



Portfolio formation with preselection using deep learning from long-term financial data

Article

Accepted Version

Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

Wang, W., Li, W., Zhang, N. and Liu, K. (2019) Portfolio formation with preselection using deep learning from long-term financial data. *Expert Systems with Applications*, 143. 113042. ISSN 0957-4174 doi: <https://doi.org/10.1016/j.eswa.2019.113042> Available at <http://centaur.reading.ac.uk/86775/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1016/j.eswa.2019.113042>

Publisher: Elsevier

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in

the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Title: Portfolio formation with preselection using deep learning from long-term financial data

Author names and affiliations:

Wuyu Wang¹, Weizi Li², Ning Zhang¹, Kecheng Liu²

¹School of Information, Central University of Finance and Economics, 39 South College Road, Haidian District, Beijing, P.R. China 100081

²Informatics Research Centre, Henley Business School, University of Reading, Reading, Berkshire RG6 6UD, UK

Email addresses: cufe_wwy@163.com (Wuyu Wang), weizi.li@henley.ac.uk (Weizi Li), 717212398@qq.com (Ning Zhang), k.liu@reading.ac.uk (Kecheng Liu)

Portfolio formation with preselection using deep learning from long-term financial data

Abstract: Portfolio theory is an important foundation for portfolio management which is a well-studied subject yet not fully conquered territory. This paper proposes a mixed method consisting of long short-term memory networks and mean-variance model for optimal portfolio formation in conjunction to the asset preselection, in which long-term dependences of financial time-series data can be captured. The experiment uses a large volume of sample data from UK Stock Exchange 100 Index between March 1994 and March 2019. In the first stage, long short-term memory networks are used to forecast the return of assets and select assets with higher potential returns. After comparing the outcomes of the long short-term memory networks against support vector machine, random forest, deep neural networks and autoregressive integrated moving average model, we discover that long short-term memory networks are appropriate for financial time-series forecasting, to beat the other benchmark models by a very clear margin. In the second stage, based on selected assets with higher returns, the mean-variance model is applied for portfolio optimisation. The validation of this methodology is carried out by comparing the proposed model with other five baseline strategies, to which the proposed model clearly outperforms others in terms of the cumulative return per year, Sharpe ratio per triennium as well as average return to the risk per month of each triennium. i.e. potential returns and risks.

Key words: asset preselection, long-term financial data, financial forecasting, portfolio optimisation

22 **Portfolio formation with preselection using deep learning from long-** 23 **term financial data**

24 **1. Introduction**

25 Portfolio management is a decision-making process in which an amount of fund is allocated
26 to multiple financial assets, and the allocation weight is constantly changed in order to maximize
27 the return and restrain the risk (Markowitz, 1952). Portfolio theory proposed by Markowitz in
28 1952, is an important foundation for portfolio management which is a well-studied subject yet not
29 fully conquered territory. There are two issues with portfolio formation. The first one is to select
30 assets with higher revenue, and another one is to determine the value composition of assets in the
31 portfolio to achieve the goal of maximal potential returns with minimal risk. Quantitative approach
32 to the portfolio formation has often been adopted in investment decisions. Based on Markowitz's
33 mean-variance (MV) model, numerous researches have discovered many model extensions and
34 supplemented plentiful reasonable insights about the portfolio formation (Tobin, 1958; Sharpe,
35 1963; Merton, 1969; Grauer and Hakansson, 1993; Liu and Loewenstein, 2002; Tu and Zhou,
36 2010; Brown and Smith, 2011; Li et al., 2013; Li et al., 2015; Bodnar et al., 2017).

37 In the portfolio optimisation process, the expected return on an asset is a crucial factor, which
38 means that a preliminary selection of assets is critical for portfolio management (Guerard Jr et al.,
39 2015). But few researches pay attention to the preselection of assets before forming a portfolio.
40 Asset selection has been a meaningful, but difficult issue in financial investment area. This line of
41 research depends on a long-term volatility of financial time-series data in the past as well as a
42 reliable performance forecasting of assets in the future (Huang, 2012). Traditional statistical
43 methods are not effective in dealing with complex, multi-dimensional and noisy time-series data

44 (Långkvist et al., 2014; Baek and Kim, 2018), while early machine learning methods, such as
45 support vector machine (SVM), principal component analysis (PCA), and artificial neural network
46 (ANN), are not most suited for learning and storing financial time-series data over a long period
47 (LeCun et al., 2015; Bao et al., 2017). This situation leads to the difficulties of financial assets
48 preselection. In fact, during the investment decision-making process, it would be unsustainable to
49 only apply complex portfolio optimisation methods without high-quality asset input (Deng and Min,
50 2013).

51 In the financial market, individual investors usually would like to know the changes in the
52 returns of their investment assets today, the possible trends in the returns tomorrow and which
53 measures should be adopted to help them possess the best portfolio (Zhang et al., 2018). Therefore,
54 incorporating forecasting theory into the portfolio formation will be promising in financial
55 investment (Kolm et al., 2014). Forecasting financial time-series is always regarded as one of the
56 most challenging tasks because of the dynamic, nonlinear, unstable and complex nature with long-
57 term fluctuations of the financial market (Chen and Hao, 2018; Paiva et al, 2019). But a reliable
58 investment decision should rely on long-term observations and patterns of behaviour of asset
59 data rather than short-term (Chourmouziadis and Chatzoglou 2016; Chong et al., 2017). In this
60 case, it is necessary to observe the change and volatility of financial data over a long time in the
61 past so as to make a good preparation for future trends forecasting and investment decisions. And
62 numerous widely accepted empirical researches suggest that financial time-series have a memory
63 of a period in the past, thus to some extent, financial markets are predictable. The behaviour of the
64 asset over a long period will significantly influence the risks and returns of a portfolio, and then
65 further affect the investment decisions (Liu and Loewenstein, 2002). However, this important

66 point is always ignored by current researches. For instance, some apply early machine learning
67 methods, GA (Huang, 2012), SVM (Huang, 2012; Paiva et al., 2019), to predict and select good
68 assets, but fail to capture long-term dependencies of financial time-series data. To overcome this
69 limitation, we present a novel method for portfolio formation in conjunction to the asset
70 preselection, in which long-term dependences of financial time-series data are duly considered.

71 The primary purpose of this paper is to construct an investment decision-making model for
72 individual investors that combines the deep learning LSTM method which concentrates on
73 capturing the long-term dependencies of the returns on assets and the Markowitz's MV method to
74 form optimal portfolios. In this respect, our study has two primary contributions which fill the gaps
75 in existing literature. Firstly, this paper develops a novel method consisting of long short-term
76 memory networks and mean-variance model (LSTM+MV) for optimal portfolio formation. This
77 method considers the long-term dependences on the fluctuations of financial market and captures
78 long-time change patterns of company stocks from the time-series data. To show the benefit of the
79 proposed method in terms of the prediction, early machine learning and statistical models are used
80 in our experiments as baselines to compare with the LSTM networks. Secondly, our proposed
81 model explores in-depth the preselection process of assets before optimal portfolio formation,
82 which guarantees high-quality inputs to the optimal portfolio. Unlike the majority of the methods
83 which aim to improve the existing portfolio management models, this paper focuses on the
84 preliminary phase of portfolio construction, i.e. the preselection of assets. Meanwhile our work
85 provides practical guidance for investors in making better investment decisions. Specifically, the
86 systematic approach present in current paper is able to help decide which assets should be part of
87 the portfolio and the value composition of assets in the portfolio.

88 The remainder of this paper is organised as follows. In Section 2, we review the development
89 of modern portfolio theory and summarise empirical work that has used deep learning to solve
90 issues corresponding to financial time-series data. In Section 3, we describe our methodology in
91 detail, i.e. data source, input variable selection, the proposed model architecture. In Section 4, we
92 present the results of the experiments and explain the results appropriately. In Section 5, we discuss
93 our key findings, implications for theory and practice, also future work.

94 **2. Theoretical background**

95 *2.1 Modern Portfolio Theory*

96 Markowitz (1952) proposes mean-variance (MV) methodology to solve portfolio selection
97 issue, which initiates the foundation of Modern Portfolio Theory (MPT). He quantifies investment
98 return and risk by expected return and variance, respectively. The main idea of MV methodology
99 is to maximize expected return keeping unchanged variance, or minimize variance keeping
100 unchanged expected return. MPT has been widely accepted and studied by researchers. Tobin
101 (1958) indicates that liquidity preference could determine how much wealth is to be invested in
102 monetary assets, and constructs an effective portfolio combined with risk-free assets as well as a
103 special type of risky assets. Sharpe (1963) puts forward the diagonal model assuming that there is
104 no interrelationship among securities so as to simplify the calculation, which significantly
105 facilitates the development of portfolio theory. Some researchers notice that multi-period portfolio
106 selection should be considered to deal with the complex financial markets. For instance, Merton
107 (1969) extends modern portfolio theory by introducing a continuous-time model in order to achieve
108 the goal of maximal expected utility within a constant planning region. Grauer and Hakansson
109 (1993) apply a discrete-time dynamic investment model to compare the MV and the quadratic

110 approximations computing the optimal portfolios. Some researches put several realistic constraints
111 into the Markowitz's MV model. For instance, Liu and Loewenstein (2002) incorporate transaction
112 cost into stock trading strategy to help maximize the investors' wealth utility. Brown and Smith
113 (2011) consider risk aversion, transaction cost, portfolio constraints into MV model and find that
114 it would be difficult to solve portfolio optimisation issues when three more assets are involved.
115 Moreover, some studies use robust optimisation techniques in portfolio management. Tu and Zhou
116 (2010) involve the financial objectives into Bayesian priors to estimate uncertain parameters and
117 they prove that Bayesian method under the objective-based priors performs better than those under
118 alternative priors in portfolio selection. Under a Bayesian estimation framework, Bodnar et al.
119 (2017) analyse the global minimum variance portfolio and consider investors' prior beliefs into the
120 portfolio decisions. On the basis of random matrix theory, Bodnar et al. (2018) evaluate the global
121 minimum variance portfolio with high-dimensional data to minimize the out-of-sample variance.

122 Furthermore, numerous scholars start to analyse portfolio issue using fuzzy set theory. Li
123 and Xu (2013) indicate that there are often fuzzy uncertainty and random uncertainty existing in
124 financial market, hence, they incorporate investors' sentiments and experts' insights into the
125 process of portfolio construction. Assuming that expected rate of returns obeys normal distribution,
126 Li et al. (2013) integrate two constraints, value at risk (VaR) and risk-free assets, into a fuzzy
127 portfolio selection model so as to find a more suitable portfolio. Li et al. (2015) put forward another
128 fuzzy portfolio selection model with background risk to obtain the effective frontier of portfolio.
129 Recently, with the development of big data and artificial intelligence technology, it is possible to
130 use computers and a large number of calculations to achieve optimal portfolio management. Huang
131 (2012) focuses on high-return stock selection using support using genetic algorithms (GAs) as well

132 as vector regression (SVR), but he ignores risk factor. Based on support vector machine (SVM),
133 Paiva et al. (2019) classify the assets to achieve a certain return and determine the components of
134 the investment portfolio. Almahdi and Yang (2017) set three optimisation objectives, annualised
135 Sharpe ratio, Sterling ratio and Calmar ratio, respectively, then choose the best performance
136 algorithm to select optimal portfolio. Yunusoglu and Selim (2013) develop expert system (ES) to
137 support portfolio managers for investment decisions. The expert system contains three stages, the
138 first stage is elimination of unacceptable stock. The second stage is to evaluate stock through a
139 comprehensive literature survey and interviews with a domain expert. The last stage is to construct
140 portfolio based on a mixed-integer linear programming model. Their results demonstrate that under
141 the different risk preference, ES performance is not particularly big difference, moreover, ES is
142 more suitable for 6 months, 9 months and 12 months of investment period.

143 It is obvious that various extensions of Markowitz's MV model help enrich the modern
144 portfolio theory and provide researchers with more research perspectives. And these extensions
145 further confirm that MV model plays an extremely significant role in portfolio management.
146 However, most of the related researches ignores the selection of high-quality assets, the stage before
147 the optimal portfolio formation. Instead, they focus more on how to improve the MV model.
148 Actually, high-quality asset input is a reliable guarantee for optimal portfolio formation during the
149 investment process. In this regard, this paper will continue to adopt the classical MV model,
150 moreover, we will study deeply the preliminary selection of assets in order to provide MV model
151 with better asset inputs. At the same time, different transaction costs will be considered for
152 simulation to visualize the performance of different models.

