The Memory of Stock Volatility

CS 230 Final Project Report

Project Category: Finance

Video: `https://youtu.be/ql4qZ6oZh3U`

Simona Diaconu, sdiaconu@stanford.edu

Introduction: Financial markets play a crucial role not only in the life of those who make a career of analyzing this field, but also in the unfolding of society given that many institutions such as private universities derive a substantial part of their resources from participating in trades. Some of the most widely used assets in this universe are stocks, and the gargantuan collective effort put into understanding their internal logic has been translated in a plethora of models attempting to describe their behavior. One obviously essential element in this respect is price; however, there are other hidden traits that have a say when it comes down to what truly matters, profits, one of them being volatility (viz. the variance of the prices) which Henri Markowitz used explicitly in the portfolio selection problem he formulated and which has been carefully considered over and over again; this feature plays a role not only in stock price modeling, but also as an asset itself (there are financial derivatives based on it). Moreover, since volatility is not observable, this underlying statistic is especially complex as there is no benchmark one could compare an estimator of it to. In this paper, we aim to investigate how well this feature, once it has attached to it a clear-cut definition, can be predicted using RNNs, and how their performance compares to that of ARIMA.

Related Work: In the existing literature, there have been several papers addressing stock prediction using LSTMs together with sentiment analysis ([1],[3],[5]), and the movement of volatility, either up or down ([4]). The first line of work is particularly interesting as it combines natural language processing (computing sentiment scores) and numerical data analysis; however, it has the shortcoming of being concerned only with predicting the direction of the price changes, not their sizes or anything quantitative that is non-binary as up/down movements are. In addition, one CS 230 final project in 2019 [6] has addressed a similar question, predicting volatility (defined as the standard deviation over a past fixed-size time window) with LSTMs. Another relatively recent paper in 2018 considers prediction of volatility with LSTMs, although it is not concerned with prediction as only errors on the training set are reported. Up to an extent, in this paper, we attempt to bring together stochastic calculus and machine learning and see if we can get the best of both worlds, the models that have a theoretical support and the computational tools that can tackle complex data structures, when it comes to predicting future volatilities.

Methods: The definitions employed for volatility are given by four stochastic models, Parkinson (P), Garman-Klass (GK), Rogers-Satchell (RS), and Yang-Zhang (YZ) referenced in [4], and they use the close (C), open (O), high (H), and low (L) prices for a window of w days:

$$P = \frac{1}{4w \log 2} \sum_{1 \leq i \leq w} (\log \frac{H_i}{L_i})^2,$$

$$GK = \frac{1}{w} \sum_{1 \leq i \leq w} [\frac{1}{2}(\log \frac{H_i}{L_i})^2 - (2\log 2 - 1)(\log \frac{C_i}{O_i})^2],$$

$$RS = \frac{1}{w} \sum_{1 \leq i \leq w} [\log \frac{H_i}{L_i} \log \frac{H_i}{C_i} - \log \frac{L_i}{O_i} \log \frac{L_i}{C_i}],$$

$$YZ = \frac{1}{w-1} \sum_{1 \leq i \leq w} [(\log \frac{O_i}{C_{i-1}})^2 - (\overline{\log \frac{O_i}{C_{i-1}}})^2] + \frac{k}{w-1} \sum_{1 \leq i \leq w} [(\log \frac{O_i}{C_i})^2 - (\overline{\log \frac{O_i}{C_i}})^2] +$$
$$+ \frac{1-k}{w} \sum_{1 \leq i \leq w} [\log \frac{H_i}{L_i} \log \frac{H_i}{C_i} - \log \frac{L_i}{O_i} \log \frac{L_i}{C_i}],$$

where

$$k = \frac{0.17}{1.17 + \frac{1}{w-1}}, \quad \overline{(Z_i)} = \frac{1}{w} \sum_{1 \leq i \leq w} Z_i.$$

