Expert Systems With Applications 189 (2022) 116093

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Investment factor timing: Harvesting the low-risk anomaly using artificial neural networks

Philipp A. Dirkx, Thomas L.A. Heil *

Zeppelin University, Chair of Empirical Finance and Econometrics, Department of Corporate Management and Economics, Am Seemooser Horn 20, 88045 Friedrichshafen, Germany

A R T I C L E I N F O

JEL classification:
C58
C63
E37
G11
G17

Keywords:
Factor investing
Low-risk anomaly
Neural networks
Long-short term memory

A B S T R A C T

We perform investment factor timing based on risk forecasts exploiting the low-risk anomaly. Among various risk measures, we find downside deviation most suited for this task. We apply Long Short Term Memory Artificial Neural Networks (LSTM ANNs) to model the relationship between macro-economic as well as financial market data and the downside deviation of factors. The LSTM ANNs allow for complex, non-linear long-term dependencies. We use LSTM-based forecasts to select high- and low-risk factors in setting up an investment strategy. The strategy succeeds in differentiating positive from negative yielding factor investments, and an accordingly constructed investment strategy outperforms every factor individually as well as LASSO and Multilayer Perceptron neural network benchmark models.

1. Introduction

Over the last two decades, a vast amount of research on well-performing investment factors has been published. Harvey and Lui (2019) document over 500 different investment factors, many of which are supposedly correlated and based on data mining. However, even if investment factors have a thorough foundation, it is questionable if investment factor returns can be expected to be stable over time. Analyzing the four renowned (Fama & French, 2015) long-short equity factors, the 10-year period between 2010 and 2019 marked the worst average performance for the investment factors, emphasizing that future investment factor returns do not necessarily resemble the past. Furthermore, all four investment factors realized returns below their historical averages over the ten years, showing that even diversifying across various investment factors would have led to low returns. To overcome sub-par results, one can use an investment factor timing approach to only focus on well-performing factors.

This paper's focus lies on innovating investment factor timing by using the Fama and French (2015) factors, building on existing literature and machine learning technology. However, instead of forecasting actual returns, we forecast risk. Risk has advantageous statistical characteristics when it comes to forecasting. Furthermore, the well-documented low-risk anomaly describes the negative correlation between risk and return, which can be used to put risk forecasts into context with returns. Prior research provides an extensive range of market- and macro-economic predictors to forecast investment factor risk. The complex, possibly non-linear relationships between the investment factors' risk metrics and the market- and macro-economic variables are modeled with Long-Short-Term-Memory (LSTM) Artificial Neural Networks (ANNs). Such an approach is required as it enables us to capture important, recurring dynamics to find and understand investment factor returns and risk characteristics. To objectify the results, we calculate a range of LSTM ANN models, each recalibrating with a different frequency with training data starting in 1994. The LSTM-based risk forecasts allow us to construct long-only investment factor timing strategies, allocating to the investment factor with the lowest risk forecast. Furthermore, we also design long-short investment factor timing strategies, which buy the investment factor with the lowest risk and, additionally, short the investment factor with the highest risk. The long-short investment factor timing strategies highlight the performance difference between low- and high-risk investment factors and achieve monthly average returns of 0.25% to 0.36% while being equity market neutral. Those results compare well to a buy-and-hold investment into any of the investment factors as well as to GARCH, Multilayer Perceptron (MLP) neural network, and LASSO benchmark models.

To the best of our knowledge, we are the first to use LSTM ANNs to perform factor timing and use risk instead of return forecasts for factor

* Corresponding author.
E-mail addresses: philipp.dirkx@oddo-bhf.com (P.A. Dirkx), thomas.heil@zu.de (T.L.A. Heil).

https://doi.org/10.1016/j.eswa.2021.116093
Received 27 April 2020; Received in revised form 6 September 2021; Accepted 12 October 2021
Available online 24 October 2021
0957-4174/

timing. Therefore, our contribution to literature is threefold. First, we contribute to the computer science literature by building a bridge between financial markets and neural networks by analyzing a specific and relevant practical use case of LSTM ANN application to stock-related risk forecasting. Second, we contribute to the finance literature by putting the focus on risk forecasting instead of returns in the context of factor timing. Thirdly, we show that factor timing is possible utilizing powerful neural networks. We combine soft computing with a clear investment strategy to build an expert system to find a low-cost solution to the complex question if factor timing is possible.

The paper proceeds as follows: Section 2 provides a literature review regarding investment factors and factor timing, risk forecasting, the low-risk anomaly, and neural networks in financial markets forecasting. Section 3 covers our research methodology and reasons, which risk metric and model we use, and how the investment factor timing strategy is put into place. Section 4 describes our model's dependent and independent variables, and Section 5 highlights our model specifications. Section 6 describes the results of our research, and Section 7 compares our results to benchmark models. Section 8 concludes.

2. Literature review

This section provides a literature overview of all essential topics for this paper. Here, we explain what investment factors are and how timing strategies can help to improve returns. Furthermore, we elaborate on why forecasting risk leads to more promising results than raw return forecasts and how the low-risk anomaly helps put risk and returns into context. Lastly, we review the suitability of ANNs for financial markets forecasting tasks.

