Modelling volatility clustering with LSTM Neural Networks
Research Proposal for a Master Thesis
at the University of St. Gallen

Lukas Schreiner
lukas.schreiner@student.unisg.ch

July 5, 2018

1 Introduction & Research Goal

Neural networks experienced an increasing popularity in recent years and many improvements in algorithms and network architecture have lead to a broad field of applications. Consequently, their potential for the financial sector is tremendous as well. When it comes to forecasting financial data of stocks, especially second order moments seem to exhibit some structure. This volatility clustering is subject to a rich family of autoregressive models that tries to catch and predict this heteroscedasticity.

In machine learning recurrent, neural networks (RNNs) are suited best for learning patterns in time series data. However, their capabilities of learning from events that a farther apart in time is quite limited. This so called "vanishing gradient" problem can be addressed by using Long Short Term Memory (LSTM) cells instead of the usual activation functions. Those LSTM Cells are capable of catching long term dependencies by using states that are able to store information.

The main goal of this master thesis is to examine whether this feature of LSTM neural networks improves the performance of volatility forecasting in comparison to usual recurrent neural networks and generally tries to gain insights about the optimal architecture and use of hyper parameters in that context.

2 Methodology & Data

In general, different models and architectures will be defined and trained on the same training set. This training set will consist of daily squared log-returns of a stock over the period of 4 years. The subsequent year will be used as a test set and the models will be used to predict daily squared returns for the next day. An error metric like root mean squared error (RMSE) will be used to to asses prediction performance of the different models. As a comparison an ARCH(1) and GARCH(1,1) as well as a naive forecast (todays variance as forecast for tomorrow's variance) will be computed as well.

In order to estimate the models, Keras[1] with a Tensorflow[2] backend will be used in order model the neural networks. For the ARCH and GARCH models, statistical software will be used.

[1] www.keras.io
[2] www.tensorflow.org

3 Structure

1. Introduction
2. Literature Review
3. Neural Networks
   (a) Feedforward Networks
   This section describes the general architecture of a feedforward network and general concepts like activation functions, back propagation, gradient descent and normalising data.
   (b) Hyperparameters, Bias and Variance
   This chapter covers briefly how hyper parameters can influence model performance and techniques to avoid overfitting or high bias like regularisation and drop out.
   (c) Recurrent Neural Networks
   This section introduces recurrent neural networks and their field of application as well as their shortcomings.
   (d) LSTM Neural Networks
   Based on the last chapter, LSTM cells are introduced.
4. Data Set
   see above
5. Models & Results
   (a) ARCH, GARCH, naive model
   (b) Recurrent neural networks & deep RNNs
   (c) LSTM, deep LSTM & bidirectional LSTMs

   Here the the specific architectures used will be explained as well as the results will be discussed.
6. Conclusion & Outlook

4 Initial Research

Recurrent neural networks were first introduced by Hopfield (1982) and have been further developed ever since. A major breakthrough, however, was a paper by Hochreiter and Schmidhuber (1997) that developed the theoretical framework for LSTM networks.

Forecasting stock prices and volatility with machine learning is a relatively young field, which results in a very thin body of literature. Tino et al. (2001) examine volatility trading models with recurrent neural networks. They find that RNNs tend of overestimate noisy data and only exhibit short term memory capabilities. A recent paper by Liu et al. (2017) reports a significant improvement in volatility prediction when including sentiment analysis in the models.

When it comes to modelling volatility with LSTM neural networks, initial research yielded almost no scientific approaches. A conference paper, written in Japanese, by Goshima et al. (2017) shows that that LSTM and RNNs indeed can produce lower MSEs when modelling financial volatility. Other approaches by Zhou et al. (2018) use search terms of the baidu search engine[3] in order to predict volatility on the stock return. They report significant improvement to the benchmark GARCH model when it comes to forecasting volatility that way.

[3] www.baidu.com

References

Goshima, K., Takahashi, H., and Terano, T. (2017). Modeling the volatility clustering with recurrent neural networks. *Abstracts of Annual Conference of Japan Society for Management Information*, 2017s:29–31.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558.

Liu, Y., Qin, Z., Li, P., and Wan, T. (2017). Stock volatility prediction using recurrent neural networks with sentiment analysis. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, pages 192–201. Springer.

Tino, P., Schittenkopf, C., and Dorffner, G. (2001). Financial volatility trading using recurrent neural networks. *IEEE Transactions on Neural Networks*, 12(4):865–874.

Zhou, Y.-L., Han, R.-J., Xu, Q., and Zhang, W.-K. (2018). Long short-term memory networks for csi300 volatility prediction with baidu search volume. *arXiv preprint arXiv:1805.11954*.