Several metrics are computed both for the training and test set (for the ARIMA model, there is no prediction for the training set, only for the test set): root-mean square error (RMSE), correlation ($corr$), R^2-statistic (R^2), and mean absolute error (MAE):

$$RMSE((x_i),(\hat{x}_i)) = \sqrt{\frac{1}{n}\sum_{1\leq i\leq n}(\hat{x}_i - x_i)^2},$$

$$corr((x_i),(\hat{x}_i)) = \frac{\frac{1}{n}\sum_{1\leq i\leq n} x_i\hat{x}_i - (\frac{1}{n}\sum_{1\leq i\leq n} x_i)(\frac{1}{n}\sum_{1\leq i\leq n}\hat{x}_i)}{\sqrt{\frac{1}{n}\sum_{1\leq i\leq n}(x_i - \overline{(x_i)})^2 \cdot \frac{1}{n}\sum_{1\leq i\leq n}(\hat{x}_i - \overline{(\hat{x}_i)})^2}},$$

$$R^2((x_i),(\hat{x}_i)) = 1 - \frac{\sum_{1\leq i\leq n}(x_i - \hat{x}_i)^2}{\sum_{1\leq i\leq n}(x_i - \overline{(x_i)})^2},$$

$$MAE((x_i),(\hat{x}_i)) = \frac{1}{n}\sum_{1\leq i\leq n}|x_i - \hat{x}_i|.$$

From the features (close, open, high, and low prices), for each stock, the four estimators above are calculated with $w \in \{10, 20, 40\}$, and two models are trained on each of them: (1) an ARIMA architecture yielding predictions iteratively for one year of data using the previous 9 years; (2) an RNN network with one LSTM layer with 10 units, a dense layer trained on 9 years of data for 100 epochs, with loss the mean squared error, optimizer the RMS prop and used to predict one year of new volatilities (we tried increasing the number of units to 20 and the number of epochs to 1000 but no considerable change emerged in any case). As stock prices have been observed to show time dependency, it is not wildly unreasonable to assume that this vestige is passed on to their volatilities as well: hence, LSTM layers are appropriate for this problem as they are designed to take into consideration past values when predicting current and future quantities. Moreover, since all the volatility estimates have incorporated explicitly past prices (given their formulae), an LSTM architecture seems to be a good fit for the task at hand inasmuch as it will encompass for each prediction a relatively long period of the past even when the layers have few units.

Each experiment is run on two different time periods, the first starting at the beginning of the dataset, and the second starting a quarter in (that is, a 5-year shift): this is for getting a sense of how sensitive the performance of these algorithms is to changes in data. As regards Python packages, ARIMA in Arch and RNNs in Tensor Flow are employed.

Dataset: The dataset comes from Yahoo Finance, it contains the four previously mentioned features for each trading day in the period May 7, 2001- May 7, 2021, and four stocks are considered: JP Morgan, Apple, Microsoft, and Wells Fargo. Here is an example of the some of the datapoints for JP Morgan:

	Date	Open	High	Low	Close
1.	5/8/01	48.75	48.98	47.5	47.75
2.	5/9/01	47.25	47.610001	47	47.25
3.	5/10/01	48	48.919998	47.259998	47.509998

Discussion: Of the two models, ARIMA does considerably worse on the test set according to all the metrics (table 1): its performance is inferior for the second period compared to the first measured by RMSE and MAE but better by correlations and R^2-statistics. The RNN results on the training set (table 2) are comparable to what is reported in [2] in which the authors analyze how LSTMs perform on stock volatility prediction for the Yang-Zhang model using SP 500 daily data but do not consider out-of-sample predictions (that is, there is no test set), while on the test set (table 3), the errors are considerably better than its counterparts on the training set. As regards the change of time period, the performance differs considerably from one case to another, the second interval having larger errors (for correlation and the R^2 statistic up to 5 percent, and for RMSE and MAE up to 8 times) (table 3) for JP Morgan. This may be partially explained by the stock crisis that hit in 2007 : the training sets are roughly $2001-2009$ and $2006-2015$, respectively, and as the side effects crisis lingered for a while, they were more prominently felt by the second period in the financial sector: Wells Fargo presents a more ambiguous performance than JP Morgan, with slight improvements on RMSE and MAE, and worse correlations and R^2-statistics for the second period relative to the first. On the other hand, Apple and Microsoft do better in second period, which might be related to the fact that these are stocks from another sector, technology (see appendix for results on these three stocks).

Concerning the length of the time window, increasing it seems to have a beneficial effect on performance, the errors going down (tables $4(a), 4(b)$) (this holds for the four stocks).

