aligned} \tag{2}$$

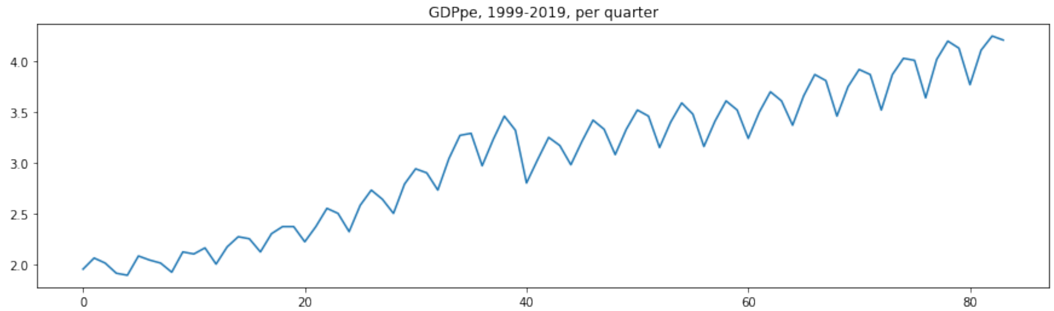
The way those models are fitted have been highly studied and we will skip this part to go directly to the conclusions of the article. To get more details on those parts you can refer to [3] and [4].

The models are fitted on an example of two sequences : Growth Domestic Product per Capita, that can be forecasted from itself (ARIMA) or from Unemployement Rate values sequence (ARIMAX).

The results show that even if the two methods are quite efficient for short term forecast, ARIMA is slightly better at forecasting GDP than ARIMAX. The conclusion of the article is that ARIMAX is better at forecasting macroeconomics series thatn ARIMA. We will begin by reproducing the experiments of the article in order to get back to their results, and then try to show that their conclusion is flawed, and that the difference of performance isn't due of a difference in the models performance, but in the tasks difficulty.

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.





2 Reproducing the results of the article

In this section, we reproduce the found ARIMA models of the article.

Let us start with the GDP. Logic and the PACF give us a seasonal parameter s=4 corresponding to a year delay. GDP per say do not benefit from the stationarity property as the Augmented Dickey-Fuller test do not permit to refute the null hypothesis. Thus we need to find which derivative is stationary.

PACF hints us to check whether $z_t = (I - B)(I - B^4)GDP_t$ is stationary, which is the case according to the ADF test. Looking now at the PACF, z_t depends mainly on z_{t-4} which leads to $p = 0$, $P = 1$. If we now fit $ARIMA(0, 1, 0)(1, 1, 0)$ to our data, we find the same numerical parameters than the article and a confirmation through the correlograms that q=Q=0 leads to a better suited model.

For the UR, the results are less straightforward. As mentionned in our notebook, we think the authors did not follow their own methodology properly and choose the model based on low AIC. Given the datas they had at the time of the article, we would have concluded to an $ARIMA(1, 0, 1)(1, 1, 0)$ fit while they conclude to an $ARIMA(0, 1, 1)(2, 0, 0)$. With the data we have now, their model can be induced properly but at their time of writing they could not affirm that $(1 - B)UR_t$ was stationary. However, both models lead to similar performance.

Once those two models obtained, we obtain similar numerical parameters for the transfert function.

3 Contesting the article conclusion

3.1 Counter Example

We thus agree with the fact that the ARIMAX performs a little less accurately on the task of predicting GDP from UR than ARIMA does by predicting GDP from the previous values of GDP, but according to us this can not lead to the conclusion that ARIMAX is better to forecast macroeconomics time series.

A trivial counter example would be to take a simple electrical system composed of an alternative generator and a bulb working effectively as a resistance of known parameter R. Assume our tools are such that we can measure voltage with a very good precision while intensity is measured poorly. Then it makes sense to extrapolate data from U(t) and then use a transfer function I(t)=U(t)/R.

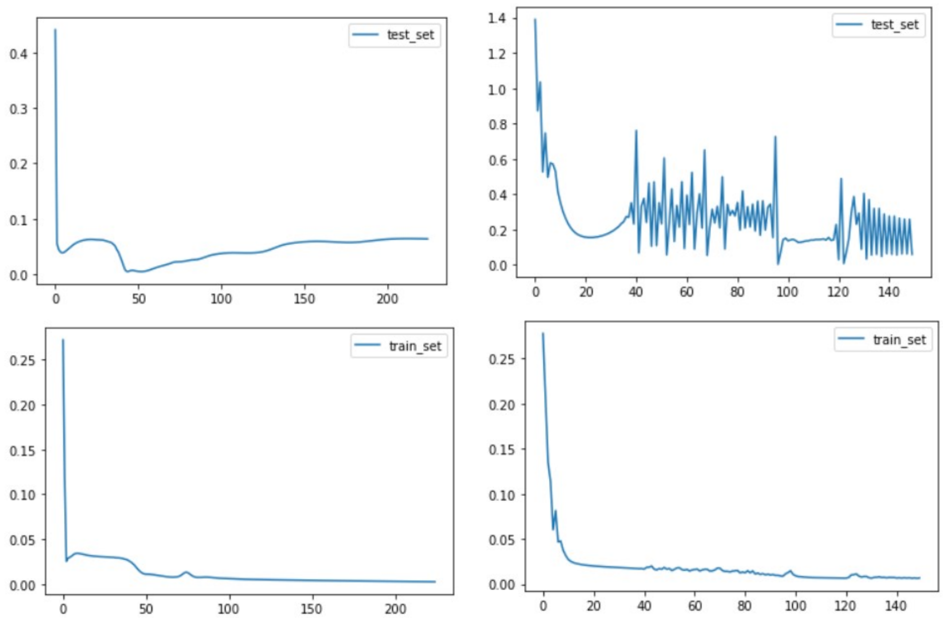


Figure 1: LSTM training to forecast a time series from previous values of the same time series. Left : predicting GDP from GDP values, Right : predicting UR from UR values

Another reformulation of this counterexample would be to take a HMM with known dependency in the hidden serie: ARIMAX is better suited to predict the noisy sequence than ARIMA.