2.1. Investment factors and timing

Sharpe (1964) set the basis for identifying factors and notes that until the writing of his article, no one has developed a thorough theory on the relationship between the return of a single asset and its risk. Until then, researchers published articles on normative models, which focus on asset choice under conditions of risk. While diversification was already known to reduce overall risk, the relationship between return and the individual risk lacks thorough research. His Capital Asset Pricing Model (CAPM) is a univariate linear regression model, which includes one risk factor, the market factor. The model describes the expected return of a stock with the risk factor market beta, which measures that stock's risk sensitivity to the overall market. After discovering the CAPM, an extensive range of risk factors, sometimes described as the factor-zoo (Cochrane, 2011), has been identified over the following decades. The nature of the risk factor discovery lies in their predictive power of stock returns. For example, the value factor relates to the price-book-ratio's predictive power for future stock returns, where stocks with a low price-book-ratio historically outperform stocks with a high price-book-ratio significantly. Suppose the outperformance of cheap over expensive stocks cannot be explained by an asset pricing model like the CAPM. In that case, the discovered factor is added to the asset pricing model as an additional risk factor to enhance the model's explanatory power. The market factor covers the market's directionality, whereas the additional risk factors are constructed as long-short portfolios to only focus on the specific characteristics. Here the long leg of the portfolio consists, e.g., of stocks with a low price-book-ratio, and the short leg of the portfolio consists of stocks with a high price-book-ratio, therefore being equity market neutral. On the one hand, the discovered factors relate to investable long-short equity portfolios with high risk-adjusted long-term returns, based on which the term "factor investing" is coined. On the other hand, those factors function as risk factors in an asset pricing model to explain the returns of a security or a basket of securities.

However, most of these factors have been discovered recently, so the track records and success of those corresponding factor investing strategies mainly exist on paper and can only be attributed to backtests. Moreover, the strong sub-par performance of the four (Fama & French, 2015) investment factors between 2010 and 2019 puts a question mark on the sustainability of investment factor performance over time. In reality, investment factor returns are time-varying.[1] Bender, Sun, Thomas, and Zdorovtsov (2018) argue that investment factors underperform during times because, without temporal payoff variation, the underlying phenomenon would be arbitraged away. Hodges, Hogan, Peterson, and Ang (2017) explain that the compensation for a specific risk explains investment factor returns. While the risk for each investment factor is different, the returns are cyclical. Therefore, there have been various studies to test whether investment styles or investment factors can be timed. On the one hand, studies focused on market-based input metrics to time investment factors. For example, those studies analyzed the relationship between investment factor returns and the VIX, momentum, insider transactions or valuation[2] (Asness, Friedman, Krail, & Liew, 2000; Chen & De Bondt, 2004; Copeland & Copeland, 1999; Desrosiers, L'Her, & Plante, 2004; Knewtson, Sias, & Whidbee, 2010). On the other hand, studies focused on timing investment factors based on macro variables. Here, macro variables and investment factor returns are usually put in a linear relationship to set up a forecasting framework (Bender et al., 2018; Bird & Casavecchia, 2008; Blin, Ielpo, Lee, & Teiletche, 2018; Hodges et al., 2017; Miller, Li, Zhou, & Giamouridis, 2015; Nalbantov, Bauer, & Sprinkhuizen-Kuyper, 2006).

2.2. Forecasting risk

Kuremoto, Kimura, Kobayashi, and Obayashi (2014) show that ANNs, especially deep belief networks, are a useful tool to forecast time series, even with chaotic behavior. We focus on forecasting risk, as stock returns lack predictability due to missing sustainability. Moreover, stock markets are characterized as random walks, which was first mentioned by Bachelier (1900). The stochastic nature of returns necessitates the use of volatility and other risk measures, which tend to have long term dependencies and possess a smaller range of outcomes than returns. Furthermore, risk exhibits a certain degree of autocorrelation and tends to cluster. Hence the risk is better forecastable. Since the preliminary work of Markowitz (1959), the awareness of risk in financial markets changed dramatically and revolutionized financial research. The modeling of volatility defines large parts of the empirical financial research for the last 30 years. The statistical foundation to measure and forecast volatility with Autoregressive Conditional Heteroskedasticity (ARCH) models from Engle (1982) and the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model by Bollerslev (1986) is still up-to-date. The initial models are accompanied by a broad range of other volatility measurements, such as implied volatility, model-free volatility, and realized volatility, well documented by Granger and Poon (2003) in a forecasting context. However, returns and potential changes in financial products' dependencies cannot be measured in infinitesimal-small time intervals. Moreover, the true volatility as a measure of variation from a financial product is, therefore, a latent process (Andersen, Bollerslev, Diebold, & Ebens, 2001). The financial crisis of 2008 shows that reasonable measurements and forecasts of volatility are a crucial element for financial decision making. Hence, the appropriate measure of future risk through volatility forecasting is of utmost importance. Especially from the standpoint that financial assets' returns are stochastic processes driven by shocks, beliefs, and irrational behavior, it is critical to not base financial decision making (solely) on return forecasts. Inversely, Lochstoer and Muir (2019) state that

[1] Table 7 shows that investment factor returns and volatility structures changed since the financial crisis.

[2] We provide a table of terminology in the Appendix explaining some finance terms and abbreviations.

volatility provides knowledge about expected returns due to their interdependence caused by price changes. Andersen and Bollerslev (1998) show that standard volatility models such as GARCH provide accurate forecasts in the short-term. Moreover, successful risk forecasting is supported by evidence that macroeconomic states and financial activity help predict stock volatility, as discussed by Schwert (1989).

