Recurrent Neural Networks vs. Classical Methods in Investment Strategies

- work in progress -

Mateusz Kijewski and Robert Ślepaczuk

December 17, 2019

Agenda

- Motivation
- Hyptotheses and research questions
- Data
- Methodology
- Literature review
- Results
- Summary
- Research etensions

Motivation

What?

- to show the most important issues affecting the overoptimisation
- to produce forecast of equities, in our case S&P500 index

Why?

- for investment purposes, in order to built **robust not overfitted** algorithmic investment strategies
- to explain the process of investment strategies testing

How?

- through the usage of standard historical methods (e.g. TS models, MAs, momentum/contrarian, TA rules or macro),
- through the combination of classical signals
- through the usage of ML/RNN/LSTM techniques for price prediction process
- through the combination of ML and classical signals

Hypotheses and research questions

1. Classical techniques (Arima, Continuation/Reversal, MAs, etc) are not valid for price prediction process
2. ML techniques like RNN/LSTM are not valid for price prediction process

not valid = not able to beat the market, i.e to obtain abnormal returns

abnormal returns = better than the benchmark and the minimum acceptance level of performance statistics (min{ARC, IR} and max{aSD, MD, MLD})

3. ML techniques are more prone to overoptimisation process than classical forecasting techniques
4. The combination of simple investment signals works the same like the diversification through additional number of basis instruments
5. The inability to set the correct value of hyperparameters affects the robustness of ML techniques

Data

- data source: https://finance.yahoo.com/ and https://fred.stlouisfed.org
- data period: 1997-10-28 - 2019-01-01
- out-of-sample period: 2001-01-01 - 2019-01-01
- training window: 252d for Arima, the length of the longestMA for MA, 252 for LSTM,
- test window: 252d for Arima and for MAs,
- optimisation criterion: IR
- frequency: daily
- fees: 0.00025 -> the equivalent of nominal transacion fee and dollar bid/ask spread

Methodology I

1. model estimation (estimation window)
2. model selection based on optimisation criterion (IR) selected on estimation or preferably on test window
3. out-of-sample tests of classical and ML time series forecasting methods by the evaluation of out-of-sample equity lines based on the performance statistics (ARC, aSD, MD, AMD, MLD, allRisk, ARCMD, ARCAMD)
4. the combination of signals from classical methods
5. the combination of signals from classical and ML methods
6. the approach enabling to select the signals from the given method during the in-sample period for out-of-sample application

Methodology. The parameters and signals

Classical methods tested - parameters

- ARIMA(p, d, q) {estimate window = test window, optimisation criterion, refresh window}
- Reversal(1d) {}
- Continuation(vector of n-day returns) {weight vector}
- MovingAverage(shortMA, longMA) {estimate window, test window, optimisation criterion, refresh}
- MacroFactor(ith factor) {transformation method, percentile window, no of in and out percentile}
- VolatilityBreakout(ith volestimator) {parameter history, percentile window, no of in and out percentile}

Signals

- whiteboardgraph-0

Methodology. The ways of testing of parametrised classical methods

Continuation, Reversal, Macro, VolBreakout

- whiteboardgraph-1 -> not proper approach -> expert parametrisation

ARIMA_rolling or anchored

- whiteboardgraph-2 -> proper but not adequate approach

MAs_rolling

- whiteboardgraph-3 -> the adequate approach

ARIMA_two winodws & two optimisation criteria

- whiteboardgraph-4 -> more advanced adequate approach {2x2 matrix}

Methodology. The performance statistics

- return: ARC,
- risk: aSD, MD, MLD, allRisk,
- risk adjusted returns: IR, ARCMD, ARCAMD
- descriptive: numbTrans, stopSignal

the minimum acceptance level of performance statistics

- min ARC $= 10\%$
- max aSD $= 20\%$
- min IR $= 1$
- max MD $= 10\%$
- max MLD $= 1$
- max allRisk $= 1$