Table 1: ARIMA test errors (P, GK, RS, YZ), $w = 10$, first (second) period, JP Morgan

RMSE	correlation	R^2-statistic	MAE
0.000138896714653 (0.000399860664917)	-0.38023921247 (0.0090096331)	-9.883192913291 (-0.06427963151026)	0.000437221079394 (0.0001380848025840)
0.0001221750671443 (0.0005289434643794)	-0.587140334576 (0.01834367484)	-25.978570390860 (-0.05352108591374)	0.000614776959617 (0.000163756323786)
0.0002719597364308 (0.0008899391495311)	-0.4383365155382 (-0.001468869462)	-19.45684156753 (-0.082570513027)	0.00119520120210 (0.000297319480149)
0.000290858520090 (0.0009217078594769)	-0.463736463610 (-0.002963698393)	-16.292700769788 (-0.12107144441193)	0.001170941031316 (0.000348903683372)

Table 2: RNN training errors (P, GK, RS, YZ), $w = 10$, first (second) period, JP Morgan

RMSE	correlation	R^2-statistic	MAE
0.001569475531578 (0.001490951180458)	9.6913084e-08 (-2.4143125e-07)	-4300.99951171875 (-4384.775390625)	0.000801385343074 (0.000706550002098)
0.001464903116226 (0.00142293143272)	1.3903849e-07 (-5.8886844e-08)	-4310.48046875 (-4603.28564453125)	0.000763883352279 (0.000679979085922)
0.002886225223541 (0.002694947719573)	1.14768504e-07 (8.2437474e-08)	-4219.35400390625 (-4128.3330078125)	0.001481869816780 (0.00126777768135)
0.003169392585754 (0.003002204179763)	-1.18762316e-07 (-5.293989e-08)	-4349.94189453125 (-4357.4921875)	0.001627562403678 (0.001443833231925)

Table 3: RNN test errors (P, GK, RS, YZ), $w = 10$, first (second) period, JP Morgan

RMSE	correlation	R^2-statistic	MAE
2.9871402450625894e-05 (0.000168608893356985)	0.9764202790603 (0.9111002587423)	0.9531214735141 (0.8221963429520)	1.7289284683495114e-05 (2.6573123484340803e-05)
2.7419776164590857e-05 (0.000232185442252155)	0.9741555429920 (0.9036587064008)	0.9478093351244 (0.8073170216287)	1.6776309585179432e-05 (3.1504082858019005e-05)
7.494881599252321e-05 (0.000370163431482547)	0.9609172529842 (0.9134948227670)	0.9227489578051 (0.8269917404279)	4.385968096871917e-05 (5.9082861426624146e-05)
6.716518897912952e-05 (0.000360623996127868)	0.972840617208 (0.923455059538)	0.94576006743 (0.846918275186)	4.066833764115236e-05 (6.487198522799058e-05)

Table 4(a): RNN test errors (P, GK, RS, YZ), $w = 20$, first (second) period, JP Morgan

RMSE	correlation	R^2-statistic	MAE
1.6051567306354768e-05 (8.464074033808171e-05)	0.9918135585757 (0.9545394643097)	0.983138988804 (0.9090834846875)	9.10170126232559e-06 (1.3587044651633604e-05)
1.5996733329449042e-05 (0.000116394221444963)	0.9894819680486 (0.9504669869623)	0.9784725143657 (0.900938695243)	8.983509025947642e-06 (1.5747782661426503e-05)
3.805522912847723e-05 (0.000185542609031061)	0.9875035357707 (0.9564559796402)	0.9747233172151 (0.912918193008)	2.0452460530511438e-05 (2.9453004277738026e-05)
3.56016983458707e-05 (0.000181684725464094)	0.9911476763513 (0.9621684935149)	0.9818142579777 (0.924347867417)	1.9361064919315088e-05 (3.3298134222218916e-05)