2.3. The low-risk anomaly

To link the volatility to their respective returns, we make use of the low-risk anomaly, which has already been discovered in the 1970's. While Black and Scholes (1972) state that the CAPM assumption of a linear relationship between risk and return does not hold, Haugen and Heins (1975) are the first to recognize and describe the anomaly. They state: "Stock portfolios with lesser variance in monthly returns have experienced greater average returns than their 'riskier' counterparts" (Haugen & Heins, 1975), p. 782). The launch of extended asset pricing models, like the three-factor model of Fama and French (1992), emphasizes the rejection of the CAPM when controlling for size. Blitz and van Vliet (2007) conduct an international study and find an inverted risk-return relationship, while further research confirms those findings with different risk metrics (Baker & Haugen, 2012; Frazzini & Pedersen, 2014). The low-risk anomaly is also documented within other asset classes, as research confirms compelling results for various fixed income segments (de Carvalho, Dugnolle, Lu, & Moulin, 2014; Houweling & van Zundert, 2017; Israel, Palhares, & Richardson, 2018) and mutual funds (Jordan & Riley, 2015). Existing research suggests that the low-risk anomaly could also be applied to selecting investment factors, i.e., perform investment factor timing. However, instead of using mid- or long-term historical risk, we forecast the risk for the period $T+1$.

2.4. Neural networks in financial market forecasting

To account for structural breaks, asymmetries, and possible changing volatility regimes, the original GARCH model by Bollerslev (1986) is transformed into numerous extensions. However, the incorporation of these nonlinearities tends to exacerbate the estimation procedure. Therefore, we use LSTM ANNs to account for possible nonlinearities, noisy structures, and long-term and recurring dependencies. Using LSTM ANNs also avoids the problem of model misspecification.

The dynamic structure of the financial return series' creates a certain unpredictability. Moreover, the intricate modeling of these complex structures is cumbersome. Therefore, we exploit the idea of soft computing in the sense of Zadeh (1994), where models utilize the limits of uncertainty to gain knowledge from understandable solutions with low computation cost. In that sense, Ibrahim (2016) explains the usefulness of soft computing to find answers for real-life problems through the approximation of models. Soft computing defines a wide range of models and ideas to approximate complex situations through models such as ANN, fuzzy logic, fuzzy neural networks, and many more. The idea of exploiting soft computing in financial markets is well summarized in the survey study by Atsalakis and Valavanis (2009), where they state that soft computing techniques provide excellent tools for highly volatile stock market series. We use the idea of soft computing to approximate uncertain and highly volatile financial market relationships. Moreover, the topic of factor timing shows to be hard to capture with standard methods used in financial market research. Ultimately, we use neural networks in combination with a clear investment strategy to build an expert system to answer the question if investment factor timing is possible. The predictive power of ANN enforces the reasonable belief that these methods bring improvements, resulting in numerous researchers applying ANNs to financial markets. Khandelwal, Satija, and Adhikari (2015) efficiently forecast exchange rates by decomposing the input data in a linear- and nonlinear part, modeling the linear component with an ARIMA model and the nonlinear part with ANN. Yao, Li, and Tan (2000) use ANNs to forecast option prices and compare them to the standard Black–Scholes option-pricing model. Ghiassi, Saidane, and Zimbra (2005) present a dynamic ANN structure that outperforms classical ANN-ARIMA combinations and Khashei and Bijari (2010) shows a hybrid ANN - autoregressive integrated moving average (ARIMA) model for time series forecasting. Moreover, various articles combine classical volatility models with ANNs, like Wang (2009), who improve stock index option price forecasts by estimating the volatility with a GJR-GARCH model. Donaldson and Kamstra (1997) introduces a nonlinear ANN-GARCH model to forecast stock return volatility in different countries, showing important structural volatility differences between them. In that context, researchers also focused on S&P 500 stock volatility forecasts, like Hamid and Iqbal (2004), showing that their forecasts are not significantly different from realized volatility and Xiong, Nichols, and Shen (2016) additionally incorporate Google trend data in their LSTM model outperforming a GARCH model. Nonetheless, Fischer and Krauss (2017) circumvent the problem of unprecise point forecasts by predicting the directional movement of the S&P 500 stocks using LSTM ANN. However, especially in financial markets, the future is uncertain, and therefore neural networks are also not spared from unpredictable market fluctuations. In contrast to returns, volatilities are more persistent, containing more information, and thus, the use of neural networks for risk forecasting is perfectly suitable. Mainly the power to provide arbitrarily nonlinear function estimations makes ANNs suitable for risk forecasting. However, ANNs suffer from optimization problems in recurrent data structures. Contrary, LSTM models presented by Hochreiter and Schmidhuber (1997) circumvent these problems. Nonetheless, despite the fit of LSTM models for risk forecasting, the literature remains sparse.

3. Methodology

In the following chapter, we review the methodological aspects of our paper in general. Firstly, we discuss risk metrics and why we chose the downside deviation. Secondly, we briefly introduce the topic of LSTM models and the reasoning why we use this type of ANN. Finally, we discuss the applied investment strategy.

3.1. Risk metrics

To verify the suitability of various risk metrics for our task, we compare three metrics: variance, semi-variance, and downside deviation. While the variance is a standard risk metric, also applied in the low-risk anomaly context, Markowitz (1959) already doubted the effectiveness of variance. Markowitz (1991) restates that semi-variance "is the more plausible measure of risk" (P. 476) due to its focus on underperformance. Variance measures the dispersion of a set of data points around their mean, and semi-variance focuses on the observations below the mean. Downside deviation resembles semi-variance, but it focuses on the returns below zero instead of below the mean returns. Consequently, semi-variance and downside deviation focus on downside risk, which a rational investor tries to limit. We calculate all risk metrics monthly based on daily data. The mathematical calculation of our realized risk metrics follows:

$$V_M = \frac{1}{T_M}\sum_{t=1}^{T_M}(r_t-\mu)^2, \tag{1}$$

$$SV_M = \frac{1}{T_M}\sum_{t=1}^{T_M}(r_t-\mu)^2 \;\; for\ r_t<\mu, \tag{2}$$

$$DD_M = \frac{1}{T_M}\sum_{t=1}^{T_M}(r_t-\mu)^2 \;\; for\ r_t<0, \tag{3}$$

where r_t denotes the daily investment factor return, μ the investment factor returns mean (monthly) and T_M is equal to the number of days in the specific month. Moreover, V_M is the variance formula, SV_M the semi-variance and DD_M the downside deviation.