Methodology. Overview of ML methods

AI

-> ML

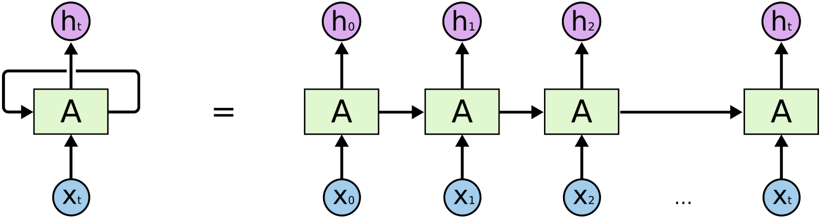
-> -> Supervised Learning

-> -> -> -> -> -> -> -> -> Neural Networks

-> -> -> -> -> -> -> -> -> -> -> -> -> -> RNN

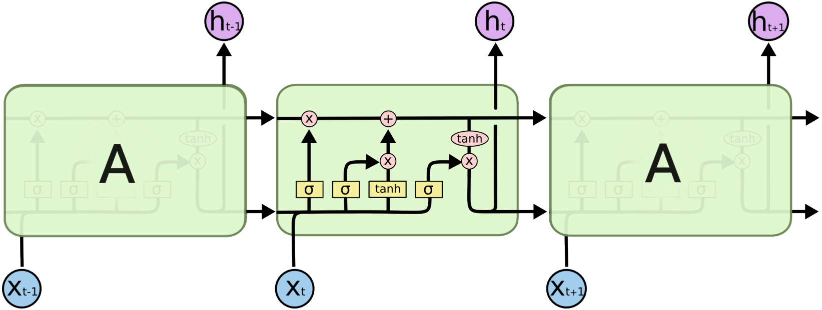
-> -> -> -> -> -> -> -> -> -> -> -> -> -> -> -> LSTM

Methodology. The logic of LSTM I



Source: https://colah.github.io/posts/2015-08-Understanding-LSTMs/

Methodology. The logic of LSTM II



Source: https://colah.github.io/posts/2015-08-Understanding-LSTMs/

Methodology. The logic of LSTM III

Types of LSTM

- Univariate LSTM
  - Vanilla LSTM
  - Stacked LSTM
  - Bidirectional LSTM
  - Convolutional LSTM
- Mulltivariate LSTM

Methodology. The parameters

ML/RNN/LSTM

- LSTM(seq, window, feature, epochs, units, activation, optimizer, loss, structure, learningrate, dropout rate)

initial values of hyperparameters

- sequence: 5
- window: 251
- epochs: 300
- units: 5
- activation: relu
- optimizer: adam
- loss: mse
- type: rolling
- dropout: 0.2
- learningrate: 0.01

Literature review

- Hochreiter and Schmidhuber (1997) -> the first introduction of LSTM. By introducing Constant Error Carousel (CEC) units, LSTM deals with the exploding and vanishing gradient problems. The initial version of LSTM block included cells, input and output gates.
- Gers and Schmidhuber (1999) -> they introduced the forget gate (also called "keep gate") into LSTM architecture, enabling the LSTM to reset its own state.
- Gers, Schmidhuber and Cummins (2000) added peephole connections (connections from the cell to the gates) into the architecture. Additionally, the output activation function was omitted
- Kyunghyun Cho et al. (2014) put forward a simplified variant called Gated recurrent unit (GRU)

Classical forecasting methods

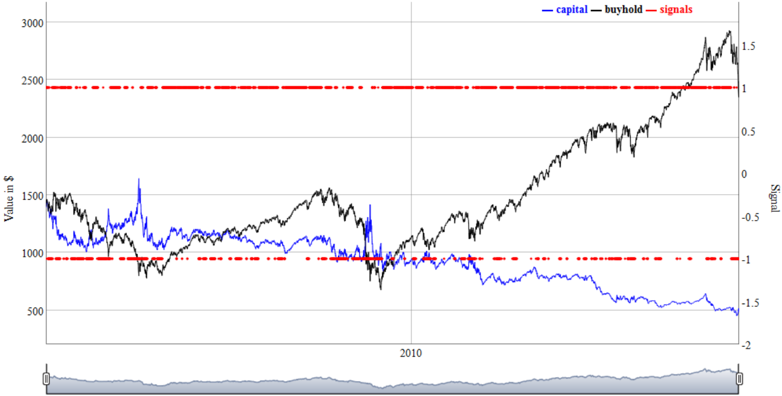
Results. Figure no A-F1. The out-of-sample equity lines and signals for ARIMA model and S&P500 index

The model predicted the trajectory of very strong move downward

Results. Figure no A-F2. The out-of-sample equity lines and signals for Reversal model and S&P500 index

The same case as with ARIMA but the performance during 2007-2009 was much better.

