Hi Sunny –

Thank you for the most recent notebook updates. I feel that we’ve now basically gone through the majority of relevant options for optimizing parameters in the VAR model. The most recent results were enlightening too. I feel its very interesting to see causality \_x and \_y leading to the best results… but I also found it equally interesting that using random tickers lead to almost as good of results. In fact, as long as we didn’t use too many tickers, I’d say accuracy for any of the methods was likely sufficient to be used for actual predictions. Another big takeaway for me is that perhaps the most important detail is the quantity of tickers used in the model where around 30 looks to be the sweet spot.

I’m hoping we can do one more set of tests where we add a couple more tweaks before considering this optimization component complete.

**Part 1** – Adding additional components to the existing workflow

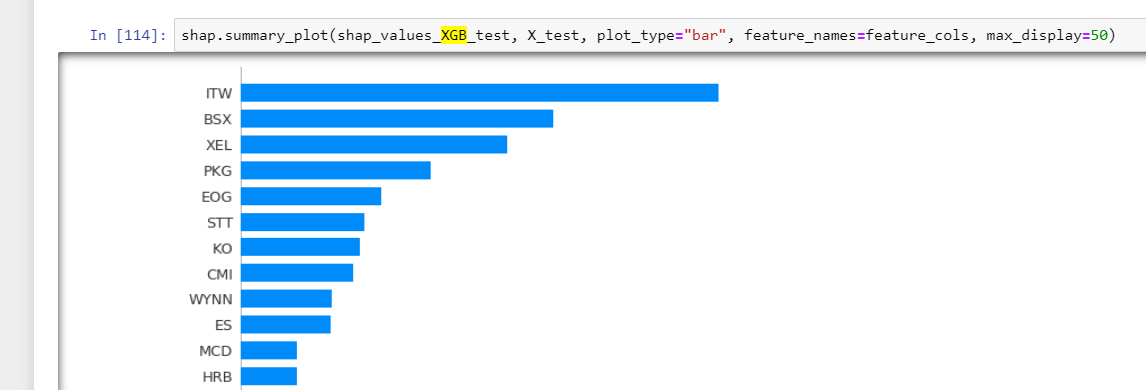
A) Adding one last filtering type for ticker selection

I’d like to try one more technique for ticker selection of our VAR models. This will add a layer of complexity to the model but I think its worth taking a look at…

Can we add XG Boost to the beginning parts of the notebook… and then use the SHAP/LIME libraries to rank the most important “features” for the XG Boost model. All the features in the model are actually stock tickers… so we’ll take the feature importance list and use it as ranking to find the most desirable and least desirable tickers to then subsequently use in our VAR model. So for example, in the scenario where we’re using only 5 tickers in the VAR model, we’d actually use the 5 tickers that XG Boost found to be most important. When we use 10 tickers in the VAR model, we’ll use the 10 tickers that XG Boost found to be most important. Etc, etc for the remaining quantities of tickers.

The important detail here will be the actual ranking of the tickers… whereas the model is always using the X amount of most desirable tickers per XG Boost’s feature importance guidance.

The below imagine was taken from our previous workbook where we attempted to apply XG Boost as a regression predictor for csi1. In this case, it found ITW, BSX, XEL (etc, etc,…) to be the most important tickers for the model.



B) Create a duplicate notebook but try to predict a different ticker’s future values instead of csi1.

I’d like to take our completed notebook and then try to use the same prediction methodology on some other tickers other than csi1. I’m just curious how the results will look. Csi1 is calculated using stock prices from many many different tickers… I’m wondering if the MSE/Bias/RMSE results will be different if we try the same prediction strategy on a different ticker… and if the model will have a harder time with other tickers (csi1 may be easier to predict since its calculation is kind of a summary of the stock market).

There are many other tickers we can try predicting for. I suggest we just pick 1 or 2. Any one will do… such as BLL, JPM, etc.

**Part 2** – New notebook which will be a quick experiment

#3 – Experimenting with VECM model type

As I understand it, VECM models are meant to predict data where cointegrated relationships exist. I believe many of the stocks and csi1 are cointegrated. In a new notebook, can we quickly stand up a VECM model with our existing data and see how the results look? No need to build it out too extensively. Maybe try a few combinations of parameter configurations. If the results seem promising and superior to those seen in our VAR models, then we can explore VECM more comprehensively.

<https://towardsdatascience.com/vector-autoregressions-vector-error-correction-multivariate-model-a69daf6ab618>

In the example from the above article, the author was able to get much better results using a VECM model compared to the same experiment using a VAR model. Granted, the experiment is different than ours, but I still think it’s worth a quick try to see if any initial results look encouraging for more investigation later.

I suggested this to be done as a separate notebook but it can be in our existing notebook if you’d prefer. Up to you.