153 *2.2 Return prediction with deep learning*

154 In recent years, with the development of big data and artificial intelligence (AI) technology,
155 more and more scholars start to use AI as support for their research solutions and prove that AI
156 methods deal with problem of nonlinear, nonstationary characteristics better than traditional
157 statistical models. For example, a number of researches based on SVM (Paiva et al., 2019), PCA
158 (Chen and Hao, 2018; Zbikowski, 2015), GA or random forest (Li and Xu, 2013; Mousavi, 2014),
159 ANN (Patel et al., 2015; Chong et al., 2017) to classify, predict and optimise complex financial
160 assets. Among these technologies, the deep learning is thought to be an appropriate method for the
161 financial time-series forecasting solution, since it is good at processing complex, high-dimensional
162 data as well as extracting abstract characteristics from mass data without depending on any
163 assumptions.

164 The deep learning method proposed by Hinton and Salakhutdinov (2006), has become a
165 leading application in the financial area, especially in predicting financial market movement and
166 processing text information. Deep learning architectures mainly include deep neural networks
167 (DNNs), deep belief networks (DBNs), recurrent neural networks (RNNs) and convolutional neural
168 networks (CNNs) (LeCun et al., 2015). Amongst them, DNNs are feedforward networks in which
169 data flows from the input layer to the output layer by their single directional forward links without
170 going backwards (Arévalo et al., 2016). Chong et al. (2017) testify that with regard to future
171 returns prediction, DNN is obviously superior to a linear autoregressive model based on data from
172 Korean stock market. Identifying the correlation between different stocks, Lachiheba and Gouider
173 (2018) come up with a DNN model with special structure to predict the trend of stock returns over
174 the next five minutes and the results manifest that the accuracy is improved to 71% considerably.

175 DBNs are composed of multiple layers of latent variables, with connections between the layers but
176 not between units within each layer (Hinton, 2009). Shen et al. (2015) construct a DBN using
177 continuous restricted Boltzmann machines to predict exchange rate and their results show that their
178 method performs better than traditional methods. Unlike feedforward neural networks, RNNs can
179 use their internal states (memory) to process sequences of inputs. For instance, Rather et al. (2015)
180 construct a novel hybrid model constituting autoregressive moving average model, exponential
181 smoothing model and RNN to obtain more accurate returns prediction. Similarly, Long et al. (2019)
182 integrate CNN and RNN into their proposed model entitled “multi-filters neural network” aiming
183 to see the trend of the stock price over time, finally, they verify the prediction accuracy of the model
184 through simulation. Long short-term memory (LSTM) networks are one of classes of recurrent
185 neural networks (RNNs), but it has the advantage to retaining information over a long time-span
186 compared with RNNs (LeCun et al., 2015; Fischer and Krauss, 2018). Kraus and Feuerriegel
187 (2017) analyse the text data using the long short-term memory (LSTM) networks, finally they
188 prove that their method increases the accuracy of the stock price prediction. Fischer and Krauss
189 (2018) take advantage of the LSTM networks to forecast stocks directional movement and their
190 results show that LSTM outperforms some classical machine learning models in this prediction
191 task. Besides, Ding et al. (2015) apply CNNs to predict the short-term and long-term influences of
192 events on stock price movements and they prove that the accuracy of the model outperforms other
193 baseline methods.

194 It is clear that the deep learning method is able to find complex structures in high-dimensional
195 financial data and acquire features through simple and non-linear modules, and then transform
196 features from lower level to higher level and more abstract features (LeCun et al., 2015). Based

197 on above literature review, it is easy to discover that the majority of the existing studies on
198 predicting assets returns based on deep learning pay more attention to improve the prediction
199 accuracy, however, few of them apply their prediction results to actual financial markets, such as
200 portfolio management, assets selection, or trading strategy, to give investors more practical
201 guidance. Actually, the high accuracy of prediction does not represent the optimal investment
202 strategy. The advantages of deep learning methods in predicting can be very helpful for decision
203 making in financial investments (Saurabh Aggarwal and Somya Aggarwal, 2017). Therefore,
204 how to combine the prediction of deep learning to help choose the optimal investment strategy is a
205 meaningful and promising research direction (Zhang et al., 2018).

206 **3. Methodology**

207 *3.1 Data*

208 The biggest challenge of prediction is to recognise a relation in financial time-series data between
209 the past and the future (Paiva et al., 2019). Since the continuity of financial stock data, the longer
210 the sample data is involved, the more likely it is to capture history information memory (Fischer
211 and Krauss, 2018; Long et al., 2019). Hence, a large amount of long-term data is required in the
212 empirical experiment (Chourmouziadis and Chatzoglou, 2016). In this research, we collect daily
213 stock data from the UK Stock Exchange 100 Index (FTSE 100) from March 1994 until March 2019,
214 covering 25 years. Since the majority of related studies have been conducted over a period of 10
215 years or less (Kara et al, 2011; Patel et al., 2015; Chen and Hao, 2018), 15 years (Paiva et al., 2019;
216 Almahdi and Yang; 2017), or 25 years (Fischer and Krauss, 2018), our samples spanning 25 years
217 can be considered to provide a sufficiently large volume data to generate statistically significant
218 results. Our sample data involves the historical series of adjusted open prices, close prices, the

219 highest prices, the lowest prices, and the trading volume of assets. Numerous scholars agree on that
 220 holding tens of thousands of different stocks as a portfolio is not realistic for individual investors
 221 (Tanaka et al., 2000; Ranguelova, 2001; Kocuk and Cornuéjols, 2018; Almahdi and Yang, 2017).
 222 For instance, Tanaka et al. (2000) select 9 securities as the sample to form the optimal portfolio.
 223 Almahdi and Yang (2017) construct a five-asset portfolio. Hence, this paper randomly chooses
 224 twenty-one stocks from FTSE 100 as sample data, which is sufficiently large for the asset
 225 preselection before forming portfolio for individual investors. The names of these sample stocks
 226 are “BP” (BP), “Barclays” (BAR), “Tesco” (TES), “Vodafone Group” (VG), “Halma” (HAL),
 227 “Johnson Matthey” (JM), “HSBC Holdings” (HH), “Sainsbury J” (SJ), “Marks & Spencer Group”
 228 (MSG), “Astrazeneca” (AST), “British American Tobacco” (BAT), “PEARSON” (PEA), “Relx”
 229 (RELX), “SSE” (SSE), “Legal & General” (LG), “Royal Bank” (RB), “Royal Dutch Shell B”
 230 (RDSB), “Sage Group” (SG), “Schroders” (SCH), “Seven Trent” (ST) and “Smiths Group” (SG).
 231 Their abbreviations are used for convenience, respectively. Table 1 shows the descriptive statistics
 232 of close prices for the 21stocks selected from FTSE 100. As can be seen, stock AST has the highest
 233 daily mean prices: 2923.12, stock LG has the lowest standard deviation: 65.36, stock VG follows,
 234 with 65.81.

235 Table 1 Descriptive statistics for sample data

Stock	Mean	Std.	Maximum	Minimum
TES	255.66	104.65	492.0	67.33
AST	2923.12	1142.28	6317.0	658.41
BAR	318.90	141.05	710.69	47.0
BP	468.58	109.99	712.0	174.5
BAT	1808.8	1468.73	5643.0	217.59
HAL	357.44	349.40	1648.0	81.5
HH	611.21	165.72	951.6	171.09
JM	1572.69	967.46	3823.0	263.85
LG	141.78	65.36	23.0	294.4
MSG	396.84	111.54	749.0	170.75

PEA	887.68	316.77	2301.79	429.5
RELX	714.0	352.18	1782.0	348.82
RB	2076.30	1877.32	6026.35	103.0
RDSB	1753.62	451.63	761.02	2841.0
SG	978.12	283.06	424.34	1801
SJ	343.75	77.27	594.0	214.6
SCH	1416.98	875.33	3773.0	346.01
ST	1257.59	536.65	2553.0	487.17
SG	978.12	283.06	1801.0	424.34
SSE	984.88	433.62	1696.0	272.5
VG	158.88	65.81	408.57	32.29

236 3.2 LSTM networks

237 LSTM networks were introduced by Hochreiter and Schmidhuber (1997) as an alternative
 238 method to learn sequential patterns. LSTM networks are one of classes of recurrent neural networks
 239 (RNNs), but it has the advantage to retaining information over a long time-span compared with
 240 RNNs (LeCun et al., 2015; Fischer and Krauss, 2018). Graves and Schmidhuber (2005)
 241 demonstrate that LSTM networks are able to overcome the previously inherent problems and
 242 memorize temporal patterns over a long period of time.

243 LSTM networks are comprised of an input layer, several hidden layers, and an output layer.
 244 The most important characteristics of LSTMs is memory cells which contained in the hidden layers.
 245 Fig. 1 illustrates the structure of an LSTM memory cell. As we can see, for each memory cell, x_t
 246 and h_t correspond to the input and hidden state respectively, at time t , and i_t , o_t and f_t , are the
 247 gates which are called input, output and forget gates, respectively, s_t is adjusting its cell state. It
 248 is worth noting that the input gate decides which data can be added into the memory cell, the output
 249 gate decides which data from the memory cell can be used as output, and the forget gate decides
 250 which data should be deleted from the memory cell. The calculations for each state and gate are
 251 performed as the following formulas.

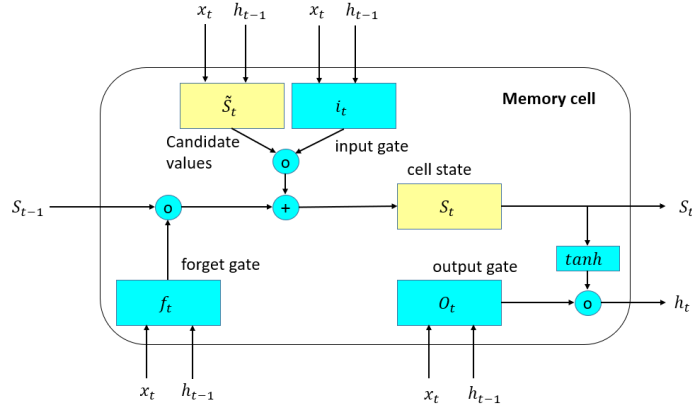


Fig.1. Structure of LSTM memory cell following Fischer and Krauss (2018)

252

253

254 $f_t = \text{sigmoid}(W_{f,x}x_t + W_{f,h}h_{t-1} + b_f)$ (1)

255 $\tilde{s}_t = \text{sigmoid}(W_{\tilde{s},x}x_t + W_{\tilde{s},h}h_{t-1} + b_{\tilde{s}})$ (2)

256 $i_t = \text{sigmoid}(W_{i,x}x_t + W_{i,h}h_{t-1} + b_i)$ (3)

257 $s_t = f_t * s_{t-1} + i_t * \tilde{s}_t$ (4)

258 $o_t = \text{sigmoid}(W_{o,x}x_t + W_{o,h}h_{t-1} + b_o)$ (5)

259 $h_t = o_t * \tanh(s_t)$ (6)

260 Where $W_{f,x}$, $W_{f,h}$, $W_{\tilde{s},x}$, $W_{\tilde{s},h}$, $W_{i,x}$, $W_{i,h}$, $W_{o,x}$ and $W_{o,h}$ are weight matrices, b_f , $b_{\tilde{s}}$,

261 b_i , and b_o are bias vectors of the respective gates. Those bias vectors are added to increase the

262 flexibility of the model to fit the data. Bias vectors $b_{\tilde{s}}$, b_i , and b_o are initialized to zero, but the

263 bias b_f for the forget gate in LSTM is initialized to 1.0 (Jozefowicz et al., 2015). The symbol of

264 * indicates element-wise multiplication. Because of this selective process of information, LSTM

265 is able to deal with longer temporal patterns.