3.2. Choice of model

Static approaches to model the relationship between investment factors and market- and macro-economic variables are not very promising since the relationship of variables is rather dynamic. In the context of investment factor timing, Bender et al. (2018) analyze a set of market- and macro-economic predictors for the periods 1972–1989 and 1990–2020 and conclude that the statistically significant predictors are not the same during the two-time spans. Their analysis implies that it is possible to find existing linear relationships during a certain sample period. However, those relationships do not necessarily hold for other sample periods. Bender et al. (2018) also stresses that a linear approach cannot adequately account for changing regimes of those relationships. Furthermore, Levin (1995) highlights the limitations of linear models since they do not account for the interaction between the explanatory variables. Investment factors and market- and macro-economic variables are expected to have a dynamic, non-seasonal recurring dependency. Furthermore, we expect the time series of the investment factor risk to have complex recurring characteristics, which cannot be modeled by normal autoregressive processes. Due to the characteristics stated above, we want to exploit artificial neural networks as a soft computing method, proven in many research fields to be powerful, well-established tools for prediction, forecasting, or classification. For example, function approximation through soft computing, especially with artificial neural networks, is applied successful to the area of public transportation engineering, showing improvements in comparison to classical statistical models (Celikoglu, 2006; Celikoglu & Cigizoglu, 2007), or integrating several neural networks for function approximation to dynamic network loading models for simulating highway traffic dynamics (Celikoglu, 2007a, 2007b, 2011). Another example would be Guan, Huang, and Zhou (2004), who use neural networks to model the incidence of hepatitis A and find that ANNs yield higher prediction accuracy than conventional statistical methods. In a similar area of research, Rey, Arcay, and Castro (2021) use soft computing techniques to find lung nodules which are the first indicators for lung carcinoma. The idea of soft computing also finds application in engineering, as Varol, Koca, Oztop, and Avci (2010) uses soft computing techniques to forecast thermal energy storage performance. The combination of several nonlinear functions with an iterative optimization process leads to arbitrary function approximation. Therefore, ANNs provide a high capability for the generalization of the data which we require for this task. The classical, well-known (supervised) feedforward multilayer perceptron (MLP) artificial neural network is a powerful tool to find possible nonlinear patterns between the input matrix X and the output y.[3] Generally speaking, artificial neural networks estimate a functional mapping $f : X \to y$. Looking for a method that finds the unknown, complex mapping between investment factor risk and macroeconomic data, ANN perfectly fits this task. However, working with time series, a focus on sequential models is required. A particular class of neural networks for sequential data are recurrent neural networks. RNNs specialize in processing sequences of variables $Y^{(1)}, \dots, Y^{(t)}$, by introducing a recurrent connection between the nodes. Therefore, they can process sequences with a long-range while detecting repeating elements by sharing the same weights across different parts of the model.[4] Mathematically, RNNs can be expressed as

$$h_i^{(t)} = f(h_i^{(t-1)}, X^{(t)}; \theta), \tag{4}$$

where $h_i^{(t)}$ is the current state (hidden layer in MLP) of the system denoted with the node index i, $X^{(t)}\ with\ X = [x_1, \dots, x_J]$ are signals or "explaining variables" and θ are the parameters, including the weights from input to hidden, hidden to hidden, hidden to output layers and a bias term. Note that Eq. (4) describes a dynamic system, with the recurrence incorporated by the element-wise, probably the same, functional transformations of each hidden layer, and hence, by feeding the output at every step in time back into the system of transformation of the following hidden layer (the outer loop). However, if the gradient is propagated over many stages, it tends to vanish or explode (Hochreiter, 1998). Hence, standard recurrent models may fail to capture possible long term dependencies. Hochreiter and Schmidhuber (1997) present long-short term memory models as a solution, outperforming RNN on an artificial data-set for long term dependencies. LSTM models incorporate, in addition to the outer loop of RNN, a self-loop within each cell of the neural network to prevent the gradient from vanishing. LSTM neural networks are a clever combination of several gated units with a self-loop within the hidden units. The input gate, output gate, and self-loop of LSTM models replace the usual nodes of RNN and transform the hidden layer into one memory cell block. Moreover, the cells are additionally connected recurrently to each other. LSTM models are a form of RNNs, but incorporate gates and a self-loop to account for long-term dependencies efficiently. The cells are structured with a forget gate, which decides which information is forgotten (providing the self-loop weights), an input gate which decides which necessary new signals to include to the memory and after internal calculation an output gate that produces the state's output, protecting other cells from unnecessary stored memory. Therefore, the current state $s^{(t)}$ is not anymore solely a weighted combination of the recent state and additional input variables. The LSTM updating rule is:

$$s_i^{(t)} = f_i^{(t)} s_i^{(t-1)} + g_i^{(t)} \sigma\left(b_i + \sum_j^J U_{i,j} x_j^{(t)} + \sum_j^J W_{i,j} h_j^{(t-1)} \right), \tag{5}$$