Table 4(b): RNN test errors (P, GK, RS, YZ), $w = 40$, first (second) period, JP Morgan			
RMSE	correlation	R^2-statistic	MAE
7.779480252342487e-06 (1.6139930349560716e-05)	0.9971845258802 (0.9975971936109)	0.994049667941 (0.9951962885111)	4.614656754220552e-06 (9.410462116071535e-06)
7.454549236144284e-06 (1.6081702908487252e-05)	0.9966905281983 (0.9973745476543)	0.9930515602915 (0.9947517502047)	4.429353687178629e-06 (9.112288688098619e-06)
1.9205958325280503e-05 (3.999711065749486e-05)	0.9957358367062 (0.9965125291035)	0.9911033999084 (0.9930285735031)	1.0907998995031158e-05 (2.0602277033364883e-05)
1.808351666383734e-05 (4.756715651803942e-05)	0.9969334403185 (0.996660030330)	0.9934959019258 (0.9933188234858)	1.0645298748302876e-05 (2.5377567106208427e-05)

Conclusion: RNNs with LSTM layers do a great job at predicting stock volatility once it is properly defined: the high correlation and R^2 statistic values obtained on the test sets suggest that this architecture learns quite well from the training data in spite of performing more poorly on the latter than on the former. This can be at least in part explained by the time memory that financial assets have. However, their learning depends also on the stability of the market, the predictions getting worse when the data fed into it comes from assets trapped in a crisis. In the future, it would be worthwhile to see if these architectures can be used in other applications of financial mathematics such as option pricing: the famous Black-Scholes formula which yields prices for financial derivatives (assets whose value is contingent on other objects such as stocks) requires an estimate for volatility and the quandary of which one to use emerges anew; historical, implied, and calibrated volatilities ([7]) are considered in the literature, and the differences between the theoretical prices given by Black-Scholes and the observed ones are modeled in an attempt to predict future prices of options. Thus, the question of whether or not LSTMs can be employed to get better volatility estimators and so simplify the errors between actual and theoretical prices is worth considering.

References:

1. *Investigation Into The Effectiveness Of Long Short Term Memory Networks For Stock Price Prediction*, Hengjian Jia, 2016, `https://arxiv.org/pdf/1603.07893.pdf`.

2. *Forecasting Volatility with LSTM Techniques*, Hemanth Kumar P, S. Basavaraj Patil, International Journal of Science and Research (IJSR), Volume 7, Issue 10, October 2018.

3. *Stock Volatility Prediction Using Recurrent Neural Networks with Sentiment Analysis*, Yifan Liu, Zengchang Qin, Pengyu Li, and Tao Wan, 2017, `https://arxiv.org/pdf/1705.02447.pdf`.

4. *Exploring the predictability of range-based volatility estimators using RNNs*, Gábor Petneházi, and József Gáll, 2018, `https://arxiv.org/pdf/1803.07152.pdf`.

5. *Stock Market Trend Prediction Using Regression Model, RNNs, and Sentiment Analysis*, Rajkumar Sujatha, Varad Abhyankar, Abhishek Gehlot, Priyank Gupta, and Sivakumar Subramaniam, *Advances in Automation, Signal Processing, Instrumentation, and Control, Lecture Notes in Electrical Engineering 700*, 2021.

6. *Stock Price Volatility Prediction with Long ShortTerm Memory Neural Networks*, Jason Sullivan, 2019, `http://cs230.stanford.edu/projects_fall_2019/reports/26254244.pdf`.

7. *Option Prices And Pricing Theory: Combining Financial Mathematics with Statistical Modeling*, Ling Chen, Tze Leung Lai, and Tiong Wee Lim, John Wiley and Sons, Inc. Volume 3, 2011.

Appendix

Apple, RNN test errors (P, GK, RS, YZ), $w = 10$, first (second) period			
RMSE	correlation	R^2-statistic	MAE
0.00015296393470745748 (3.21455808866939e-05)	0.8310907735164114 (0.9664077969738177)	0.6837874590062526 (0.932906921443444)	2.566770450707418e-05 (1.8359907236921434e-05)
0.00020945726974648873 (3.4267631239963867e-05)	0.815654937621129 (0.9647416586680734)	0.6576253195381913 (0.9295646877685392)	2.975137632675967e-05 (1.7560073579756076e-05)
0.0004014607588336153 (7.086504595324285e-05)	0.8170347735609803 (0.9632001660549172)	0.6606354258264384 (0.926455417906974)	6.323274893829923e-05 (3.930525603622243e-05)
0.00034938712130966657 (0.00010800323655755142)	0.8506968103465676 (0.9471171527229814)	0.7151316859796639 (0.8943460442845458)	6.690088561035922e-05 (5.15877498765938e-05)