266 3.3 Mean-variance model

267 Mean-variance (MV) model proposed by Markowitz (1952) in order to solve optimal portfolio

268 selection issue, which initiates the foundation of Modern Portfolio Theory (MPT). In this model,

269 investment return and risk are quantified by expected return and variance, respectively. Santos and

270 Tessari (2012) hold the view that the core of the portfolio selection for investors is to decide which
 271 portfolio is the best on the basis of risk and expected returns. Hereby, rational investors always
 272 prefer the lower risk portfolios with constant expected returns or the higher expected return
 273 portfolios with constant risk level. To solve this issue, a set of optimal solutions is generated, named
 274 an efficient investment frontier. The model can be described by the following formulas:

$$275 \quad \underset{w_1, \dots, w_n}{Min} \sum_{i=1}^n \sum_{j=1}^n w_i w_j \delta_{ij} \quad (7)$$

$$276 \quad \underset{w_1, \dots, w_n}{Max} \sum_{i=1}^n w_i \mu_i \quad (8)$$

$$277 \quad \text{Subject to: } \begin{cases} \sum_{i=1}^n w_i = 1 \\ 0 \leq w_i \leq 1, \forall i = 1, \dots, n \end{cases} \quad (9)$$

278 Where w_i and w_j represent the initial value invested in the portfolio or asset i and asset j .
 279 δ_{ij} specifies covariance between assets i and asset j . μ_i is expected return on asset i . Following
 280 Paiva et al. (2019), a variable λ called risk aversion coefficient is integrated into the model to
 281 depict investors' behavior corresponding to the risk investment choices. A mono-objective
 282 formulation is as following:

$$283 \quad \underset{w_1, \dots, w_n}{Min} \lambda \left[\sum_{i=1}^n \sum_{j=1}^n w_i w_j \delta_{ij} \right] - (1 - \lambda) \left[\sum_{i=1}^n w_i \mu_i \right] \quad (10)$$

$$284 \quad \text{Subject to: } \begin{cases} \sum_{i=1}^n w_i = 1 \\ 0 \leq w_i \leq 1, \forall i = 1, \dots, n \end{cases} \quad (11)$$

285 As a result, a group of optimal portfolios constitute an effective frontier can be derived and
 286 introduced to the investor. So, the investor could select the point among these possible solutions
 287 according to his or her risk preference.

288 **3.4 Proposed model: LSTM+MV**

289 Many researches always ignore the fact that the purpose of forecasting financial market is not
 290 to show off the accuracy of a model but to apply these good results into the real market so as to
 291 give investors more practical and meaningful guidance. During the investment decision-making

292 process, high-quality asset input would be very helpful for the optimal portfolio formation. Given
293 the important role that MV method plays in portfolio management area, we will continue to adopt
294 this classical model, moreover, we will study deeply the preliminary selection of assets in order to
295 provide MV model with better asset inputs. In this regard, this study puts forward a mixed method
296 named LSTM+MV combining the advantages of deep learning LSTM method in time-series
297 forecasting with the effectiveness of MV model in portfolio optimisation, aiming to improve the
298 investment decision-making process.

299 There are two stages in our proposed model. In the first stage, LSTM method is applied to
300 predict the return of the sample stocks in the next period. All the predicted results will be sorted in
301 descending order and the top stocks will enter into the next phase. In the second stage, the
302 Markowitz's MV model will be used to obtain the capital allocation proportion for each stock that
303 has been entered.

304 *3.4.1 Input variable selection*

305 The selection of input variables is extremely necessary for time-series prediction tasks. In the
306 light of previous literatures, technical indicators are effective features to describe and reflect the
307 real market situation. For instance, Chen and Hao (2018) suggest that Exponential Moving
308 Average (EMA), Relative Strength Index (RSI) and Momentum Index (MoM) are correlated with
309 changes in stock market. Kara et al (2011) select ten technical indicators as input feature subsets.
310 Also, financial time-series forecasting is always explained by the lagged observations. For example,
311 Fischer and Krauss (2018) use a return time sequence length of 240 for training. Paiva et al. (2019)
312 use several lagged variables of return as inputs to predict the future return of stocks. Hereby, after
313 referring to the views of domain papers, we make feature selection by recursive feature elimination

314 (RFE). To be specific, RFE works by recursively removing features and building a model on those
 315 features that remain. It uses the model accuracy to identify which features contribute the most to
 316 predicting the target feature (return in $t + 1$ period). We use RFE with the logistic regression
 317 algorithm to select the features with a ratio greater than 0.3. Fig. 2 shows the results of feature
 318 selection using RFE. We finally choose twenty important indicators as input variables, including
 319 five technical indicators and fifteen lagged observations about return. The values of all technical
 320 indicators are standardized in the range of $(-1, +1)$, in order to avoid the errors caused by different
 321 indicators of different numerical ranges. Table 2 summarises the selected input variables. Among
 322 the variables are return measures based on open, close, high, low prices, and volume. A brief
 323 explanation of each indicator is as following.

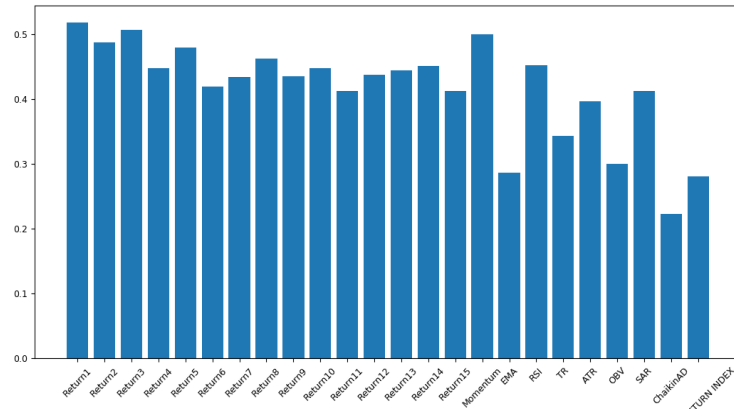


Fig. 2. Feature selection results

324
 325
 326
 327

Table 2 Input features summary

Attribute	Details	Attribute	Details
1	$r_1 = \ln\left(\frac{\text{close price}_i}{\text{close price}_{i-1}}\right)$	11	$r_{11} = \ln\left(\frac{\text{high price}_{i-3}}{\text{open price}_{i-3}}\right)$
2	$r_2 = \ln\left(\frac{\text{close price}_{i-1}}{\text{close price}_{i-2}}\right)$	12	$r_{12} = \ln\left(\frac{\text{low price}_i}{\text{open price}_i}\right)$
3	$r_3 = \ln\left(\frac{\text{close price}_{i-2}}{\text{close price}_{i-3}}\right)$	13	$r_{13} = \ln\left(\frac{\text{low price}_{i-1}}{\text{open price}_{i-1}}\right)$
4	$r_4 = \ln\left(\frac{\text{close price}_{i-3}}{\text{close price}_{i-4}}\right)$	14	$r_{14} = \ln\left(\frac{\text{low price}_{i-2}}{\text{open price}_{i-2}}\right)$
5	$r_5 = \ln\left(\frac{\text{high price}_i}{\text{open price}_i}\right)$	15	$r_{15} = \ln\left(\frac{\text{low price}_{i-3}}{\text{open price}_{i-3}}\right)$
6	$r_6 = \ln\left(\frac{\text{high price}_i}{\text{open price}_{i-1}}\right)$	16	Relative Strength Index (close price, period =14)
7	$r_7 = \ln\left(\frac{\text{high price}_i}{\text{open price}_{i-2}}\right)$	17	Momentum Index (close price, period =10)
8	$r_8 = \ln\left(\frac{\text{high price}_i}{\text{open price}_{i-3}}\right)$	18	True range (high, low, and

9	$r_9 = \ln\left(\frac{\text{high price}_{t-1}}{\text{open price}_{t-1}}\right)$	19	close price) Average true range (high, low and close price, period = 14))
10	$r_{10} = \ln\left(\frac{\text{high price}_{t-2}}{\text{open price}_{t-2}}\right)$	20	Parabolic SAR (high and low price, acceleration = 0.02, maximum = 0)

328

329 1) Simple return

330 Set P_t^i as the price process of stock i at time t , with $i \in \{1, 2, \dots, n\}$ and $R_t^{m,i}$ as the simple
 331 return for a stock i over t periods, i.e., $R_t^{m,i} = \frac{P_t^i}{P_{t-m}^i}$.

332 2) Relative Strength Index (RSI)

333 RSI, a momentum indicator, is able to measure the magnitude of the rise and fall in prices
 334 recently. It is very effective in assessing the overbought/oversold condition of an asset. According
 335 to the parameters of this indicator in existing researches (Paiva et al., 2019; Chen and Hao, 2018;
 336 Patel et al., 2015), this paper set the period as 14.

337 3) Momentum Index (MoM)

338 MoM is an extremely popular indicator measuring a security's rate-of-change, which refers to
 339 the force or speed of movement. Following existing researches (Paiva et al., 2019; Chen and Hao,
 340 2018; Patel et al., 2015), in this paper, the period is set to 10.

341 4) True range (TR)

342 TR is the maximum change in the price of the day compared to the previous day.

343 5) Average true range (ATR)

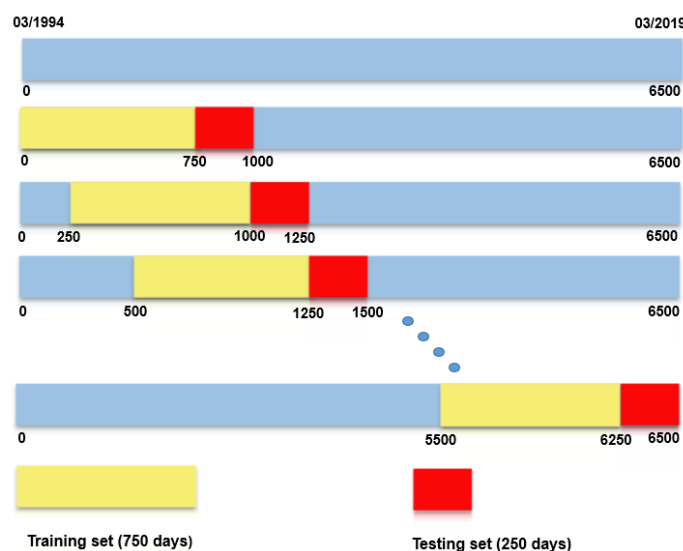
344 ATR is a technical analysis indicator that reflects market volatility through decomposing the
 345 entire range of an asset price for a period.

346 6) Parabolic SAR

347 The parabolic SAR is used to determine the direction in which asset prices rise or fall, besides,
 348 it will remind us when the direction of the price changes, in another words, it will adjust as prices
 349 change so as to attract investors' attention.

350 3.4.2 Generation of training and testing sets

351 Since the continuity of time-series data, we consider each training-testing set as a “study
 352 period”, involving a training period of 750 days and a testing period of 250 days (Fischer and
 353 Krauss, 2018). We divide our sample data from March 1994 until March 2019 into twenty-two
 354 study periods with overlapping training-testing sets. In each study period, the data in the first 750
 355 days is used for training with rolling windows, the rest data fully out-of-sample in the last 250 days
 356 is performed for testing based on the trained parameters. Then, the entire network will roll forward
 357 250 days, leading to twenty-two non-overlapping testing sets. Details can be seen in Fig. 3. The
 358 blue area represents the whole span of our sample, from March 1994 until March 2019. The yellow
 359 area indicates the training set, 750 days. The red area is the testing set, 250 days. The red and yellow
 360 areas together form our “study period”, 1000 days.



361
 362

Fig. 3. overlapping training-testing sets

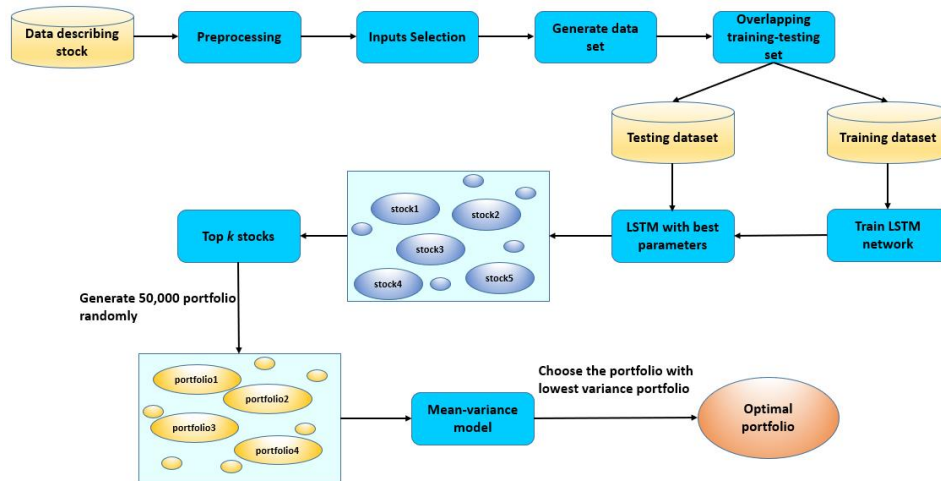
363 3.4.3 *Process of optimal portfolio formation*