with the hidden layer $h^{(t-1)}$ now as a component producing the current state. Additionally, f_t denotes the forget gate, g_i the input gate for external variables, and b, U and W the biases, input weights, and recurrent weights, respectively. Additionally, $\sigma(\cdot)$ stands for the sigmoid function creating gates by producing values between 0 and 1, and the hidden unit $h^{(t-1)}$ includes the outputs of all other cells of the previous time. Hence, $h^{(t)}$ is the output of the current cell. Moreover, the forget gate and input gate follow the same idea of multiplicative transformation, stating:

$$f_i^{(t)} = \sigma\left(b_i^f + \sum_j^J U_{i,j}^f x_j^{(t)} + \sum_j^J W_{i,j}^f h_j^{(t-1)} \right), \tag{6}$$

$$g_i^{(t)} = \sigma\left(b_i^g + \sum_j^J U_{i,j}^g x_j^{(t)} + \sum_j^J W_{i,j}^g h_j^{(t-1)} \right). \tag{7}$$

Additionally, the subsequent hidden unit $h^{(t)}$ can be gated with another sigmoid function, the output gate, ultimately producing the output:

$$h_i^{(t)} = tanh(s_i^{(t)}) q_i^{(t)}, \tag{8}$$

where $tanh$ is the hyperbolic tangent function and $q_i^{(t)}$, the output gate, is given by:

$$q_i^{(t)} = \sigma\left(b_i^q + \sum_j^J U_{i,j}^q x_j^{(t)} + \sum_j^J W_{i,j}^q h_j^{(t-1)} \right). \tag{9}$$

LSTM models circumvent the problem of long-term memory by carrying only important information with the sequence through all memory blocks. With this advantage of long-term memory in mind, we expect LSTM ANNs to generate the best results. Therefore, we apply LSTM ANNs, which provide an edge over linear approaches. LSTM ANNs improve the modeling effort due to their ability to learn dynamic and complex relationships. Generally speaking, deep learning methods excel in tasks with highly nonlinear and complex structures, whereas standard statistical tools cannot often detect these dynamics if not modeled correctly. Using neural networks' ability to approximate

[3] For this section we follow the overview on this topic by Goodfellow, Bengio, and Courville (2016) closely, as well as Hochreiter and Schmidhuber (1997).

[4] However, weight sharing is only feasible if there is no dependence between the variables within the time t.

arbitrary functions, we try to overcome the limitations of standard regression models. Moreover, ANNs shine at tasks where knowledge-based modeling fails to capture possible dependencies, which would be the challenge in the case of investment factor timing. LSTM models' utilization should enable us to capture valuable, recurring dynamics to understand investment factor return and risk characteristics better.

3.3. Investment strategy

The investment strategy aims to show that investment factor timing is possible, especially during times when average investment factor returns are challenging. Additionally, it provides a natural measure for the performance of the ANN. Every month, the strategy invests in the investment factor, which has the lowest forecasted risk for the month $T + 1$. This way, the investor limits downside risk, and in connection with the low-risk anomaly, this strategy should result in relatively high absolute as well as risk-adjusted return. In contrast, we do the opposite exercise for the investment factor, which has the highest risk for the month $T + 1$. As a result, we also construct a long-short investment factor strategy, investing long into the investment factor with the lowest risk in the month $T+1$ and short into the investment factor with the highest risk in $T+1$. This long-short investment factor strategy highlights the performance spread of low- and high-risk investment factors. Besides, we adjust the short leg weight according to the difference in risk to invest in a risk-neutral way. Mathematically that is:

$$r_{LS,t} = r_{L,t} - \left(\frac{risk_{L,t}}{risk_{S,t}}\right) r_{St}, \tag{10}$$

where LS refers to the long-short portfolio, L to the long investment factor, and S to the short investment factor. To illustrate which risk metric is applicable in the context of the low-risk anomaly, Fig. 1 shows the cumulative log performance of three long-short investment factor portfolios. In more detail, each investing long into the factor with the lowest risk and short into the factor with the highest risk, based on the T + 1 risk metrics' perfect hindsight. Furthermore, Fig. 1 also shows the cumulative average investment factor performance, which yields an average annual return of 2.3% over the 25 years. However, the last ten years show a cumulative negative return of −6.6%. Interestingly, downside deviation strongly outperforms the other two risk metrics, yielding an average annual return of 9.6%, while variance only achieves 0.3% and semi-variance is even negative with −2.2% on average. The downside deviation metric's strong outperformance is related to its characteristics and the short time frame of only 21 trading days between rebalancing. Downside deviation only considers outright negative returns, while the two other risk metrics consider the mean return. To increase the probability of achieving a positive return over a short time frame, one has to limit large negative returns. Large deviations of above zero returns can also drive a high standard deviation or semivariance. A high downside deviation is forcibly driven by returns from the far left of the return distribution. When large negative daily returns occur, a one-month period is in most cases too short to overcome the drawdown, therefore leading most likely to a negative monthly return. On the other hand, if the downside deviation is low, the 21-day period most likely did not have daily returns from the far left of the return distribution. The probability of having a positive monthly return is therefore higher. Fig. 2 uses a Monte Carlo Simulation to randomly construct 10,000 21-day time series, sourced from the SMB, HML, RMW, and CMA[5] factor time series between 1993 and 2019. For all 10,000 21-day simulations, standard deviation, downside deviation, and semivariance are calculated and grouped into quartiles on the x-axis (low to high risk). At the same time, the y-axis shows the average daily returns of those buckets. Fig. 3 performs the same exercise; however, for a hypothetical period of 260 trading days instead of 21. The two figures illustrate how low and high-risk metrics readings correlate with high and low average returns. Downside deviation and semivariance show a clear negative relationship with average returns, while the standard deviation shows no apparent connection. Fig. 2 indicates that the downside deviation shows the most considerable difference in average returns between the first and fourth quartile over 21 days. Fig. 3 highlights that the difference in average returns between the first and fourth quartile goes down when doing the same exercise for 260 days, indicating that the risk metrics have more explanatory power for short time frames than for longer ones. To conclude, the more pronounced sensitivity of downside deviation to negative returns over short time frames makes it the most suitable risk metric to differentiate between high and low returns, which is why we only forecast downside deviation in this paper.