Apple, RNN test errors (P, GK, RS, YZ), $w = 20$, first (second) period			
RMSE	correlation	R^2-statistic	MAE
7.706467172471237e-05 (1.6386615422912926e-05)	0.9570295330119716 (0.9860112711559085)	0.9148046794155876 (0.9720629223683157)	1.3194785269859639e-05 (9.256352574404248e-06)
0.00010518744312536211 (1.776276640521462e-05)	0.9527191624443838 (0.9848123583417356)	0.9066307959007217 (0.9696606479724272)	1.5157771277715867e-05 (9.08535134223172e-06)
0.00020175127784667285 (3.65676866247253e-05)	0.9518217309359356 (0.9830396698817991)	0.9050258523284959 (0.9660615371662723)	3.240095557288722e-05 (2.020856319926725e-05)
0.00017582563743721833 (5.2487931377917825e-05)	0.9612901275935698 (0.9800808257645233)	0.922952425622331 (0.9599136890634719)	3.488073731877615e-05 (2.6844760275118668e-05)

Apple, RNN test errors (P, GK, RS, YZ), $w = 40$, first (second) period			
RMSE	correlation	R^2-statistic	MAE
3.841285648872765e-05 (7.771663979834723e-06)	0.9832913574110245 (0.9948379240142832)	0.9664755091347056 (0.9895567128850541)	6.359621957037862e-06 (4.711645165350415e-06)
5.251158230822344e-05 (8.491342889277314e-06)	0.9806191440172874 (0.9939278237470669)	0.9612441720535386 (0.9877959162643446)	7.236372891181965e-06 (4.421192912326988e-06)
0.00010058224682034783 (1.6576761497869006e-05)	0.9811073805007401 (0.9930615771095084)	0.9622155704207552 (0.9859601868660183)	1.5337961925466065e-05 (9.389408460004402e-06)
8.785014335200718e-05 (2.5307649744776252e-05)	0.9847841537235412 (0.9906437042037559)	0.9693444356201574 (0.9810455834149379)	1.655118665013031e-05 (1.285676094225435e-05)

Microsoft, RNN test errors (P, GK, RS, YZ), $w = 10$, first (second) period			
RMSE	correlation	R^2-statistic	MAE
3.564101093685655e-05 (2.1156847603287777e-05)	0.9387976057814746 (0.9804076635377423)	0.8791375186174435 (0.9608184295692365)	1.4982181333775427e-05 (1.2277842937357762e-05)
4.20402082431236e-05 (2.2391778739696683e-05)	0.9330715842282864 (0.9809422494340229)	0.8684858019123854 (0.9618888800122376)	1.6572202092263652e-05 (1.3099819011461635e-05)
9.203667449984862e-05 (4.2183934725477865e-05)	0.9353590000013569 (0.9804617378031019)	0.8725919446058247 (0.9609162116376713)	3.5443769026175666e-05 (2.443346104127166e-05)
8.315102345250488e-05 (4.502693648411387e-05)	0.9472464298350614 (0.9793742522701796)	0.895801058317839 (0.9587364242148642)	3.554228697029609e-05 (2.677690864071389e-05)

Microsoft, RNN test errors (P, GK, RS, YZ), $w = 20$, first (second) period			
RMSE	correlation	R^2-statistic	MAE
1.7631381965006673e-05 (1.2241902662507685e-05)	0.9812998649315211 (0.9916026532069072)	0.9627646614182291 (0.983191002644053)	7.4144557686009245e-06 (7.345404484460021e-06)
2.062554017133609e-05 (1.2433121034132148e-05)	0.9790862483629409 (0.9923749365158634)	0.9584222171077915 (0.9847377884190229)	7.900148273776267e-06 (7.242576337934148e-06)
4.447844339164231e-05 (2.4162530274648143e-05)	0.9802797371514285 (0.991038933364416)	0.9607615338295888 (0.9820682736782872)	1.745647314660633e-05 (1.4270128826339437e-05)
4.0309649704787416e-05 (2.5796959450541797e-05)	0.9847675603008997 (0.9907775491525213)	0.9696452195365886 (0.9815525665524261)	1.7378575635708352e-05 (1.5481306285598747e-05)