364 The proposed model LSTM+MV in this paper is on the basis of technical analysis as well as
365 the historical asset prices identification. On this account, we follow the assumption of Fama (1965)
366 who holds the view that history behaviour trend of the price change in individual assets are inclined
367 to repeat in the future. The primary objective of the LSTM method here is to forecast the relative
368 return rate of each stock in $t + 1$ trading day on the basis of information before time t . In LSTM
369 networks of our proposed model, some sequences of input features are required for training, that is,
370 the values of input features at points in consecutive time. With regards to the training of the LSTM
371 networks, three advanced methods are applied through Keras. First, Adam (Kingma and Ba, 2014)
372 is used as the optimiser to improve the neural network. This selection is inspired from some existing
373 researches (Kingma and Ba, 2014; Reimers and Gurevych; 2017; Kraus and Feuerriegel, 2017), as
374 they testify that Adam is appropriate for deep LSTM networks and has a better performance in
375 optimising the neural network. Second, referring to Srivastava et al. (2014), we make use of dropout
376 regularization technique on the hidden layer. In this case, randomly selected neurons are dropped
377 during training times, along with corresponding input and output connections, which is able to
378 reduce overfitting efficiently (Srivastava et al., 2014; Fischer and Krauss, 2018). In the case of
379 Adam optimiser, we also carry an initial experiment using a part of the sample, the result shows
380 that the model performance decreases as the dropout rate increases, hence, we set the dropout rate
381 relatively low as 0.1. Third, we perform random search method to dynamically find a good
382 combination of hyperparameters based on the above settings. Many empirical evidences have
383 shown the effectiveness of random search in optimising the parameters (Bergstra and Bengio, 2012;
384 Greff et al., 2017). The random search samples the following hyperparameters: (1) the sequence

385 length, ranging from 30 to 250; (2) the number of epochs, ranging from 10 to 100, (3) neuron
386 activation function; (4) the number of neurons per hidden layer, ranging from 2 to 200. Finally, the
387 specified topology of the LSTM network is confirmed. We set 20 features and 72 timesteps in input
388 layer. And in LSTM layer, we set 60 hidden neurons and 0.1 for dropout rate. In dense layer, we
389 apply 16 neurons and relu activation function. Also, we set one neuron and sigmoid activation
390 function in output layer, which is a standard configuration (Fischer and Krauss, 2018). Since the
391 optimal sequence length is 72, approximately covering the data of three testing months. Thus,
392 overlapping sequences of 72 consecutives are generated. In total, 22 study period contain about
393 429,000 of those sequences, in which approximately 321,750 are utilized for in-sample training,
394 and 107,250 are utilized for out-of-sample predictions. For each study period, there are about
395 19,500 of those sequences. Suppose that we would like to find whether an asset has the potential
396 to reach higher return in $t + 1$. Then, we will collect all data of that asset before the trading session
397 at t_0 in order to achieve this goal. According to LSTM principles, the data series from previous
398 days would be put into the model to implement experiment.

399 Once all the assets are predicted, one by one, we will rank all stocks for each period $t + 1$ in
400 descending order of this predicted return. Only the top k of the ranking with the higher return
401 assets that are considered to qualify to enter into the next phase. The purpose of the second stage is
402 to obtain the capital allocation proportion for each asset. And the Markowitz's MV model will be
403 used to carry on this stage. It is worth clarifying that the proposed model does not take into account
404 investors' risk preference and risk-free assets, thus, the portfolios exclusively compose of risky
405 assets. According to the way of Malkiel (2007) letting a blindfolded monkey throw darts at a
406 newspaper's financial pages, we also create a function in python to randomly generate 50,000

407 portfolios. From a statistical perspective, 50,000 random portfolios basically cover most possible
 408 portfolios with different weights and can be regarded representative enough (Fischer and Krauss,
 409 2018). Furthermore, all these 50,000 portfolios will be screened in accordance with MV
 410 optimisation rules so that better portfolio can be find. In the end, the available resources will be
 411 allocated to the portfolio with the lowest variance. As such, when the assets and the respective
 412 investment proportions are confirmed, the next step is to allocate capital at the opening of the next
 413 trading day. We will go long the top k assets during the investment day. The detailed process of
 414 the proposed method is shown in Fig. 4.



415
 416 Fig. 4. The scheme of proposed model

417 **3.4.4 Benchmark models for prediction: SVM, RAF, DNN and ARIMA**

418 In order to benchmark the LSTM, three representative machine learning models, support
 419 vector machine, random forest, deep neural network, as well as a traditional statistical model named
 420 Autoregressive Integrated Moving Average that is often applied for time-series prediction. We will
 421 introduce the principles of each model in the following paragraphs.

422 **Support Vector Machine:** This technique aims to solve issues related to classification,
 423 regression estimation, pattern recognition and time series (Paiva et al., 2019). Support vector

424 regression (SVR), proposed by Drucker et al. (1997), is a version of support vector machine (SVM)
425 for regression. SVR is able to deal with continuous values and find the best regression hyperplanes
426 in order to estimate the dependent variable value (Loureiro et al., 2018).

427 Random Forest: The algorithm derives from the decision trees and is developed to improve the
428 accuracy of decision trees and overcome the high sensitivity to small changes in data. It is generally
429 accepted that it is an advanced machine learning model that usually gets good results and seldom
430 needs tuning (Fischer and Krauss, 2018).

431 Deep neural network: DNN is consisted of multiple hidden layers, one input and one output
432 layer (Loureiro et al., 2018). To be specific, this paper applies a feedforward neural network with
433 20 input neurons and the activation with relu (Li and Yuan, 2017), 30 neurons in the first hidden
434 layer, 3 neurons in the second hidden layer (Fischer and Krauss, 2018), and one neuron in the output
435 layer. Dropout is set to 0.2.

436 Autoregressive Integrated Moving Average model: ARIMA is a classical econometrics model,
437 fitted to predict time-series data in future, and ARIMAX extends ARIMA model by including
438 exogenous variables (Pektas and Cigizoglu, 2013). This paper uses ARIMAX model as one of
439 baseline models.

440 *3.4.5 Baseline strategies for portfolio formation*

441 In reality, except for MV model, equal-weighted portfolio and Black-Litterman (BL) model
442 are also popular. It is worth noting that we originally used the BL model as one of the baselines,
443 but in the end, we found that we could not get a prominent and consistent result to explain. Maybe
444 the parameters of different models need to be adjusted or due to some other reasons we have not
445 figured out. Therefore, we decide not to discuss BL model in this paper. These following baseline

446 strategies are based on the LSTM+MV model proposed in the prior section and used to compare
447 with this model's changes and performance.

448 **(1) Alternative model: Machine learning + MV**

449 This kind of model's design is similar with the logical structure of the LSTM+MV model. The
450 main objective is to find out whether different prediction results of asset return will have an impact
451 on the formation of the final optimal portfolio. To be specific, assets returns in $t + 1$ will be
452 predicted by one machine learning method with better forecasting performance in the in the first
453 stage, and assets with higher return in the future will be chosen into the second stage. Notice that
454 the number of assets selected must be as same as the number defined in the LSTM+MV model.
455 The second stage, portfolio optimisation, applying the Markowitz's MV method is maintained.

456 **(2) Alternative model: Machine learning + 1/N**

457 The objective of this baseline strategy is to examine the portfolio optimisation effect between
458 MV and 1/N (equal-weighted), in the case of the same initial selection of assets. Specifically, one
459 machine learning method with better forecasting performance in the first stage will be used to
460 predict assets returns in $t + 1$, and then rank these assets according to the predicted results. Finally,
461 the top k assets will enter into the second stage and receive the same proportion of investment.
462 Notice that k should be consistent with the number defined in the LSTM+MV model.

463 **(3) Alternative model: Random+ MV or 1/N**

464 This kind of baseline strategy differs from the previous baselines in terms of the asset
465 preselection phase. The asset preselection is randomly undertaken without relying on any
466 predictions, but the number of assets should be same as the number defined by the other models.
467 To be specific, we will randomly select a certain number of assets from all our samples and then

468 apply Markowitz's MV method or 1/N optimisation separately to optimise the portfolio. The
 469 objective of this kind of baseline strategy is to examine the necessity of asset preselection using
 470 machine learning.

471 4. Experiments and Results

472 4.1 Results analysis in the first stage: prediction

473 In this section, we use five criteria to evaluate predictive accuracy, mean square error (MSE),
 474 root-mean-square error (RMSE), mean absolute percentage error (MAPE), mean absolute error
 475 (MAE) and coefficient of determination (R^2). Tables 3 to 5 summarise the best results achieved for
 476 each model applied according to the different evaluation metrics employed. As can be seen from
 477 three tables, the majority of indicators corresponding to LSTM model perform better than the index
 478 value of other models, but several exceptions also exist. For example, the MAE and MAPE
 479 indicator of stock SSE where the prediction result of LSTM is larger than that of SVM and DNN
 480 respectively. Another example is that the R^2 of 3 stocks (BP, JM, SG) predicting by SVM are
 481 higher than that of LSTM, and the R^2 of 2 stocks (BAT, PEA) using RAF are higher than that of
 482 LSTM too.

483

Table 3 Comparison of prediction performance

Stock	LSTM					SVM				
	MSE	RMSE	MAPE	MAE	R^2	MSE	RMSE	MAPE	MAE	R^2
TES	0.0031	0.0557	67.99	0.0364	0.4209	0.0042	0.0651	167.92	0.0427	0.3108
AST	0.0032	0.0568	165.53	0.0324	0.2806	0.0089	0.0942	190.68	0.0512	0.1856
BAR	0.0007	0.0265	6.63	0.0159	0.2631	0.0050	0.0608	22.47	0.0335	0.1200
BP	0.0053	0.0727	123.00	0.0413	0.1121	0.0054	0.0732	114.88	0.0450	0.1379
BAT	0.0019	0.0439	7.76	0.0296	0.1259	0.0064	0.0731	25.50	0.0404	0.0862
HAL	0.0050	0.0709	266.49	0.0378	0.2288	0.0054	0.0735	221.27	0.0397	0.1359
HH	0.0015	0.0390	25.10	0.0214	0.2395	0.0024	0.0484	55.71	0.0290	0.1349
JM	0.0063	0.0797	221.26	0.0495	0.1327	0.0099	0.0997	226.03	0.0431	0.1718
LG	0.0009	0.0293	17.42	0.0155	0.1585	0.0010	0.0317	18.71	0.0176	0.1210
MSG	0.0029	0.0540	17.06	0.0301	0.2630	0.0040	0.0629	19.62	0.0390	0.2141
PEA	0.0018	0.0426	9.06	0.0227	0.1100	0.0026	0.0513	10.95	0.0323	0.1816
REL	0.0028	0.0532	35.08	0.0267	0.1557	0.0029	0.0539	37.78	0.0297	0.1145

RB	0.0002	0.0146	1.98	0.0094	0.4578	0.0003	0.0162	2.22	0.0104	0.3768
RDSB	0.0068	0.0825	158.40	0.0417	0.2960	0.0062	0.0785	154.90	0.0417	0.1093
SG	0.0039	0.0622	74.90	0.0302	0.2545	0.0047	0.0684	71.50	0.0367	0.3592
SJ	0.0028	0.0525	16.29	0.0258	0.6139	0.0031	0.0552	17.00	0.0279	0.1218
SCH	0.0011	0.0329	265.81	0.0189	0.1638	0.0014	0.0368	232.70	0.0224	0.1553
ST	0.0046	0.0680	43.34	0.0388	0.1441	0.0070	0.0837	77.83	0.0450	0.1454
SG	0.0052	0.0726	61.96	0.0394	0.2809	0.0077	0.0876	94.77	0.0510	0.2001
SSE	0.0041	0.0642	196.59	0.0346	0.4572	0.0042	0.0645	245.40	0.0335	0.2977
VG	0.0045	0.0672	138.70	0.0385	0.3448	0.0071	0.0844	237.70	0.0469	0.2810