4. Data

When having hundreds of investment factors to choose from, a specific rigidity is required to make a choice. We focus on the investment factors, which have been used by Fama and French (2015) in their five-factor model: size, value, profitability, and investment. The fifth factor is the market factor, which is excluded from our analysis. The market factor reflects the overall equity market and is, therefore, not a long-short equity factor. Forecasting the equity market's direction is an entirely different topic and exceeds the scope of this study. Due to the US's uniformity, we focus on the US market and economy in our paper. Daily and monthly US investment factor returns are obtained from Kenneth R. French's website. The literature explains investment factor premiums by the compensation for certain risks. Therefore, we use literature on risk factors explaining the investment factors' existence and literature on investment factor timing to link investment factors to market- and macro-economic variables. Much research has been done on the economic explanation of investment factor premiums. Small companies are generally more exposed to the economic cycle and are therefore more likely to suffer from refinancing- and inflation shocks (Hahn & Lee, 2006; Kelly, 2003; Nielsen, 2015; Penman, Reggiani, Richardson, & Tuna, 2018; Sarwar, Mateus, & N., 2015; Steiner, 2009; Zhang, Hopkins, Satchell, & Schwob, 2009). Cheap companies tend to have less leverage and higher dividends, which is why value stocks are less sensitive to refinancing- and inflation shocks (Gregory, Harris, & Michou, 2001; Hahn & Lee, 2006; Zhang et al., 2009). Highly profitable companies are expected to be more exposed to economic risks and changing interest rates due to the tendency of high operational leverage and long-run dependency of profit growth (Kisser, 2014; Lam, Wang, & Wei, 2014; Novy-Marx, 2013). Companies, which invest heavily, tend to be affected by the economic cycle due to debt or equity financing of investment operations (Fama & French, 2006; Titman, Wei, & Xie, 2004). While the theoretical drivers of investment factor returns are overlapping, investment factor returns have time-varying pay-offs, which come with a rather low cross-correlation. To extend the list of explanatory variables that underlie those cited papers, we also make use of other investment factor timing articles (Bender et al., 2018; Nalbantov et al., 2006) to arrive at a list of 24 market- and macro-economic variables, shown in Table 1. All explanatory data is obtained from Bloomberg, starting in January 1994 until December 2019 in monthly frequency. We lag certain variables to ensure point in time data, i.e., the inputs construct a valid model without hindsight bias. Furthermore, we feed the ANN with lags of the downside deviation, and make first, second, and third differences, squared, and cubic data-points of all independent variables, as the neural network decides which external variables explain best in specific periods. Moreover, we transform the downside deviation measure in two ways. Firstly, we transform all data points to the power of^{-1} in order to make the output and graphs more logical. Therefore, higher values are better. Secondly, we square the

[5] These are the four factors discussed in the Fama and French (2015) model. We provide further information in the Appendix.

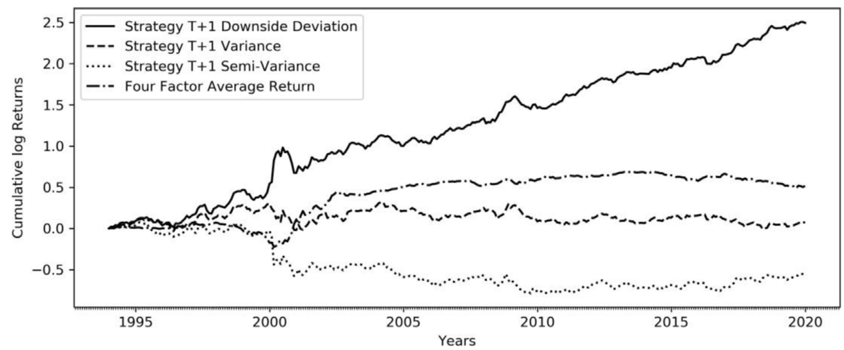

Fig. 1. Long-short investment factor timing strategies based on perfect hindsight in comparison to average investment factor. **Description:** This chart aims to illustrate, which risk metric is worthwhile forecasting in context of the investment strategy. Downside deviation, variance and semi-variance are being used, while the investment strategy always invests long into the investment factor with the lowest risk in month $T+1$ and invests short the investment factor with the highest risk in month $T+1$, while adjusting the exposure for the difference in risk. The chart shows the cumulative log returns from 12/1993 until 12/2019 in comparison to the average of the four investment factor returns.