Results. Figure no A-F3. The out-of-sample equity lines and signals for Continuation model and S&P500 index



The continuation system did not work.

Results. Figure no A-F4. The out-of-sample equity lines and signals for MovingAverage model and S&P500 index

MAs approach produced better results which were not affected by global financial crisis.

Results. Figure no A-F5. The out-of-sample equity lines and signals for Unemployment model and S&P500 index

One of the best approaches but mainly because of using adequate macro indicator.

Results. Figure no A-F6. The out-of-sample equity lines and signals for VolatilityBreakout model and S&P500 index

Once again, the example of simple investment technique no affected by the direction of the current trand.

Results. Table no A-T1. The performance statistics for 6 tested strategies, the benchmark, and the combined model

finName	ARC	IR	aSD	MD	AMD	MLD	allRisk	ARCMD	ARCAMD	numbTrans	stopSignal
buyhold	**2.77**	**0.1445**	**19.1643**	**64.33**	**16.90**	**7.1548**	**14.9070**	**0.0431**	**0.1639**	**1**	**0**
arima	1.10	0.0752	14.6264	49.34	12.36	9.6825	8.6366	0.0223	0.0890	1497	2552
reversal	5.65	0.2961	19.0799	60.18	15.76	9.4563	17.1122	0.0939	0.3585	2533	3
continuation	-5.55	-0.2894	19.1746	79.39	19.27	16.4206	48.1685	-0.0699	-0.2880	388	0
movingaverage	4.90	0.2623	18.6838	38.69	15.17	5.2421	5.7485	0.1266	0.3230	33	195
unemployment	7.10	0.4693	15.1295	31.54	11.02	6.1468	3.2323	0.2251	0.6443	378	2402
volbreakout	2.64	0.1817	14.5273	48.42	13.19	7.4960	6.9548	0.0545	0.2002	32	1329
combined	**3.96**	**0.4690**	**8.4428**	**16.14**	**6.86**	**6.0159**	**0.5624**	**0.2454**	**0.5773**	**3209**	**579**

The combination of signals form simple methods shows the potential of using of N various preferably not correlated investment methods.

The main emphasis should be put on risk.

Results. Figure no A-F7. The out-of-sample equity lines and signals for the combined model and S&P500 index

The main advantage reveals at the time of market turmoils.

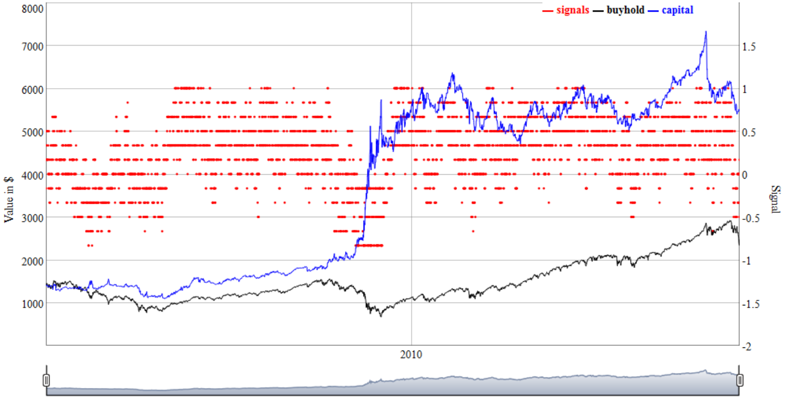
Results. Table no A-T2. The performance statistics for 6 tested strategies, the benchmark, and the combined model with and without leverage

finName	ARC	IR	aSD	MD	AMD	MLD	allRisk	ARCMD	ARCAMD	numbTrans	stopSignal
buyhold	**2.77**	**0.1445**	**19.1643**	**64.33**	**16.90**	**7.1548**	**14.9070**	**0.0431**	**0.1639**	**1**	**0**
arima	1.10	0.0752	14.6264	49.34	12.36	9.6825	8.6366	0.0223	0.0890	1497	2552
reversal	5.65	0.2961	19.0799	60.18	15.76	9.4563	17.1122	0.0939	0.3585	2533	3
continuation	-5.55	-0.2894	19.1746	79.39	19.27	16.4206	48.1685	-0.0699	-0.2880	388	0
movingaverage	4.90	0.2623	18.6838	38.69	15.17	5.2421	5.7485	0.1266	0.3230	33	195
unemployment	7.10	0.4693	15.1295	31.54	11.02	6.1468	3.2323	0.2251	0.6443	378	2402
volbreakout	2.64	0.1817	14.5273	48.42	13.19	7.4960	6.9548	0.0545	0.2002	32	1329
combined	**3.96**	**0.4690**	**8.4428**	**16.14**	**6.86**	**6.0159**	**0.5624**	**0.2454**	**0.5773**	**3209**	**579**
combined lev2	**7.20**	**0.4275**	**16.8437**	**30.92**	**13.49**	**6.3810**	**4.4831**	**0.2329**	**0.5337**	**3209**	**579**

The leverage on the level of 200%.