484

485

Table 4 Comparison of prediction performance

Stock	DNN					RAF					
	MSE	RMSE	MAPE	MAE	R ²	MSE	RMSE	MAPE	MAE	MSE	R ²
TES	0.0130	0.1160	317.26	0.0590	0.1300	0.0058	0.0758	189.56	0.0526	0.0058	0.3593
AST	0.0100	0.1010	194.48	0.0480	0.1410	0.0046	0.0676	162.97	0.0409	0.0046	0.2119
BAR	0.0020	0.0450	10.04	0.0220	0.0800	0.0056	0.0643	23.25	0.0370	0.0056	0.1563
BP	0.0150	0.1210	484.08	0.0590	0.1020	0.0213	0.1402	653.1	0.0817	0.0213	0.1261
BAT	0.0040	0.0630	10.44	0.0340	0.1770	0.0025	0.0495	8.87	0.0285	0.0025	0.2855
HAL	0.0080	0.0900	273.28	0.0460	0.1120	0.0062	0.0787	234.59	0.0413	0.0062	0.2100
HH	0.0060	0.0750	89.25	0.0380	0.1340	0.0069	0.0832	49.11	0.0472	0.0069	0.2047
JM	0.0150	0.1220	495.23	0.0660	0.1210	0.0256	0.1503	296.76	0.0847	0.0256	0.1026
LG	0.0030	0.0540	34.60	0.0230	0.6890	0.0023	0.0475	24.14	0.0282	0.0023	0.3095
MSG	0.0060	0.0760	35.77	0.0390	0.1630	0.0045	0.0674	31.37	0.0362	0.0045	0.2057
PEA	0.0060	0.0780	17.88	0.0340	0.0360	0.0020	0.0442	9.69	0.0265	0.0020	0.2403
REL	0.0047	0.0684	53.22	0.0407	0.1130	0.0046	0.0678	43.69	0.0404	0.0046	0.1153
RB	0.0010	0.0300	3.71	0.0150	0.1530	0.0002	0.0140	1.91	0.0085	0.0002	0.3735
RDSB	0.0170	0.1310	221.60	0.0670	0.1030	0.0095	0.0972	155.13	0.0666	0.0095	0.1051
SG	0.0060	0.0790	95.74	0.0380	0.1293	0.0045	0.0672	81.20	0.0470	0.0045	0.1064
SJ	0.0090	0.0950	38.24	0.0450	0.1407	0.0037	0.0607	19.67	0.0363	0.0037	0.2391
SCH	0.0030	0.0550	270.87	0.0260	0.0840	0.0014	0.0368	291.32	0.0244	0.0014	0.1528
ST	0.0090	0.0930	122.99	0.0500	0.0826	0.0064	0.0797	48.82	0.0474	0.0064	0.1515
SG	0.0120	0.1080	196.68	0.0600	0.1973	0.0062	0.0784	64.88	0.0460	0.0062	0.1032
SSE	0.0090	0.0940	191.07	0.0460	0.2891	0.0050	0.0703	312.05	0.0407	0.0050	0.1439
VG	0.0090	0.0950	236.85	0.0490	0.0100	0.0060	0.0771	186.13	0.0450	0.0060	0.2093

486

487

Table 5 Comparison of prediction performance

Stock	ARIMA				
	MSE	RMSE	MAPE	MAE	R ²
TES	32.22	5.68	8618.9	5.89	0.2427
AST	171.13	13.08	1222.3	10.37	0.1717
BAR	7.90	2.81	389.45	2.56	0.1399
BP	14.83	3.85	7780.1	3.94	0.1038
BAT	1.67	1.29	200.90	0.80	0.0444
HAL	50.50	7.11	19703	5.79	0.0805

HH	3.89	1.97	888.75	1.36	0.0772
JM	2.63	1.62	2701.7	1.05	0.0176
LG	3.08	1.75	518.96	1.53	0.1103
MSG	8.75	2.96	639.18	1.72	0.1648
PEA	7.25	2.69	416.42	2.43	0.1239
REL	67.52	2.60	1065.3	2.14	0.1534
RB	2.25	1.50	214.61	1.56	0.2945
RDSB	11.52	3.39	2297.2	1.99	0.0678
SG	3.50	1.87	884.67	1.66	0.0570
SJ	11.70	3.42	795.99	3.54	0.1286
SCH	39.27	6.27	34652	5.74	0.1363
ST	40.70	6.38	3414.5	6.49	0.3774
SG	51.26	7.16	5028.2	4.07	0.2443
SSE	2.66	1.63	6030.1	1.53	0.0299
VG	1.26	1.12	3880.9	0.86	0.0137

488 Mean square error (MSE): It is an indicator measuring the average squared difference between
489 the observed values and predicted values. From Table 3, Table 4 and Table 5, we find the following
490 average MSE result: 0.0033 for LSTM model, 0.0047 for SVM model, 0.008 for DNN model, 0.64
491 for RAF model and 25.49 for ARIMA model.

492 Root-mean-square error (RMSE): It is another effective indicator measuring differences
493 between the observed values and predicted values. As can be seen from Tables 3 to 5, LSTM
494 exhibits favourable mean RMSE 0.0543, followed by SVM (0.0649), then the indicator for DNN
495 and RAF equals to 0.0852 and 0.0723, but 3.817 for ARIMA model.

496 Mean absolute percentage error (MAPE): It measures the prediction deviation proportion in
497 terms of the true value. After comparing different models in terms of MAPE, we can get the average
498 results: 91.44 for LSTM model, 106.93 for SVM model, 161.58 for DNN model, 137.53 for RAF
499 model and 2560 for ARIMA model.

500 Mean absolute error (MAE): It is a measure of accuracy of a forecasting method. We see that
501 the LSTM has the lowest mean MAE of 0.0303, followed by SVM (0.0361), 0.0431 for DNN
502 model, 0.0432 for RAF model and 3.1913 for ARIMA model.

503 Coefficient of determination (R^2): This is a measure of how well the model can be explained.
504 The R^2 of RAF, SVM and DNN is a little higher than that of LSTM in terms of several stocks,
505 but on average, the LSTM model has the highest R^2 of 0.2621, followed by RAF (0.1958) and
506 SVM (0.1886), 0.1518 for DNN model and 0.0699 for ARIMA model. We can see that R^2 of
507 LSTM ranges from 0.1100 to 0.6139, similar to several existing financial researches (Gatev et al.,
508 2006; Fischer and Krauss, 2018). To be specific, the main purpose of the preselection phase is to
509 forecast the return of assets and select assets with higher potential returns. Unlike researches on
510 explanatory modelling aiming to explain causal relationships and the importance of each indicator,
511 predictive modelling is primarily concerned with accuracy and error in order to predict future
512 observations (Shmueli, 2010; Gandhmal and Kumar, 2019). In this case, the effectiveness of this
513 kind of model is primarily determined by accuracy measures, such as RMSE and MSE, rather than
514 the value of R^2 (Alexander et al., 2015; Gandhmal and Kumar, 2019).

515 With regard to stock market prediction, MSE, RMSE, MAPE, and MAE are generally
516 regarded as popular performance metrics since they can clearly present the average model
517 prediction error (Kao et al., 2013; Weng et al., 2018; Gandhmal and Kumar, 2019). For several
518 other works, it is difficult to evaluate these metrics through direct comparison due to the difference
519 of datasets. But we can compare the results with widely used methods in related researches. From
520 Tables 3 to 5, the average values of MSE, RMSE, MAE for LSTM model are 0.0033, 0.0543 and
521 0.0303 respectively, which have showed superior performance in forecasting stock returns against
522 existing works (Ticknor, 2013; Sadaei et al., 2016; Weng et al., 2018; Gandhmal and Kumar, 2019).

523 In conclusion, the LSTM model predictions are superior to other baseline methods in both
524 accuracy and direction. And the predicted performance of SVM and RAF is second only to LSTM,

525 but far better than DNN and ARIMA model. Besides, traditional statistics model ARIMA performs
526 worst. For example, for stock TES, the MSE in ARIMA equals to 32.21, which is 5000 times bigger
527 than MSE (0.0031) in LSTM.

528 ***4.2 Results analysis in the second stage: optimal portfolio formation***

529 *4.2.1 Determination of the portfolio size*

530 Firstly, we analyse the characteristics of portfolios consisting of k assets. Most of researches
531 corresponding to portfolio formation for individual investors focus on only fewer than 10 assets
532 (Kocuk and Cornuéjols, 2018; Tanaka et al., 2000; Almahdi and Yang, 2017), because holding too
533 many different stocks is hard for an individual investor to manage. Rangelova (2001) indicate that
534 individual investors usually hold three or four stocks in their account on average. Paiva et al., (2019)
535 discover that the portfolio with seven assets performs better than others with different numbers of
536 assets. Hereby, assuming an individual investor holding less than or equal to 10 assets is realistic.
537 Based on the above discussion, we choose $k \in \{4, 5, 6, 7, 8, 9, 10\}$, and then compare the
538 performance of the model LSTM+MV with the other baseline strategies according to the
539 dimensions annualised standard deviation, annualised mean return, annualised Sharpe ratio, and
540 Sortino ratio before transaction costs.

541 As can be seen from Fig. 5, there are four subgraphs. Specifically, the Y-axis of four sub-
542 graphs represents mean return, standard deviation, Sharpe ratio and Sortino ratio, the X-axis of four
543 subgraphs represents the same meaning, that is, different models with different portfolio sizes.
544 From Fig. 5, it is clear that irrespective of the portfolio size k , the LSTM+MV shows greater
545 performance than the other strategies in three dimensions of annualised mean return, Sharpe ratio
546 and Sortino ratio. To be specific, annualised returns prior to transaction costs are at 0.16, compared

547 to 0.09 for the LSTM+1/N, 0.11 for the SVM+MV, 0.07 for the SVM+1/N, 0.09 for the RAF+MV
548 and 0.06 for RAF+1/N for $k = 8$. For other portfolio sizes, like $k = 10$, the LSTM+MV also
549 achieves the highest mean returns per year. In regard to annualised standard deviation, a risk metric,
550 differences among models are not obvious, the LSTM +MV is on a similar level as the other models,
551 with slightly higher values for $k = 6, 7, 8$, thus we could not distinguish which models are good
552 or bad on this metric easily. In this study, we set risk-free ratio as 0.0125, according to the British
553 treasuring bill rate in recent 10 years. With respect to Sharpe ratio, return per unit of risk, is highest
554 for the LSTM+MV. For example, when $k = 9$, Sharpe ratio before transaction cost is 0.58,
555 compared to 0.40 for the LSTM+1/N, 0.46 for the SVM+MV, 0.36 for the SVM+1/N, 0.38 for the
556 RAF+MV, 0.28 for RAF+1/N. Sortino ratio, measuring the risk-adjusted return of an investment
557 portfolio. A clear advantage of the LSTM+MV can be seen for portfolios of each size. From the
558 perspective of different portfolio sizes, it is easy to find that the four indicators perform better
559 overall in each model when $k = 10$ than other sizes. Specifically, in model LSTM+MV, the
560 portfolio with $k = 10$ not only has a high mean return 0.136, Sharpe ratio 0.58 and Sortino ratio
561 13.7, but also has a lower standard deviation 0.21. And the same is true for the analysis of other
562 models. From the above analysis, we focus the portfolio with $k = 10$ in our subsequent analyses,
563 which is also consistent with the research of Fischer and Krauss (2018).

564

565

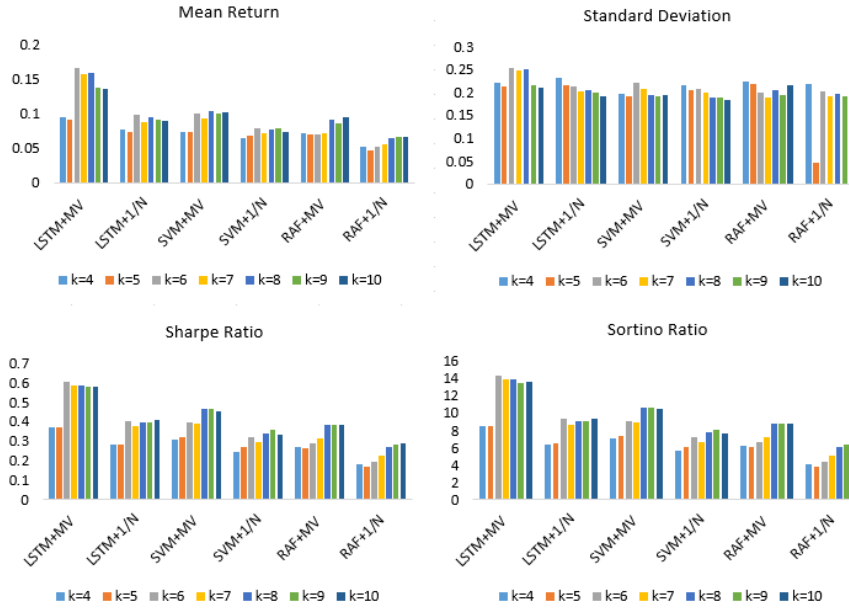


Fig. 5. Annualised performance characteristics for portfolios of different sizes

566
567

568 *4.2.2 Details on financial performance*

569 It is worth clarifying that this paper only considers brokerage cost as transaction cost because
570 the investor is able to control brokerage cost directly (Paiva et al., 2019). According to Brooks et
571 al. (2001), brokerage costs for purchasing and selling the stocks of FTSE 100 index is from 0.00bps
572 to 0.30 bps. Referring to the parameters of several empirical research (Almahdi and Yang, 2016;
573 Guerard Jr et al., 2015; Paiva et al., 2019), we decide to simulate transaction costs as 0.10 bps,
574 0.05 bps to present the results finally. Tables 6 to 8 provide insights of the financial performance
575 of the LSTM+MV, compared to the baselines, without transaction cost, including transaction cost
576 (0.1bps, 0.05bps) separately. Hence, Panel A, B and C depict daily return characteristics, daily risk
577 characteristics and annualised risk-return metrics respectively.

