Predicting Stock Returns based on quarterly SEC Reports using Deep Neural Networks

Author: Daniel Winkler
Supervisors: Florian Huber & Christian Hotz-Behofsits

1 Predicting Stock Returns

Introduction

- Predicting stock returns is, like, really hard! (Welch and Goyal, 2007)
- Often a simple Random Walk or AR(1) outperforms complex models
- Company specific "fundamentals" improve predictions marginally
- ... so do market indicators
- **What else could be used as predictors?**
- Think beyond numeric data!

Model TEXT as a sequence of characters

- Basic idea (Krause et al., 2016):

$$p(c_1, c_2, \ldots, c_T) = p(c_1)p(c_2|c_1)p(c_3|c_1, c_2) \ldots p(c_T|c_1, \ldots, c_{T-1})$$

- c_i being the i-th character in the TEXT
- Minimize the prediction error of the next character given all previous ones
- Can be solved efficiently using gradient descent → "Deep Learning"

2 Deep Learning

Quick Introduction

- Similar to maximum likelihood (Goodfellow et al., 2016, pp. 128f.)!
- Adjust coefficient matrices until negative log-likelihood is minimized
- However, we have **multiple** coefficient matrices which are linear
- Inbetween each operation there is a non-linearity so matrices are identified
- Deep in the sense that we do not see where the change happens and why!

3 Proposed Model

Recurrent Neural Network

- Class of DL models for sequences

$$h_t = f_a(W_h h_{t-1} + W_x x_t), \tag{1}$$

- Specifically: multiplicative long short-term memory (mLSTM)
- Achieves state of the art in sentiment discovery (Radford, Józefowicz, and Sutskever, 2017)
- I'll show that it can also "learn" about company reports

Econometric Framework

- Dynamic Panel regression

$$r_{kt} = \alpha_k + \phi r_{kt-1} + F_{kt}\beta_1 + S_{kt}\beta_2 + \varepsilon_{kt}, \quad \varepsilon_{kt} \sim N(0, \sigma^2) \tag{2}$$

- S_{kt} is the final cell state; F_{kt} the fundamentals
- Using shrinkage priors (Piironen and Vehtari, 2017) to get rid of unnecessary coefficients
- Estimate using Hamiltonian Monte Carlo with Stan
- Evaluated using log-predictive density and RMSE (Geweke and Amisano, 2010)

4 Results

- Overall models with text information + AR best
- So mLSTM does extract relevant information
- Shrinkage is major problem for many predictors. Suggestions?
- Texts do not work for all companies. How to identify which?
- Improvement is marginal. We average over quarters.
- How about 1-2 weeks after publishing of report?
- Using only companies whose reports matter?

References

Geweke, John and Gianni Amisano (2010). "Comparing and evaluating Bayesian predictive distributions of asset returns". In: *International Journal of Forecasting* 26.2, pp. 216–230.

Goodfellow, Ian et al. (2016). *Deep learning*. Vol. 1. MIT press Cambridge.

Krause, Ben et al. (2016). "Multiplicative LSTM for sequence modelling". In: *arXiv preprint arXiv:1609.07959*.

Loughran, Tim and Bill McDonald (2011). "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks". In: *The Journal of Finance* 66.1, pp. 35–65.

Piironen, Juho and Aki Vehtari (2017). "Sparsity information and regularization in the horseshoe and other shrinkage priors". In: *Electronic Journal of Statistics* 11.2, pp. 5018–5051.

Radford, Alec, Rafal Józefowicz, and Ilya Sutskever (2017). "Learning to Generate Reviews and Discovering Sentiment". In: *CoRR* abs/1704.01444. arXiv: `1704.01444`. URL: `http://arxiv.org/abs/1704.01444`.

Welch, Ivo and Amit Goyal (2007). "A comprehensive look at the empirical performance of equity premium prediction". In: *The Review of Financial Studies* 21.4, pp. 1455–1508.