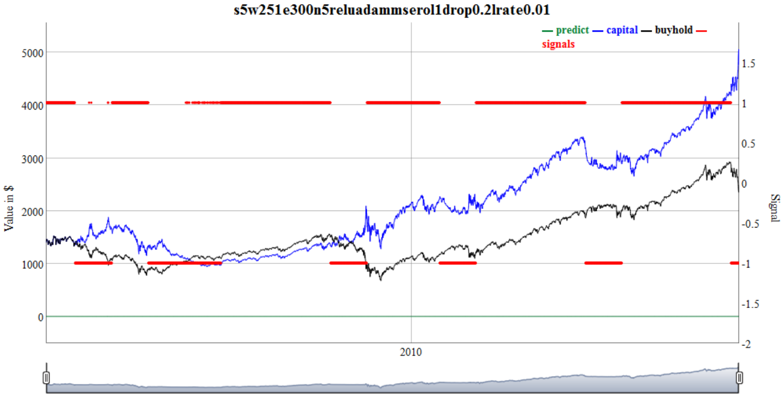
Diversified investment through combined signals from various techniques can be leveraged and can obtain risk-adjusted returns still better than benchmark investment.

Results. Figure no A-F8. The out-of-sample equity lines and signals for the combined model with leverage and S&P500 index

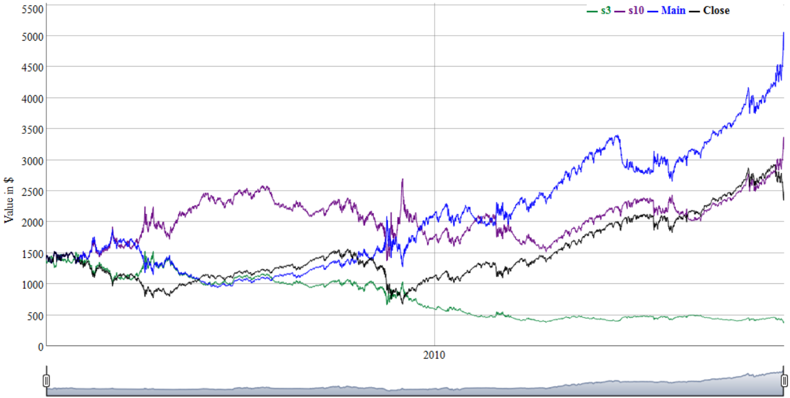


LSTM forecasting methods

Results. Figure no B-F1. Core assumption

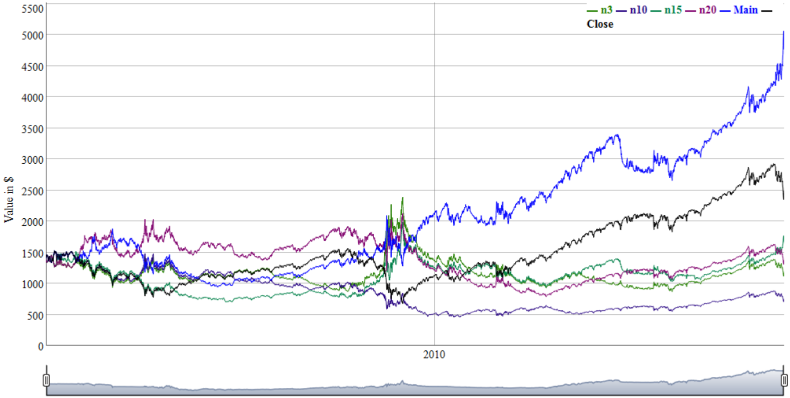


Are these results robust to various hyperparameters assumed at the beginning?



It seems that we selected the best sequence at the beginning. The selection was based on the literature review and practitioners recommendation.