Microsoft, RNN test errors (P, GK, RS, YZ), $w = 40$, first (second) period			
RMSE	correlation	R^2-statistic	MAE
7.69738285720477e-06 (5.483056472028684e-06)	0.9945702902131267 (0.9977667986902673)	0.9890618450577138 (0.995531388752066)	3.613616713614597e-06 (3.4277332456526724e-06)
9.321059028295646e-06 (5.8899788795790415e-06)	0.9931404474230873 (0.99770483856673189)	0.9861920594489084 (0.9954064874977848)	3.893803264336086e-06 (3.62843252917172e-06)
2.062628083403754e-05 (1.1802090555473796e-05)	0.9934027313732354 (0.996940534358906)	0.9867371400346748 (0.9938775702059605)	8.644497896168358e-06 (7.046562971115997e-06)
1.879765166504271e-05 (1.2978555658828372e-05)	0.9949812868360618 (0.9967582213728111)	0.9898595183261838 (0.9935156806417117)	8.755296093212352e-06 (7.886761581424905e-06)

Wells Fargo, RNN test errors (P, GK, RS, YZ), $w = 10$, first (second) period			
RMSE	correlation	R^2-statistic	MAE
5.571234852050461e-05 (5.561428737194254e-05)	0.9628114120947617 (0.9815047971753333)	0.926452374014095 (0.9630067297553343)	2.8476182417999285e-05 (2.9145855282936162e-05)
6.298514406642468e-05 (5.633045346415101e-05)	0.9469195300673479 (0.9810265785747361)	0.8959021653349755 (0.9620515280440193)	2.6502981185621326e-05 (2.8044330868581642e-05)
0.00012900377409485878 (0.00014071870564404744)	0.9479305300902554 (0.9771238804921317)	0.8979824013183484 (0.9542274327819833)	6.40517720173656e-05 (6.408361415923565e-05)
0.00012442386234835825 (0.00014936067952946085)	0.9567151707656012 (0.97825017033362874)	0.9148211932215526 (0.9564926428377593)	6.377398621251595e-05 (7.160347079123564e-05)

Wells Fargo, RNN test errors (P, GK, RS, YZ), $w = 20$, first (second) period			
RMSE	correlation	R^2-statistic	MAE
2.928234013804138e-05 (3.129489046207015e-05)	0.9853344791545285 (0.991988051117345)	0.9706601632438991 (0.9839720746374188)	1.4855882839592976e-05 (1.6567469947648303e-05)
3.287280935923768e-05 (3.102478351394808e-05)	0.9795037891847588 (0.9922444506394844)	0.9591540268928663 (0.9844884131934233)	1.3900625302331599e-05 (1.6142762075754263e-05)
6.651502802327218e-05 (7.692285220807961e-05)	0.9802105869572063 (0.9900242625229897)	0.9605647958756771 (0.9800435898124928)	3.1323586720696385e-05 (3.6186158928433025e-05)
6.474666796765793e-05 (8.109334297941735e-05)	0.9842279619662253 (0.9912346513910635)	0.9684606044002557 (0.982469062713154)	3.226129587923999e-05 (3.9586057298007705e-05)

Wells Fargo, RNN test errors (P, GK, RS, YZ), $w = 40$, first (second) period			
RMSE	correlation	R^2-statistic	MAE
1.4436224604084768e-05 (1.400688195032641e-05)	0.9934392916085982 (0.9976329545030961)	0.9866930511612085 (0.9952616268565343)	7.446901759417956e-06 (8.015005603045809e-06)
1.6167568030462527e-05 (1.449208988075388e-05)	0.9910972513778039 (0.9975654647167184)	0.9820054614378435 (0.9951298067864365)	6.791704517703714e-06 (7.995308297285032e-06)
3.230807564383198e-05 (3.683806429418567e-05)	0.9921360283013375 (0.9964228280222183)	0.9841023737666983 (0.992838469311333)	1.5748263766578145e-05 (1.7500577989045983e-05)
3.190651924811874e-05 (4.043314830094787e-05)	0.9937627215171356 (0.9968670543622768)	0.9873214413931504 (0.9937286435817938)	1.6581776662940997e-05 (1.996561423134712e-05)