3.2 Can GDP be easily inferred from UR

A better approach to those results would be to understand why that model do not yield great results. By taking this model, the authors assumed an underlying linear dependency between those two series. While stating that decreasing GDP leads to let down and thus UR increasing, it is by no way clear why the dependency would be linear. On top of that, there is no evidence whatsoever that the relation can be easily computed.

So using ARIMAX here presuppose we can both predict UR thanks to its past and infer GDP from UR and its past.

Let us handle the first assumption : the model we found in the previous reproduction of results is clearly not as good as the one we found for GDP. Even using other techniques such as LSTM, with the same architecture, show that the two tasks have very different difficulties.

The model we used is a simple one-LSTM-cell, one linear layer arhitecture, that takes as input the 4 last values of a sequence (GDP or UR) and predicts the next value of GDP. At first we trained the model on a similar task predicting the next value of a time series with previous values of the same time series (to check that the model is adapted to the task of forecasting time series). The results of the training are shown in 1. The model is adapted to the task of forecating time series from previous values of the same time series with a final RMSE on the test set of 4e-3 for the GDP prediction and 2e-3 for the UR prediction, and a nice training behaviour.

The second assumption is that we can deduce GDP from UR and its past. This seems also a very hard problem. For instance, we tried to predict GDP using VARIMA on either GDP or (GDP and UR) and the first one gave us better results.

We can also see the results on the LSTM used to predict GDP from UR on 2. Two things can be highlighted. First, the best performance reached is much poorer : the final RMSE on the test set is 0.17. Moreover, the training was much more instable, and we can see that the model is much more eager to overfit (in less than 20 epochs we can already see an overfitting behaviour), which is

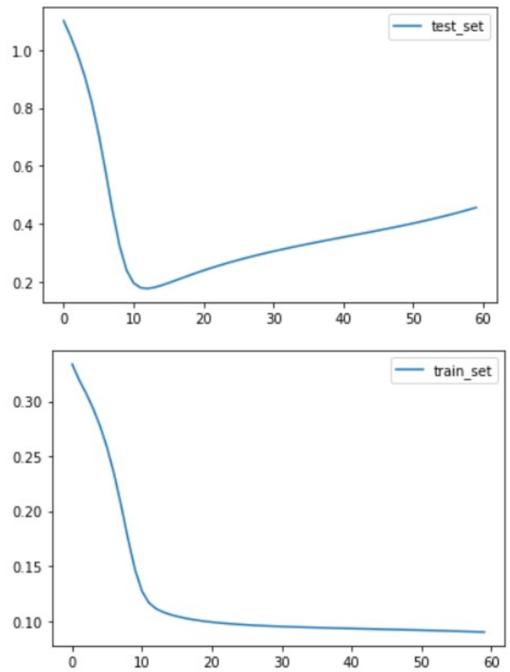


Figure 2: LSTM training to forecast GDPvalues from previous values of UR time series.

caracteristic of the difficulty of learning general rules on the datasets, and a trend to try and learn "by heart" the training samples.

Those experiments show that the very same model has much more difficulties to forecast GDP from UR than forecasting GDP from GDP or UR from UR. The training process is much more likely to overfit (which is characteristic of a difficult task : the model will perform better by learning the training samples by heart instead of inferring general rules).

To sum it up, both assumption can be refuted and The performances highlighted in [1] by Durka isn't due to the difference of model performance but rather due to the difference of task difficulty.

4 Conclusion

The article [1] of Durka is showing two methods of forecasting Growth Per Capita values : one is inferring GDP values from previous GDP values (ARIMA) and the other one is inferring GDP value from previous UR values. The article is showing that ARIMA is performing better than ARIMAX on the task.

However, we have shown that the conclusion of the article "ARIMA is better than ARIMAX to forecast macroeconomics time series" is flawed, and that the difference of performance is only due to a difference in task difficulty. In other situations, we could show that ARIMAX would perform better at forecasting than ARIMA. Those two models are performing different task, so we shouldn't try and compare their performances directly, but try to compare the two tasks, and chose the more adapted one. Then the correct conclusion would not be a statement "Among the two models, which is better?" but rather the statement "Among the two tasks, which is easier?".

The code can be found at [2].

References

[1] Pastoreková Silvia Ďurka Peter. Arima vs. arimax – which approach is better to analyze and forecast macroeconomic time series? volume 2, pages 136–140, 2012.

[2] Antoine Doize Alexandre Andreini. Ast project. `https://github.com/antoinedoize/AST_project_MVA`, 2021.

[3] G. M. – Reinsel G. C. Box, G. E. P Jenkins. Time Series Analysis: Forecasting and Control. 2008.

[4] L Rublíková, E. – Marek. Linear transfer function model for outflow rates. *Ekonomické rozhl'ady*, 30:457–466, 2001.