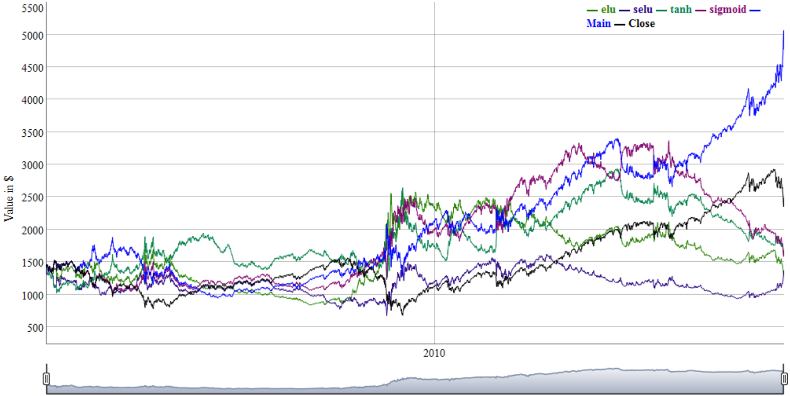
Resuls. Figure no B-F3. Units



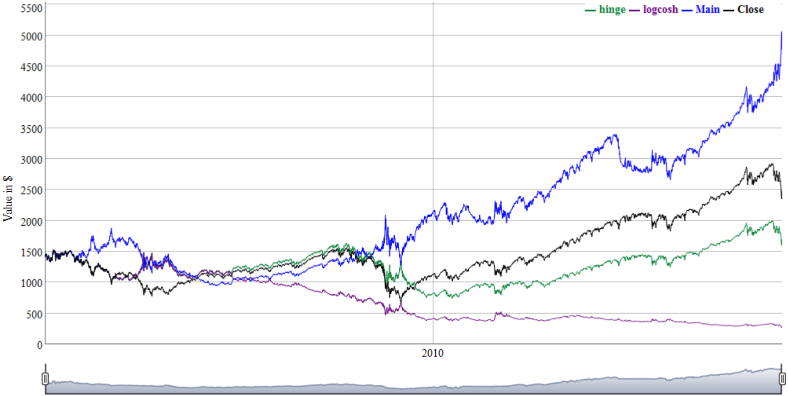
Once again, we selected the best sequence at the beginning. The selection was based on practitioners recommendation.