578

Table 6 Performance characteristics for portfolios without transaction cost

	LSTM+MV	LSTM+1/N	SVM+MV	SVM+1/N	RAF+MV	RAF+1/N
A Mean return	0.0005	0.0004	0.0004	0.0003	0.0004	0.0003
Standard deviation	0.0134	0.0121	0.0124	0.0116	0.0137	0.0118
Maximum	0.1003	0.1100	0.0953	0.0965	0.1241	0.1052
Minimum	-0.0748	-0.0953	-0.0744	-0.1014	-0.0749	-0.1005
B 1-percent VaR	0.0330	0.0337	0.0336	0.0327	0.0385	0.0328
1-percent CVaR	0.0439	0.0446	0.0451	0.0427	0.0514	0.0443

5-percent VaR	0.0207	0.0189	0.0188	0.0178	0.0207	0.0188
5-percent CVaR	0.0306	0.0282	0.0285	0.0271	0.0318	0.0277
Maximum drawdown	2.5277	2.4182	2.9685	2.2441	2.5612	2.1990
C Mean return	0.1367	0.0913	0.1022	0.0743	0.0963	0.0676
Standard deviation	0.2125	0.1919	0.1963	0.1844	0.2176	0.1878
Sharpe ratio	0.5845	0.4105	0.4569	0.3354	0.3852	0.2932
Sortino ratio	13.7078	9.3844	10.4918	7.6352	8.8549	6.6693

579

580

Table 7 Performance characteristics for portfolios including transaction cost (0.05bps)

	LSTM+MV	LSTM+1/N	SVM+MV	SVM+1/N	RAF+MV	RAF+1/N
A Mean return	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003
Standard deviation	0.0125	0.0139	0.0155	0.0125	0.0139	0.0134
Maximum	0.1152	0.1397	0.1535	0.1031	0.1186	0.1302
Minimum	-0.1027	-0.1746	-0.2575	-0.1322	-0.1387	-0.1698
B 1-percent VaR	0.0341	0.0368	0.0408	0.0337	0.0384	0.0365
1-percent CVaR	0.0472	0.2119	0.0598	0.1897	0.0529	0.2047
5-percent VaR	0.0196	0.0210	0.0232	0.0196	0.0209	0.0202
5-percent CVaR	0.0293	0.0424	0.0352	0.0379	0.0321	0.0409
Maximum drawdown	2.3442	2.5068	2.7753	2.9920	3.6550	2.3043
C Mean return	0.0765	0.0789	0.0792	0.0780	0.0691	0.0630
Standard deviation	0.1988	0.2203	0.2462	0.1988	0.2207	0.2129
Sharpe ratio	0.3218	0.2978	0.2710	0.3294	0.2567	0.2374
Sortino ratio	0.0906	0.0834	0.0753	0.0926	0.0720	0.0667

581

582

Table 8 Performance characteristics for portfolios including transaction cost (0.1bps)

	LSTM+MV	LSTM+1/N	SVM+MV	SVM+1/N	RAF+MV	RAF+1/N
A Mean return	0.0003	0.0003	0.0003	0.0003	0.0002	0.0003
Standard deviation	0.0149	0.0140	0.0153	0.0126	0.0155	0.0135
Maximum	0.1524	0.1411	0.1445	0.1041	0.1658	0.1314
Minimum	-0.2124	-0.1762	-0.2340	-0.1334	-0.2541	-0.1714
B 1-percent VaR	0.0388	0.0372	0.0405	0.0340	0.0408	0.0368
1-percent CVaR	0.0582	0.0212	0.0600	0.1915	0.0612	0.2066
5-percent VaR	0.0222	0.2140	0.0228	0.0198	0.0229	0.0204
5-percent CVaR	0.0344	0.0428	0.0352	0.0383	0.0356	0.0413
Maximum drawdown	2.7861	2.5068	3.0653	2.9920	2.5326	2.3043
C Mean return	0.0763	0.0787	0.0705	0.0787	0.0616	0.0636
Standard deviation	0.2366	0.2224	0.2434	0.2007	0.2468	0.2149
Sharpe ratio	0.2697	0.2975	0.2384	0.3300	0.1990	0.2379
Sortino ratio	0.0752	0.0833	0.0662	0.0924	0.0553	0.0665

583

Return characteristics: In panel A of Table 6, we can see that the LSTM+MV exhibits

584 favourable daily mean return 0.0005, and the SVM+MV has the lowest standard deviation as
585 0.0116. After including transaction cost 0.05 bps, in panel A of Table 6, we can find that all the
586 models have almost same daily return 0.0003, the LSTM+MV model and SVM+1/N model have
587 the lowest standard deviation. After including transaction cost 0.1 bps, in panel A of Table 7,
588 SVM+1/N model has a better risk level, daily standard deviation equals to 0.0126.

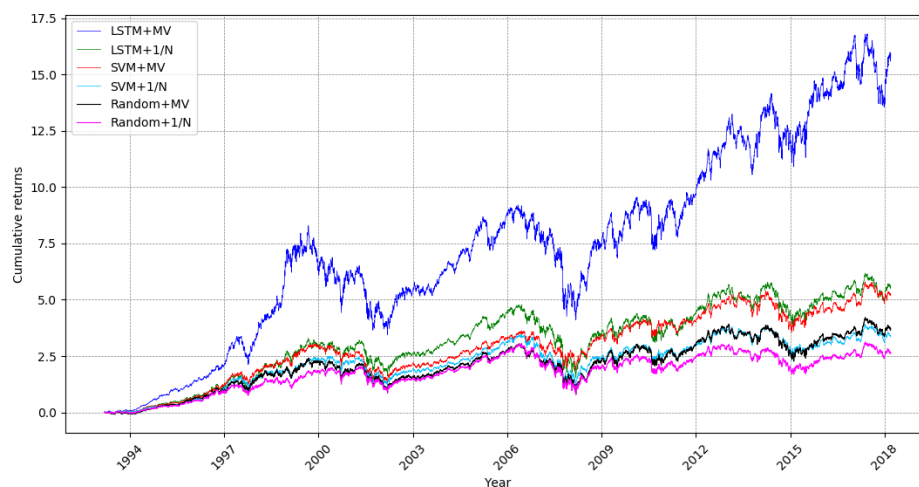
589 *Risk characteristics:* In panel B of Table 6, Table 7 and Table 8, we can see a mixed picture
590 corresponding to risk characteristics. Before transaction cost, SVM+1/N achieved the best place
591 with a 1-percent VaR of 0.0327, 5-percent VaR of 0.0178, 1-percent CVaR of 0.0427 and 5-percent
592 CVaR of 0.0271. After including transaction cost 0.05 bps, the LSTM+MV performs better, with
593 1-percent CVaR of 0.0472, 5-percent VaR of 0.0196 and 5-percent CVaR of 0.0293. After including
594 transaction cost 0.1 bps, in terms of 1-percent VaR, SVM+1/N model has the lowest value.
595 LSTM+1/N achieves the lowest 1-percent VaR, SVM+1/N performs best for 5-percent VaR.

596 *Annualised risk-return metrics:* In panel C of Table 6, Table 7 and Table 8, we discuss risk-
597 return metrics on an annualised basis. It is clear that the LSTM+MV achieves the highest annualised
598 returns of 0.1367 without transaction costs, followed by the SVM+MV (0.1022). SVM+MV and
599 SVM+1/N perform best in terms of annualised mean returns with transaction cost 0.05 bps and 0.1
600 bps. The Sharpe ratio measures excess return using standard deviation and can be explained as the
601 return per unit of risk. We find that the LSTM+MV achieves the highest level of 0.5845, with the
602 SVM+MV coming in second with 0.4569. After transaction cost 0.1 bps and 0.05 bps, SVM+1/N
603 gets the highest Sharpe ratio at 0.3294 and 0.3300 respectively. In addition, SVM+1/N achieves
604 the first place in terms of standard deviation and Sortino ratio, followed by LSTM+MV (0.05 bps)
605 and LSTM+1/N (0.1 bps) respectively.

606 From a financial perspective, we can find that the LSTM+MV, SVM+MV, LSTM+1/N and
607 SVM+1/N outperform the RAF+MV and RAF+1/N in terms of the return, risk or risk-return
608 metrics. In order to compare these models further, we are thus able to choose these four more
609 competitive strategies to visualize performance over time, i.e., from March 1994 to March 2019.

610 4.2.3 Visualization on financial performance

611 In this section, we select 4 models, LSTM+MV, SVM+MV, LSTM+1/N and SVM+1/N, that
612 perform better in the previous section to display their performance for further comparisons. Besides,
613 we also consider Random+MV and Random+1/N as comparison models to examine the necessity
614 of using machine learning for asset pre-selection and further verify whether our proposed method
615 is effective comparing with other portfolio data sets. Fig. 6 presents the cumulative return for each
616 model without transaction cost. The LSTM+MV model has an obviously higher result and achieves
617 cumulative return of 15.9 approximately. The profitability of the LSTM+1/N model follows, with
618 5.7, and then the SVM+MV, with 5.5. And the Random+MV and the SVM+1/N keep similar at
619 about 3.3, the Random+1/N is the lowest, with 2.5. Furthermore, we should also figure out how the
620 LSTM+MV and other models behave at different levels of transaction costs.

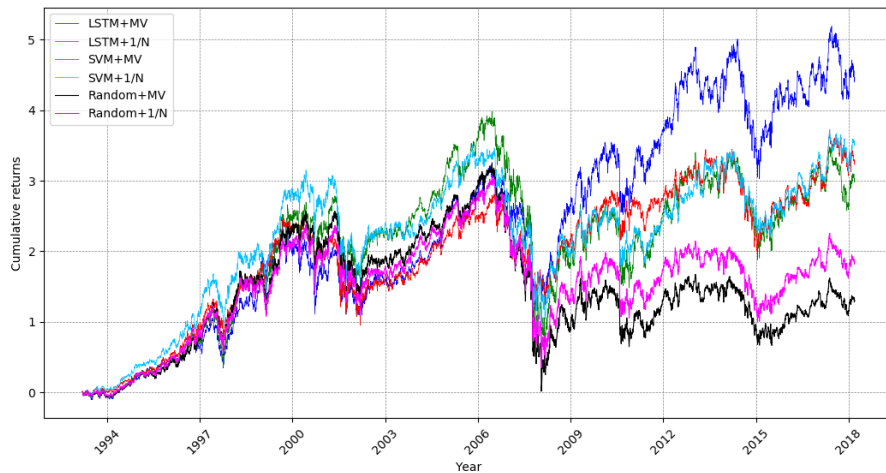


621
622

Fig. 6. cumulative return without transaction cost

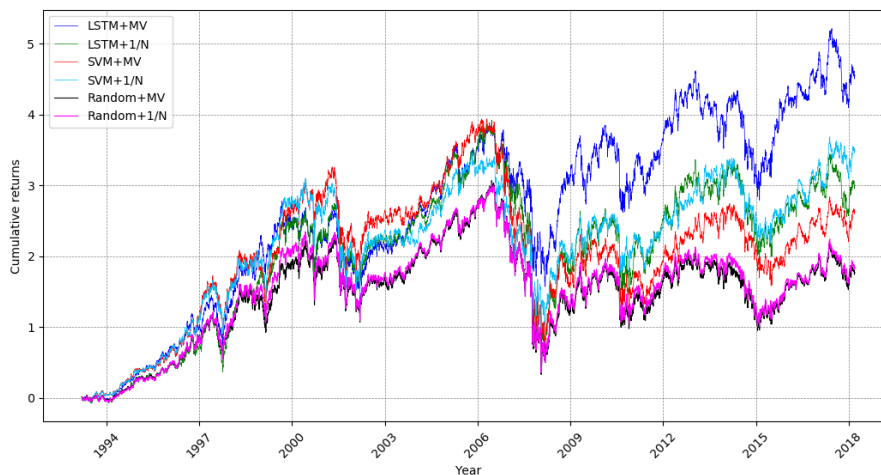
623 Figures 7 and 8 depicts the simulations of the cumulative returns considering transaction costs

624 of 0.05bps and 0.10 bps, respectively, and the accumulated returns are strongly decreased. But in
625 general, the LSTM+MV model still maintains a better accumulated return. The cumulative return
626 with a transaction cost of 0.05 bps is about 4.6, while for a transaction cost of 0.10 bps it is 4.5.