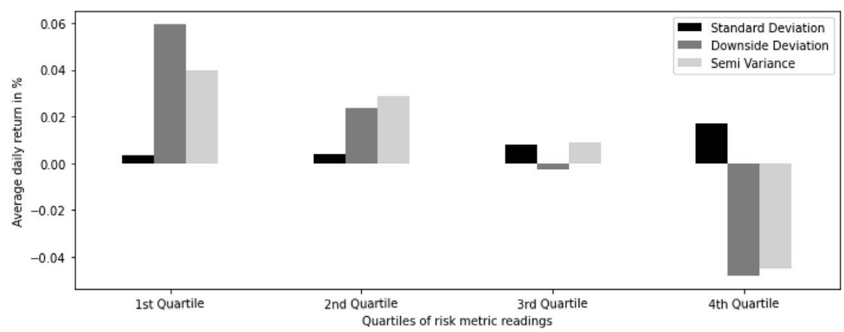

Fig. 2. Average daily returns of simulated 21-day periods. **Description:** This figure aims to illustrate, how the risk metrics correlate with average daily returns on a 21-day horizon. The analysis is based on a Monte Carlo Simulation with 10,000 bootstraps, where the returns are randomly sourced from the daily return series of the SMB, HML, RMW and CMA factors from 12/1993 until 12/2019. The average returns of those 10,000 time series are then grouped by the three risk metrics standard deviation, downside deviation and semi-variance.

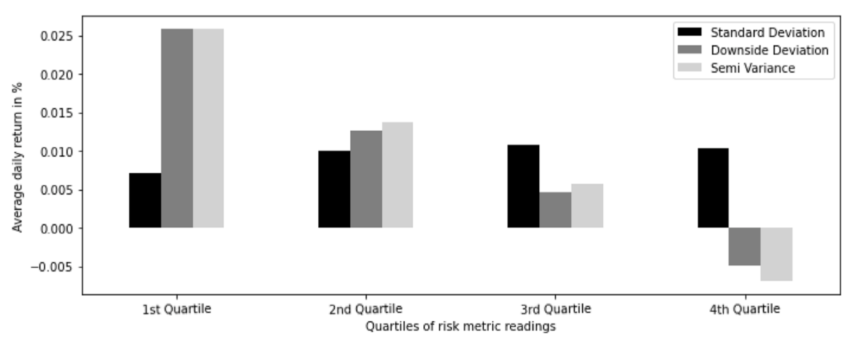

Fig. 3. Average daily returns of simulated 260-day periods. **Description:** This figure aims to illustrate, how the risk metrics correlate with average daily returns on a 260-day horizon. The analysis is based on a Monte Carlo Simulation with 10,000 bootstraps, where the returns are randomly sourced from the daily return series of the SMB, HML, RMW and CMA factors from 12/1993 until 12/2019. The average returns of those 10,000 time series are then grouped by the three risk metrics standard deviation, downside deviation and semi-variance.

reciprocal to provide a larger range of values, helping the model and investment strategy to disentangle high from low risk.

The data-set reaches from May 1994 until December 2019 with monthly frequency and is split into a training period of slightly more than 15 years, while the forecasting set is ten years. We start in May 1994 due to the availability of explanatory variables. Therefore, we train the model on 60% of the data and forecast the last 40%. Fig. 4 shows the monthly rolling average 10-year annual returns for the four (Fama & French, 2015) investment factors since the early 1960s. As mentioned before, investment factor returns appear to be somewhat cyclic, and 2019 marks the worst 10-year average investment factor performance. Therefore, we find it especially interesting to construct a model, which successfully times investment factors during the ten years between 2010 and 2019.

5. Model specification

To detect a proper hyperparameter selection for our LSTM model, we apply a step-by-step tuning procedure based on a grid-search with 3-fold cross-validation and the negative mean squared error (NMSE) loss. We use the years from 1994 to 2010 to tune the parameters accordingly. Therefore, the model never sees the forecasting period from 2010 to 2019. For each subsequent tuning, the already optimized hyperparameters are used for the next step. Specifically, we show results for selecting the batch size, number of epochs, optimizer, activation function, dropout rate, number of nodes, and number of layers. We focus on the procedure and graphs for the tuning of the LSTM for the HML investment factor risk, as the same procedure was applied to the other time-series. Moreover, the results show almost similar

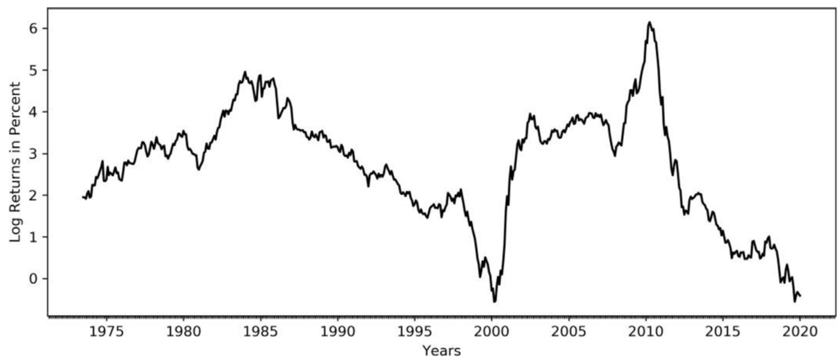

Fig. 4. 10 year rolling average annual investment factor return. **Description:** 10-year monthly rolling average annual log returns of SMB, HML, RMW and CMA from 06/1973 until 12/2019.

Table 1
Overview of independent variables.

Market variables	USD Libor3 Months
	USD Government Bonds 3 Months
	USD Government Bonds 2 Years
	USD Government Bonds 10 Years
	USD investment grade corporate credit spreads
	USD high yield corporate credit spreads
	USD VIX index
	USD index
	Gold price
	S&P 500 P/E ratio
	S&P 500 6 month return
	S&P 500 EPS YoY
Macro-economic variables	US industrial production MoM
	US retail sales MoM
	ISM manufacturing index
	Philadelphia fed business outlook survey
	US COI MoM
	US Core CPI MoM
	Personal income MOM
	AAII bullish, bearish and neutral index
	Aruoba–Diebold–Scotti business conditions index

Description: Categorization into market and macro-economic variables. All macro-economic data-points are lagged by one data-point, to ensure point in time data. YoY refers to year-over-year change and MoM refers to month-over-month change.