$$units = (window - sequence)/(9 * (sequence + theoutputlength))$$

Results. Figure no B-F4. Activation

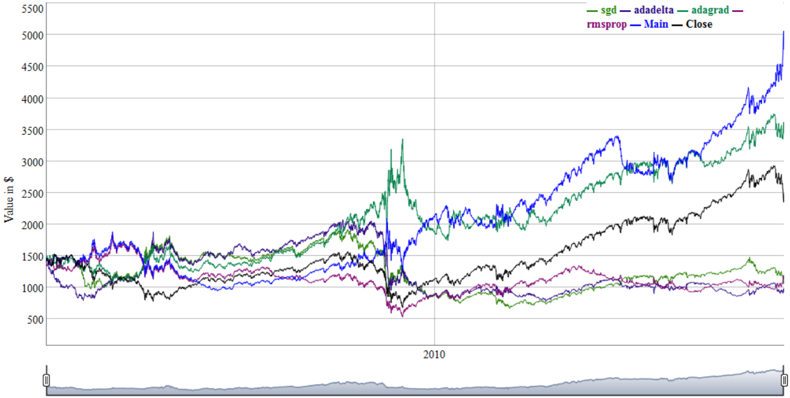


not robust to activation



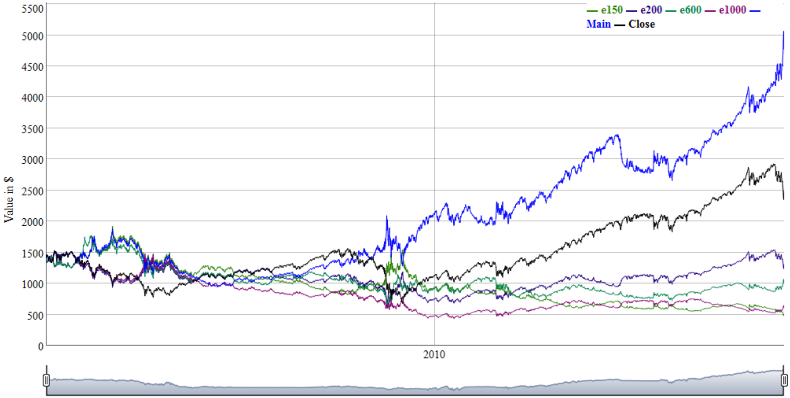
not robust to loss

Results. Figure B-F6. Optimizer



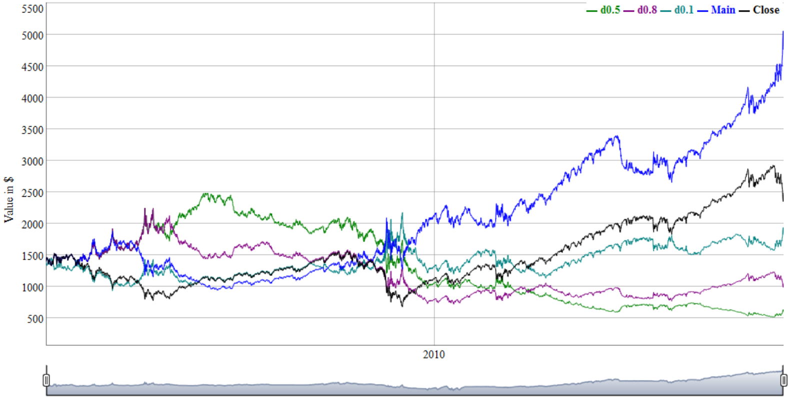
not robust to optimizer

Results. Figure B-F7. Epochs



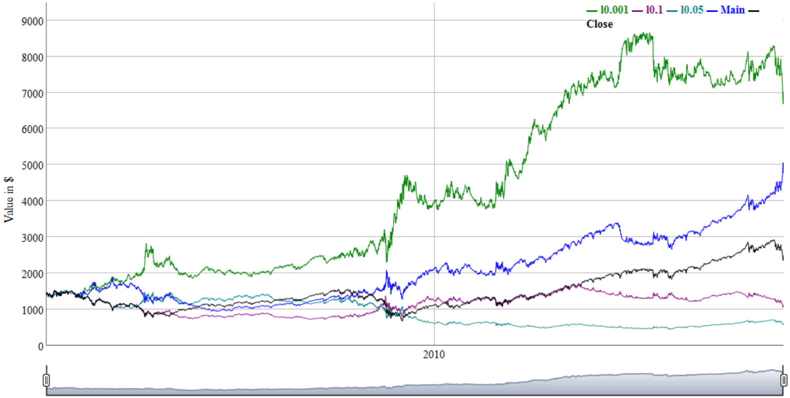
not robust to epochs

Results. Figure B-F8. DropOut



not robust to dropout rate

Results. Figure B-F9. LearningRate



not robust to learning rate

Results. Table no B-T1. The performance statistics for LSTM and the benchmark

finName	ARC	IR	aSD	MD	AMD	MLD	allRisk	ARCMD	ARCAMD	numbTrans	stopSignal
buyhold	**2.77**	**0.1445**	**19.1643**	**64.33**	**16.90**	**7.1548**	**14.9070**	**0.0431**	**0.1639**	**1**	**0**
s5w251e300n5reluadammserol1drop0.2lrate0.01	6.48	0.3388	19.1237	52.66	15.44	7.0357	10.9397	0.1231	0.4197	103	0
s3w251e300n5reluadammserol1drop0.2lrate0.01	-6.65	-0.3469	19.1690	82.25	18.91	16.4206	48.9571	-0.0809	-0.3517	47	0
s10w251e300n5reluadammserol1drop0.2lrate0.01	4.33	0.2260	19.1577	51.65	16.06	8.8532	14.0689	0.0838	0.2696	63	0
s5w251e300n3reluadammserol1drop0.2lrate0.01	-0.90	-0.0471	19.1173	67.27	17.60	9.8016	22.1849	-0.0134	-0.0511	101	0
s5w251e300n10reluadammserol1drop0.2lrate0.01	-3.31	-0.1726	19.1799	76.67	18.55	18.7302	51.0926	-0.0432	-0.1784	66	0
s5w251e300n15reluadammserol1drop0.2lrate0.01	0.92	0.0480	19.1606	57.31	17.82	9.8016	19.1798	0.0161	0.0516	370	0
s5w251e300n20reluadammserol1drop0.2lrate0.01	-0.14	-0.0073	19.1534	69.57	17.09	9.8016	22.3207	-0.0020	-0.0082	99	0
s5w251e300n5eluadammserol1drop0.2lrate0.01	-0.20	-0.0105	19.1292	56.75	17.35	9.4722	17.8407	-0.0035	-0.0115	518	0
s5w251e300n5seluadammserol1drop0.2lrate0.01	-0.15	-0.0079	19.0884	59.24	18.14	9.0317	18.5264	-0.0025	-0.0083	682	0
s5w251e300n5tanhadammserol1drop0.2lrate0.01	0.92	0.0481	19.1353	47.56	17.57	4.8294	7.7222	0.0193	0.0524	526	0
s5w251e300n5sigmoidadammserol1drop0.2lrate0.01	0.95	0.0497	19.0988	54.53	18.20	8.1151	15.3818	0.0174	0.0522	380	0
s5w251e300n5reluadamhingerol1drop0.2lrate0.01	1.01	0.0527	19.1784	61.77	17.05	9.5635	19.3166	0.0164	0.0592	98	0
s5w251e300n5reluadamlogcoshrol1drop0.2lrate0.01	-8.40	-0.4385	19.1571	87.48	20.26	18.7302	63.5946	-0.0960	-0.4146	12	0

