Report on Manuscript Number:

CSEM-D-16-00137

Fast and adaptive cointegration based model for forecasting high frequency financial time series

I. Summary

In this paper authors aim to extend the VECM approach to the case of high frequency data, called adaptive VECM (AVECM). More specifically, authors aim to analyze the forecasting accuracy for the VECM under the case of high frequency data. Two main findings are presented in this paper. First, authors show that the number of cointegration relationships varies with the length of historical data. Second, authors show that AVECM out perform others models (such ARIMA and random walk model) in term of forecasting.

II. General comments:

1. The paper is very well written.
2. The paper has a real contribution from a technical point of view. However, I'm not convinced by whether determining long run equilibrium in high frequency data make sense. In other words, when we have 5 minutes data, what is the long term and what is the short term? This is my principle subtlety for this analysis.
3. In line to my previous comments, authors did not motivate any where in the paper, the importance of long-run equilibrium and short term interaction for economic and/or financial point of view, under high frequency data. In low frequency data (monthly, quarterly) the long run has an economic interpretation (for example economic cycle, Phillips curve relationship, and others economic background). Authors should highlight whether this information coming from long-run equilibrium for the case of high frequency data are useful from a theoretical point of view. In other words, whether this information are useful for policymakers, investors, hedgers and economic agents?

III. Specific comments:

4. Authors say "*In order to show that the number of cointegration vectors depends on the amount L of historical data and the number of lags p in the VECM, we used a grid search. We defined a grid of possible values for L and p. L goes throughout [2, 14] hours (1 hour = 360 data points) with a step size of 4 hours and p throughout [1, 5] with step size of 1.*" My comment is as follows: what are the motivations behind the choice of these different grids. For example, for the choice of L goes through [2, 14] hours, what are the raisons, as in the financial markets trading is made through cession of 6 hours or 24hours.
5. How you can interpret the figure 1. For example, the probability of having one cointegration relationship is equal to one for 10 hours.
6. After presenting the PC ratio formula, authors run an experiment with L go through [700, 1500] data points. For 700 data points is approximately corresponds to 2 hours. However, why the choice of 1500 data points and not 5000 which corresponds to 14 hours?
7. For figure 2, authors interpret the first one "We found that better cointegration percentage leads to better performance accuracy in terms of reducing MSE ». What bout the second figure with L. indeed, we observe that MSE is lowest with L equal to 1050. How you explain this result?
8. The empirical application is well done.
9. The authors used four currencies markets in order to analysis their relationship based on the proposed approach AVECM. However, the economic analysis still very scarce or absent.

IV. Conclusion

It is an interesting paper with very important technical contribution. However, my two major comments on the utility of analyzing the long run relationship for high frequency data.