627
628 Fig. 7. cumulative return including transaction cost (0.05 bps)

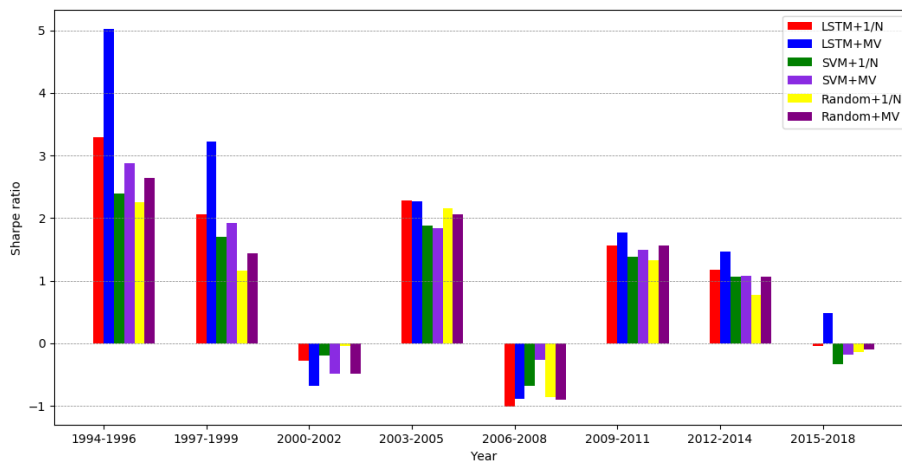
629



630
631 Fig. 8. cumulative return including transaction cost (0.1 bps)

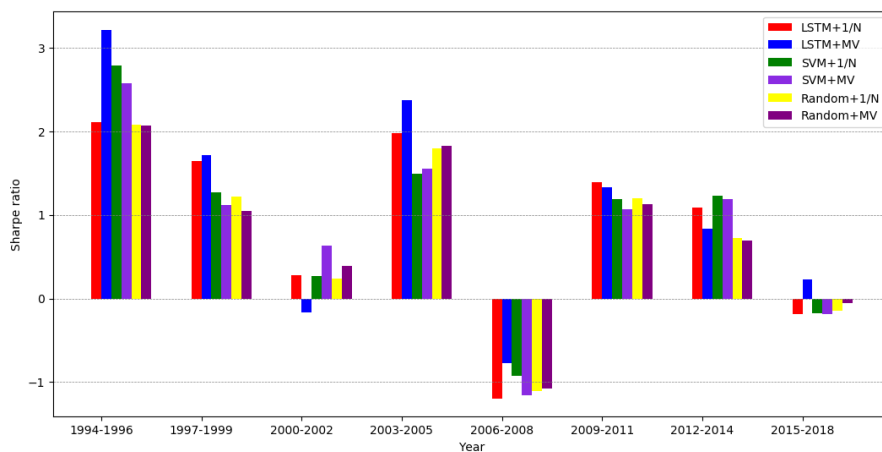
632 From the comparison of the cumulative returns between the LSTM+MV model and the other
633 baseline strategies, we can discover that LSTM+MV performs much better than other baselines in
634 terms of return metrics. Another idea which is inspired from this is that we would like to see the
635 results when integrating risks and whether the good performance only occurs during a certain
636 period of time. As shown in Fig. 9, we use the Sharpe ratio performance of each model every three

637 years. We can observe that, of the eight surveyed triennia, six of them show that the Sharpe ratio
 638 of the LSTM+MV model has a better result than other models during the corresponding periods.
 639 Figures 10 and 11 present the Sharpe ratio per triennium with transaction costs. The LSTM+MV
 640 model, with transaction costs of 0.05 bps, behaves better. Specifically, among the eight surveyed
 641 triennia, five of them have a higher Sharpe ratio in LSTM+MV model than other models. After
 642 including transaction cost 0.1 bps, only half of the surveyed period shows a greater result of the
 643 LSTM+MV model.



644
 645

Fig. 9. Sharpe ratio of each triennium without transaction costs



646
 647
 648

Fig. 10. Sharpe ratio of each triennium including transaction costs (0.05 bps)

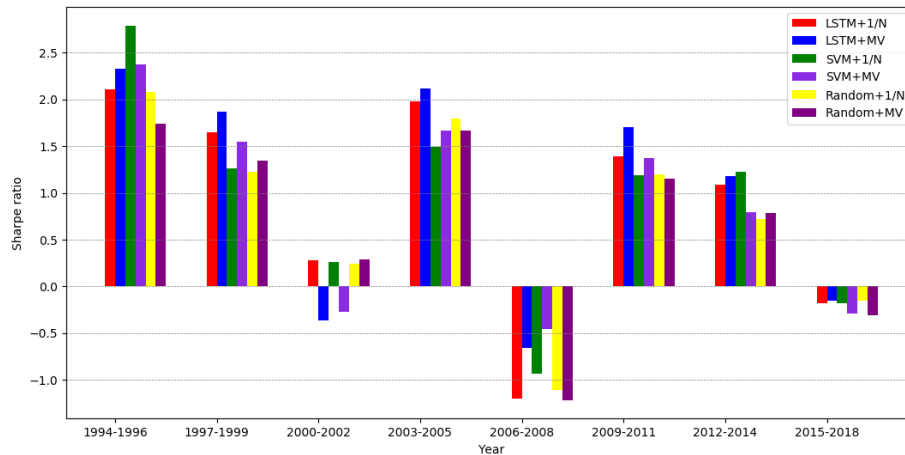


Fig. 11. Sharpe ratio of each triennium including transaction costs (0.1 bps)

649
650

651 Fig. 12 depicts the result of average return to the risk per month of each triennium per model
 652 without transaction costs. Apparently, the LSTM+MV model obtains a remarkable performance for
 653 the return-risk ratio during most study period. We also discover the average results as followings:
 654 0.2670 for the LSTM+MV model, 0.1966 for the LSTM+1/N model, 0.1808 for the SVM+MV,
 655 0.1581 for the SVM+1/N model, 0.1593 for the Random+MV, and 0.1458 for the Random+1/N
 656 model. The LSTM+MV model stops having the highest value during period 2006-2008, and this
 657 result coincides with the financial crisis and troubled political.

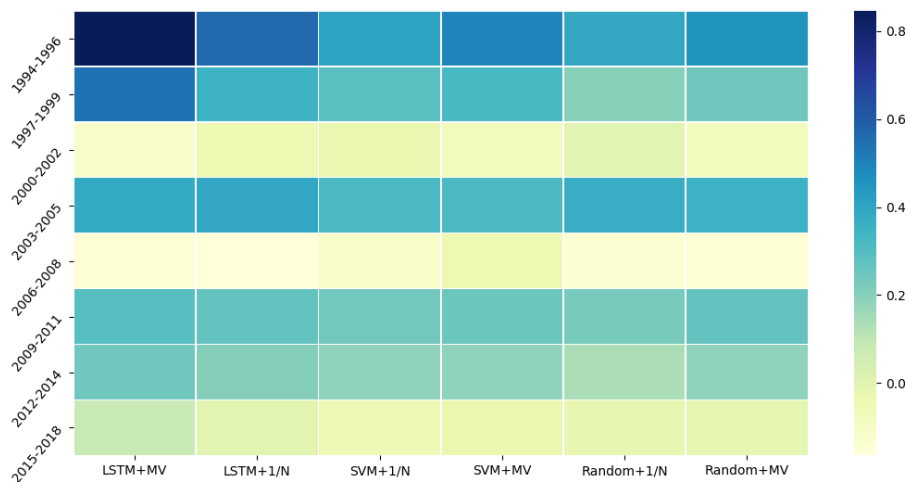


Fig. 12. Average return to the risk per month of each triennium without transaction costs

658
659

660 5. Discussion and Conclusions

661 5.1 Discussion for key findings

662 This paper puts forward an investment decision model entitled LSTM+MV. Based on the

663 LSTM method, predict and select assets with a higher daily return of gain, then integrate this
664 prediction with the MV diversification method to compose the optimal portfolio. Our study results
665 in several important findings.

666 First of all, LSTM networks are applied to achieve the financial time-series prediction
667 empirical application on big data volume. Specifically, we create an appropriate prediction task,
668 divide whole sample data set into 22 overlapping training-testing sets, normalize the input features
669 in order to facilitate model training, find an appropriate LSTM architecture for forecasting. After
670 comparing the outcomes of the LSTM against SVM, RAF, DNN as well as ARIMA, we discover
671 the LSTM networks are appropriate for financial time-series forecasting, to beat the other early
672 machine learning models and the statistics model by a very clear margin.

673 Secondly, for individual investors, holding 10 assets is realistic and helps them maintain better
674 returns with the same level of risk. In this case, the LSTM+MV, SVM+MV, LSTM+1/N and
675 SVM+1/N outperform the Random+MV and Random+1/N in terms of the return, risk or risk-return
676 metrics. Among these results, we further display their performance in accordance cumulative return
677 per year, Sharpe ratio per triennium as well as average return to the risk per month of each triennium.

678 Finally, for cumulative return performance without transaction costs, the LSTM+MV model
679 is significantly better than the other baseline models. A three-year Sharpe ratio experiment also
680 confirms the better performance of the LSTM+MV model. After including transaction costs, the
681 LSTM+MV model still outperforms the other models with a better outcome. In that case, the
682 applicability of the model's implementation may depend on the amount of money invested by
683 investors.

684 *5.2 Theoretical implications*

685 This research enriches the theoretical literature on the stock return prediction and portfolio
686 management. First of all, the portfolio formation method proposed in this paper is able to capture
687 the long-term dependences of financial time-series data fluctuation, which fills the gap in
688 corresponding portfolio optimisation researches paying insufficient attention to the continuity and
689 memory characteristics of financial time-series data. To be specific, this paper compares the
690 forecasting outcomes of the LSTM with SVM, RAF, DNN as well as ARIMA to demonstrate the
691 accuracy and feasibility of LSTM networks in predicting financial time-series more convincingly.

692 Second, the preselection process of assets is incorporated into the optimal portfolio formation.
693 Instead changing and improving the Markowitz' MV model, this paper puts effort into the
694 preliminary phase of portfolio construction to ensure that the portfolio is composed of assets with
695 high-return in the beginning. Specifically, our study demonstrates that the proposed model
696 LSTM+MV is able to help individual investors obtain remarkable outcomes for the cumulative
697 returns as well as risk-adjusted return for majority of periods. The merger of the return forecasting
698 and portfolio optimisation processes may provide a new perspective for research in fintech area.

699 *5.3 Practical implications*

700 The study also provides several practical implications. For portfolio managers, this paper puts
701 forwards a practical method for optimal portfolio selection that can help improve day investments.
702 Following this model, managers can pick assets with higher return based on the predicting results
703 in real market, and then apply MV model to reduce risk level so that keep investments safe and
704 beneficial. For individual investors, this method is able to systematically help them to make
705 decisions for investing. In another words, tell them which assets they should hold and how much
706 to invest in each asset to achieve the goal of maximal potential return with minimal risk.

707 **5.4 Limitations and future work**

708 Although this research provides useful insights, there are some limitations in this study, which
709 provide opportunities for further research. First, five technical indicators and fifteen lagged
710 variables are used as input features to predict the return in the future, however, there are some other
711 external environment factors, such as government policies, interest rates, public events and so forth
712 that have an impact on financial market can also be considered as the input indicators to the models
713 (Christou et al., 2017). In addition, the study uses the asset data in only one country of UK. Due
714 to the different political environment and economic backgrounds, we cannot ensure whether the
715 proposed method is suitable for the stock markets from other countries. Thus, in future research,
716 asset data from more countries should be used for experiments and comparisons to further testify
717 the applicability and establish the boundaries of the proposed model.

718 **Acknowledgements**

719 Wuyu Wang, Weizi Li and Kecheng Liu conceptualised the idea; Wuyu Wang and Weizi Li
720 completed the experiments; Wuyu Wang analysed the results and wrote the paper, Kecheng Liu,
721 Ning Zhang and Weizi Li provided language help and proof reading the paper.