Results. Table no B-T2. The performance statistics for LSTM and the benchmark

	finName	ARC	IR	aSD	MD	AMD	MLD	allRisk	ARCMD	ARCAMD	numbTrans	stopSignal
1	buyhold	2.77	0.1445	19.1643	64.33	16.90	7.1548	14.9070	0.0431	0.1639	1	0
14	s5w251e300n5relusgdmserol1drop0.2lrate0.01	-1.19	-0.0620	19.1808	70.78	17.81	11.2103	27.1056	-0.0168	-0.0668	434	0
15	s5w251e300n5reluadadeltamserol1drop0.2lrate0.01	-2.19	-0.1140	19.2188	68.86	19.18	11.3333	28.7672	-0.0318	-0.1142	542	0
16	s5w251e300n5reluadagradmserol1drop0.2lrate0.01	4.78	0.2498	19.1368	49.59	15.13	8.8214	12.6660	0.0964	0.3159	504	0
17	s5w251e300n5relurmspropmserol1drop0.2lrate0.01	-1.16	-0.0606	19.1472	76.03	18.18	17.2500	45.6534	-0.0153	-0.0638	14	0
18	s5w251e150n5reluadammserol1drop0.2lrate0.01	-5.34	-0.2790	19.1422	81.58	19.05	17.2500	51.3168	-0.0655	-0.2803	26	0
19	s5w251e200n5reluadammserol1drop0.2lrate0.01	-0.35	-0.0183	19.1724	63.77	17.84	17.8532	38.9407	-0.0055	-0.0196	45	0
20	s5w251e600n5reluadammserol1drop0.2lrate0.01	-1.73	-0.0905	19.1201	69.58	17.95	17.7937	42.4918	-0.0249	-0.0964	328	0
21	s5w251e1000n5reluadammserol1drop0.2lrate0.01	-4.37	-0.2277	19.1923	77.83	18.50	18.7302	51.7593	-0.0561	-0.2362	9	0
22	s5w251e300n5reluadammserol1drop0.5lrate0.01	-4.47	-0.2333	19.1630	83.79	18.00	14.8651	42.9631	-0.0533	-0.2483	9	0
23	s5w251e300n5reluadammserol1drop0.8lrate0.01	-1.53	-0.0798	19.1789	73.80	17.65	16.4206	41.0217	-0.0207	-0.0867	10	0
24	s5w251e300n5reluadammserol1drop0.1lrate0.01	1.42	0.0740	19.1790	51.80	16.66	9.8016	16.2229	0.0274	0.0852	12	0
25	s5w251e300n5reluadammserol1drop0.2lrate0.001	8.46	0.4434	19.0787	36.20	13.75	5.7500	5.4604	0.2337	0.6153	583	0
26	s5w251e300n5reluadammserol1drop0.2lrate0.1	-1.17	-0.0612	19.1157	57.82	18.92	12.9405	27.0608	-0.0202	-0.0618	92	0
27	s5w251e300n5reluadammserol1drop0.2lrate0.05	-4.40	-0.2295	19.1741	78.73	18.17	18.7302	51.3751	-0.0559	-0.2422	50	0

Technical slide-additional research question

The estimation time for each model

- MAs/Continuation/Reversal/Macro/VolBreakout: less than 1s for ALL
- ARIMA: 0.002s for 1 out 108m*5000d = 540000
- LSTM model: 10minutes for 1 out of 27model

Which parameter affect the calculation time the most (LSTM vs classical strategies)?

- Classical: the number of ARIMA models
- LSTM: the number of Epochs

Results. LSTM. What could went wrong?

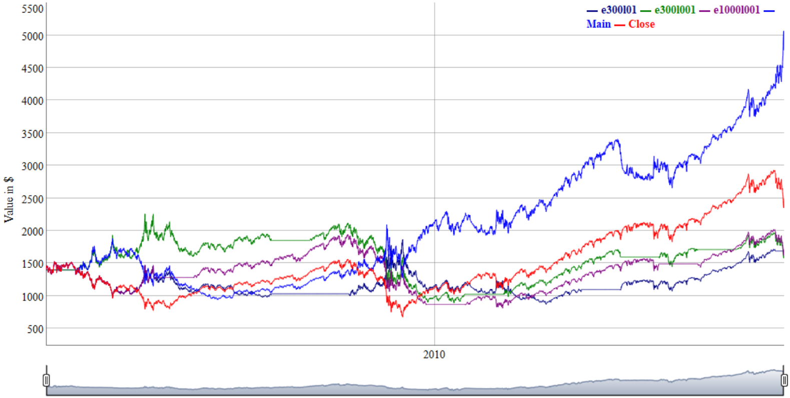
the possible overoptimistion during LSTM fitting

- the solution:
  - **Regularization (L1, L2)**
  - Random search
  - Early stopping

the structure of the model

- the solution:
  - more layers
  - more units
  - different type of LSTM

Results. Figure B-F10. Regularization



not affected by additional regularization

Summary

1. There is no any basis to reject the first hypothesis concerning the classical methods
2. There is no any basis to reject the second hypothesis concerning LSTM model
3. The out-of-sample results suggest that once again we do not have any basis to reject the third hypothesis
4. The combined results for 6 classical methods do not allow to reject this hypothesis
5. our initial results unable to strongly refer to this hypothesis

Research extensions

- further investigation of hidden parameters in RNN/LSTM
- alternative ways of setting of initial values of hyperparameters
- identifying which parameters are responsible for possible ovefitting
- comparing daily with HF results
- preparing forecast simultaneously for the set of equity indices

References I

- Hochreiter, S.; Schmidhuber, J. (1997). "Long Short-Term Memory". Neural Computation. 9 (8): 1735–1780
- Gers, F. A.; Schmidhuber, J. (2001). "LSTM Recurrent Networks Learn Simple Context Free and Context Sensitive Languages", IEEE Transactions on Neural Networks. 12 (6): 1333–1340
- Felix A. Gers; Jürgen Schmidhuber; Fred Cummins (2000). "Learning to Forget: Continual Prediction with LSTM". Neural Computation. 12 (10): 2451–2471
- Klaus Greff; Rupesh Kumar Srivastava; Jan Koutník; Bas R. Steunebrink; Jürgen Schmidhuber (2015). "LSTM: A Search Space Odyssey". IEEE Transactions on Neural Networks and Learning Systems. 28 (10): 2222–2232.

- Cho, Kyunghyun; van Merrienboer, Bart; Gulcehre, Caglar; Bahdanau, Dzmitry; Bougares, Fethi; Schwenk, Holger; Bengio, Yoshua (2014). "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation"
- Zhang, G.P., 2003. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, pp.159-175.
- Grinblatt, M., Titman, S. and Wermers, R., 1995. Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior. The American economic review, pp.1088-1105.
- Bollinger, J., 2002. Bollinger on Bollinger bands. McGraw Hill Professional.
- Campbell, R., 2008. Art as a financial investment. The Journal of Alternative Investments, 10(4), pp.64-81.

Thank you for your attention !

Mateusz Kijewski, mj.kijewski@student.uw.edu.pl

and

Robert Ślepaczuk, rslepaczuk@wne.uw.edu.pl,

Faculty of Economic Sciences,

University of Warsaw,

Department of Quantitative Finance

Quantiative Finance Research Group