722 **References**

- 723 Aggarwal, S., & Aggarwal, S. (2017). Deep Investment in Financial Markets using Deep
724 Learning Models. *International Journal of Computer Applications*, 162(2), 0975-8887.
- 725 Alexander, D. L. J., Tropsha, A., & Winkler, D. A. (2015). Beware of R^2 : Simple,
726 Unambiguous Assessment of the Prediction Accuracy of QSAR and QSPR Models.
727 *Journal of Chemical Information and Modeling*, 55, 1316-1322.
- 728 Almahdi, S., & Yang, S. Y. (2017). An adaptive portfolio trading system: A risk-return portfolio
729 optimisation using recurrent reinforcement learning with expected maximum drawdown.

730 *Expert Systems With Applications*, 87, 267-279.

731 Arévalo, A., Niño, J., Hernández, G., & Sandoval, J. (2016). High-frequency trading strategy
732 based on deep neural networks. *International conference on intelligent computing*, 424-
733 436.

734 Baek, Y., & Kim, H. Y. (2018). ModAugNet: A new forecasting framework for stock market
735 index value with an overfitting prevention LSTM module and a prediction LSTM module.
736 *Expert Systems With Applications*, 113, 457- 480.

737 Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using
738 stacked autoencoders and long-short term memory. *PLoS ONE*, 12(7), 1-24.

739 Bergstra, J., & Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. *Journal*
740 *of Machine Learning Research*, (3), 281-305.

741 Bodnar, T., Mazur, S., & Okhrin, Y. (2017). Bayesian estimation of the global minimum
742 variance portfolio. *European Journal of Operational Research*, 256, 292-307.

743 Bodnar, T., Parolya, T., & Schmid, W. (2018). Estimation of the global minimum variance
744 portfolio in high dimensions. *European Journal of Operational Research*, 266, 371-390.

745 Brooks, C., Rew, A. G., Ritson, S. (2001). A trading strategy based on the lead-lag relationship
746 between the spot index and futures contract for the FTSE 100. *International Journal of*
747 *Forecasting*, 17, 31-44.

748 Brown. D. B., & Smith, J. E. (2011). Dynamic Portfolio Optimisation with Transaction Costs:
749 Heuristics and Dual Bounds. *Management Science*, 57(10), 1752-1770.

750 Chen, Y. J., & Hao, Y. J. (2018). Integrating principle component analysis and weighted support
751 vector machine for stock trading signals prediction. *Neurocomputing*, (321), pp. 381- 402.

752 Chong, E., Han, C., & Park, F. C. (2017). Deep learning networks for stock market analysis
753 and prediction: Methodology, data representations, and case studies. *Expert Systems With*
754 *Applications*, (83), 187-205.

755 Chourmouziadis, K., & Chatzoglou, P. D. (2016). An intelligent short term stock trading fuzzy
756 system for assisting investors in portfolio management. *Expert Systems with Applications*,
757 (43), 298-311.

758 Christou, C., Cunado, J., Gupta, R., & Hassapis, C. (2017). Economic policy uncertainty and
759 stock market returns in PacificRim countries: Evidence based on a Bayesian panel VAR
760 model. *Journal of Multinational Financial Management*, (40), 92-102.

761 Chuluun, T. (2017). Global portfolio investment network and stock market comovement.
762 *Global Finance Journal*, (33), 51- 68.

763 Deng, S. J., & Min, X. (2013). Applied Optimization in Global Efficient Portfolio Construction
764 Using Earning Forecasts, *Journal of Investing*, 22(4), 104-114.

765 Ding, X., Zhang, Y., Liu, T., & Duan, J. (2015). Deep learning for event-driven stock prediction.
766 *Proceedings of the Twenty-Fourth International Joint Conference on Artificial*
767 *Intelligence*.

768 Drucker, H., Burges, C. J., Kaufman, L., Smola, A., & Vapnik, V. (1997). Support Vector
769 Regression Machines. *Advances in Neural Information Processing Systems*, 9, 155-161.

770 Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for
771 financial market predictions. *European Journal of Operational Research*, 270, 654-669.

772 Gatev, E., Goetzmann, W. N., & Rouwenhorst, K. G. (2006). Pairs trading: Performance of a
773 relative-value arbitrage rule. *Review of Financial Studies*, 19(3), 797-827.

774 Gandhmal, D. P., & Kumar, K. (2019). Systematic analysis and review of stock market
775 prediction techniques. *Computer Science Review*, 34, 100190

776 Grauer, R. R., & Hakansson, N. H. (1993). On the Use of Mean-Variance and Quadratic
777 Approximations in Implementing Dynamic Investment Strategies: A Comparison of
778 Returns and Investment Policies. *Management Science*, 39(7), 856-871.

779 Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional
780 LSTM and other neural network architectures. *Neural Networks*, 18(5), 602-610.

781 Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2017). LSTM:
782 A Search Space Odyssey, *IEEE Transactions on Neural Networks and Learning Systems*,
783 28(10), 2222-22232.

784 Guerard Jr, J., B. Markowitz, H., & Xu, G. L. (2015). Earnings forecasting in a global stock
785 selection model and efficient portfolio construction and management. *International*
786 *Journal of Forecasting*, 31, 550-560.

787 Hinton, G. (2009). Deep belief networks. *Scholarpedia*, 4(5), 5947.

788 Hinton, G.E., & Salakhutdinov, R. R. (2006). Reducing the Dimensionality of Data with Neural
789 Networks. *Science*, 313(5786), 504-507.

790 Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8),
791 1735-1780.

792 Huang, C. F. (2012). A hybrid stock selection model using genetic algorithms and support
793 vector regression. *Applied Soft Computing*, 12, 807-818.

794 Jozefowicz, R., Zaremba, W., Sutskever, I. (2015). An empirical exploration of recurrent
795 network architectures. *In Proceedings of the 32nd International Conference on Machine*

796 *Learning*, 2342-2350.

797 Lachiheba, O., & Gouiderb, M. S. (2018). A hierarchical Deep neural network design for stock
798 returns prediction. *Procedia Computer Science*, (126), 264-272.

799 Långkvist, M., Karlsson, L. & Loutfi, A. (2014). A Review of Unsupervised Feature Learning
800 and Deep Learning for Time - series Modeling. *Pattern Recognition Letters*, (42), 11-24.

801 LeCun, Y., Bengio, Y., & Hinton, G. 2015. Deep Learning, *Nature*, 521(7553), 436- 444.

802 Li, J., & Xu, J. (2013). Multi-objective portfolio selection model with fuzzy random returns
803 and a compromise approach-based genetic algorithm. *Information Sciences*, 220, 507- 521.

804 Li, T., Zhang, W., & Xu, W. (2013). Fuzzy possibilistic portfolio selection model with VaR
805 constraint and risk-free investment. *Economic Modelling*, 31, 12-17.

806 Li, T., Zhang, W., & Xu, W. (2015). A fuzzy portfolio selection model with background risk.
807 *Applied Mathematics and Computation*, 256, 505-513.

808 Li, Y., Yuan, Y. (2017). Convergence Analysis of Two-layer Neural Networks with ReLU
809 Activation. *Neural Information Processing Systems Conference*.

810 Liu, H., & Loewenstein, M. (2002). Optimal Portfolio Selection with Transaction Costs and
811 Finite Horizons. *The Review of Financial Studies Summer*, 15(3), 805-835.

812 Long, W., Lu, Z., & Cui, L. (2019). Deep learning-based feature engineering for stock price
813 movement prediction. *Knowledge-Based Systems*, (164), 163-173.

814 Loureiro, A.L.D., Miguéis, V.L., & Silva, L. F. M. (2018). Exploring the use of deep neural
815 networks for sales forecasting in fashion retail. *Decision Support Systems*, 114, 81-93.

816 Kara, Y., Boyacioglu, M. A., & Baykan, O. K. (2011). Predicting direction of stock price index
817 movement using artificial neural networks and support vector machines: The sample of

818 the Istanbul Stock Exchange. *Expert Systems with Applications*, (38), 5311-5319.

819 Kao, L. J., Chiu, C. C., Lu, C. J., & Yang, J. L. (2013). Integration of nonlinear independent
820 component analysis and support vector regression for stock price forecasting.
821 *Neurocomputing*, 99, 534- 542.

822 Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. CoRR, arXiv
823 preprint arXiv:1412.6980.

824 Kocuk, B., & Cornuéjols, G. (2018). Incorporating Black-Litterman views in portfolio
825 construction when stock returns are a mixture of normals, *Omega*, In press.

826 Kolm, P. N., Tütüncü, R., & Fabozzi, F. J. (2014). 60 Years of portfolio optimisation: Practical
827 challenges and current trends. *European Journal of Operational Research*, (234), 356-
828 371.

829 Kraus, M., & Feuerriegel, S. (2017). Decision support from financial disclosures with deep
830 neural networks and transfer learning. *Decision Support Systems*, (104), 38-48.

831 Malkiel, B. G. (2007). *A random walk down Wall Street: The time-tested strategy for successful*
832 *investing*. WW Norton & Company.

833 Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7(1), 77- 91.

834 Merton, R. C. (1969). Lifetime Portfolio Selection under Uncertainty: The Continuous-Time
835 Case. *The review of Economics and Statistics*, 51(3), 247-257.

836 Paiva, F. D. A, Cardoso, R. T. N., Hanaoka, G. P., & Duarte, W. M. (2019). Decision-making
837 for financial trading: A fusion approach of machine learning and portfolio selection.
838 *Expert Systems with Applications*, (115), 635-655.

839 Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index

840 movement using Trend Deterministic Data Preparation and machine learning techniques.
841 *Expert Systems with Applications*, (42), 259-268.

842 Pektas, A. O., & Cigizoglu, H. K. (2013). ANN hybrid model versus ARIMA and ARIMAX
843 models of runoff coefficient. *Journal of Hydrology*, 500, 21-36.

844 Rather, A. M. Agarwal, A., & Sastry, V. N. (2015). Recurrent neural network and a hybrid
845 model for prediction of stock returns. *Expert Systems with Applications*, 42, 3234-3241.

846 Rangelova, E. (2001). Disposition Effect and Firm Size: New Evidence on Individual Investor
847 Trading Activity. SSRN eLibrary.

848 Reimers, N., & Gurevych, I. (2017). Optimal Hyperparameters for Deep LSTM-Networks for
849 Sequence Labeling Tasks. CoRR, arXiv preprint arXiv:1707.06799.

850 Sadaei, H. J., Enayatifar, R., Lee, M. H., & Mahmud, M. (2016). A hybrid model based on
851 differential fuzzy logic relationships and imperialist competitive algorithm for stock
852 market forecasting, *Applied Soft Computing*, 40, 132-149.

853 Santos, A. A. P., & Tessari, C. (2012). Quantitative portfolio optimisation techniques applied
854 to the Brazilian stock market. *Revista Brasileira de Finanças*, 10, 369-394.

855 Sharpe, W. F. (1963). A Simplified Model for Portfolio Analysis. *Management Science*, 9(2),
856 277-293.

857 Shmueli, G. (2010). To Explain or to Predict? *Statistical Science*, 25(3), 289-310.

858 Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout:
859 A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine*
860 *Learning Research*, 15, 1929-1958.

861 Tanaka, H., Guoa, P., Turksen, B. (2000). Portfolio selection based on fuzzy probabilities and

862 possibility distributions. *Fuzzy Sets and Systems*, 111, 387-397.

863 Ticknor, J. L. (2013). A Bayesian regularized artificial neural network for stock market
864 forecasting. *Expert Systems with Applications*, 40 (14), 5501-5506.

865 Tobin, J. (1958). Liquidity Preference as Behavior Towards Risk. *The Review of Economic*
866 *Studies*, 25(2), 65-86.

867 Tu, J., Zhou, G. (2010). Incorporating Economic Objectives into Bayesian Priors: Portfolio
868 Choice under Parameter Uncertain. *Journal of Financial and Quantitative Analysis*, 45(4),
869 959-986.

870 Weng, B., Lu, L., Wang, X., Megahed, F. M., & Martinez, W. (2018). Predicting short-term
871 stock prices using ensemble methods and online data sources. *Expert Systems With*
872 *Applications*, 112, 258- 273.

873 Yunusoglu, M. G., & Selim, H. (2013). A fuzzy rule based expert system for stock evaluation
874 and portfolio construction: An application to Istanbul Stock Exchange. *Expert Systems*
875 *with Applications*, (40), 908-920.

876 Zbikowski, K. (2015). Using Volume Weighted Support Vector Machines with walk forward
877 testing and feature selection for the purpose of creating stock trading strategy. *Expert*
878 *Systems with Applications*, 42, 1797-1805.

879 Zhang, Y., Li, X., & Guo, S. (2018). Portfolio selection problems with Markowitz's mean-
880 variance framework: a review of literature. *Fuzzy Optimal Decision Making*, 17, 125-158.