Table 2
Results for optimizer and activation function.

Panel A: Comparison Activation Functions

	relu	tanh	sigmoid	linear
NMSE	-3765.81	-4080.88	-4565.15	-3821.51
STD	3180.64	3565.62	3875.22	3348.78

Panel B: Comparison Optimizer

	SGD	RMSprop	Adagrad	Adadelta	Adam	Adamax	Nadam
NMSE	NaN	-4175.56	-4247.33	-5239.61	-3458.16	-3627.70	-3520.08
STD	NaN	3309.77	3566.39	4144.59	2803.73	2942.15	2999.00

Description: Negative mean squared error and Standard Deviation for the 3-fold cross-validation tuning the optimizer and activation function.

Table 3
MSE of different LSTM layer numbers.

	Epoch 1	Epoch 2	Epoch 3	Epoch 4	Epoch 5	Average
One Layer	310.28	378.29	378.29	475.11	440.39	462.38
Two Layer	341.36	301.91	661.11	430.87	522.81	451.61
Three Layer	370.15	161.47	389.34	222.56	725.72	373.85
Four Layer	336.71	575.39	349.40	628.73	668.49	511.75
Five Layer	156.15	1087.80	894.01	512.06	841.45	698.29
Six Layer	301.49	318.51	717.85	467.77	645.73	490.27

Description: Mean squared error for different numbers of layers on the last 15% of training data (End 2008 until December 2009).

results for all time-series. Therefore, and to provide the networks' uniformity, the same hyperparameters are used for all LSTM models. Firstly, we simultaneously check different batch sizes and the number of epochs as LSTM models are sensitive to those hyperparameters. We find that independent from the batch size, a smaller number of epochs provide the highest NMSE, ending in five epochs for our LSTM models. Additionally, we use a batch size of twenty. Secondly, we check for the best activation function and optimizer for the model. Table 2 shows the "Adam" by Kingma and Ba (2014) and rectified linear unit (ReLu) as the most suited optimizer and activation function, respectively. Moreover, Stochastic Gradient Descent (SGD) results in an infinite NMSE.

Thirdly, we check for a good dropout rate to ensure stability and prevent overfitting, as discussed by Srivastava, Hinton, Krizhevsky, Sutskever, and Salakhutdinov (2014). However, the grid-search for the best dropout rate was most affected by the neural network's stochastic nature, resulting in different outcomes. After several calculations of the grid-search, we find a dropout of 20% or 30% most robust. Fourthly, we try to find a well-suited number of nodes for the neural network covering the complex relationship between input and output variables. As provided in Fig. 5, peaks of the NMSE are around 100 and 250 nodes, where we decide to sacrifice a minimal higher NMSE for more nodes, resulting in 256 nodes at each layer of the LSTM. We solve weight initialization with the "He-uniform initializer" from He, Zhang, Ren, and Sun (2015). Ultimately, we split the years from 1994 to 2010 in an 85/15 train test split, having the last 15% of the months as a validation forecasting period, and choose the best number of LSTM layers on the test MSE (see Table 3)[6]. We apply a full dense layer with 200 and one layer with one node in the end to summarize the LSTM layers and project the output to one variable. Ultimately, we apply the network, as shown in Fig. 6 to forecast the risk and build the investment strategy. To increase our results' stability, we estimate every LSTM 100 times, producing an ensemble of forecasts and average over them. We base this procedure on the idea of model averaging, well documented by Claeskens and Hjort (2008). Moreover, we discuss the achieved stability with corresponding graphs further in Section 6.2.

6. Results

This section describes the statistical characteristics of the four investment factors: size, value, profitability, and investment. Afterwards, the results of the ANN LSTM downside deviation forecasts are presented, including the results of the investment strategy.

[6] However, as the other hyperparameters are tuned on a three-layer LSTM, the results are only plausible. Nonetheless, it documents that the tuning process worked.

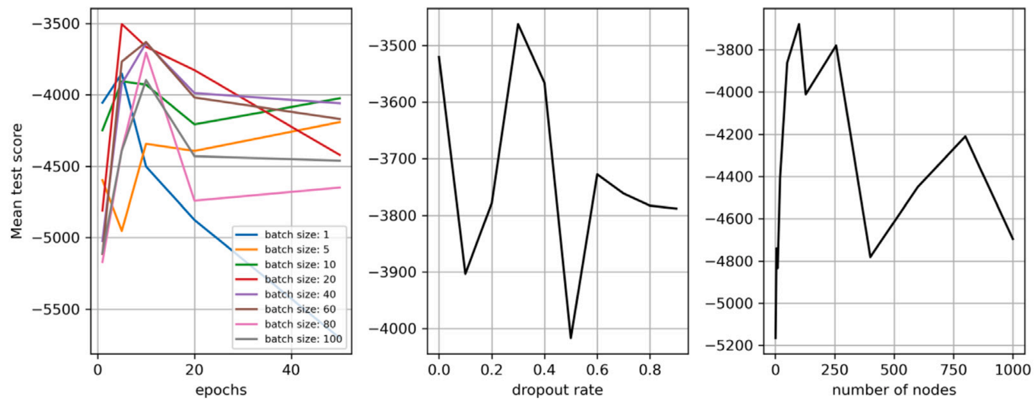

Fig. 5. Negative mean squared error graphs for hyper-parameter optimization. **Description:** These figures aims to illustrate the hyperparameter decision for the LSTM based on a 3-fold cross-validation. The loss was measured with a negative mean squared error.

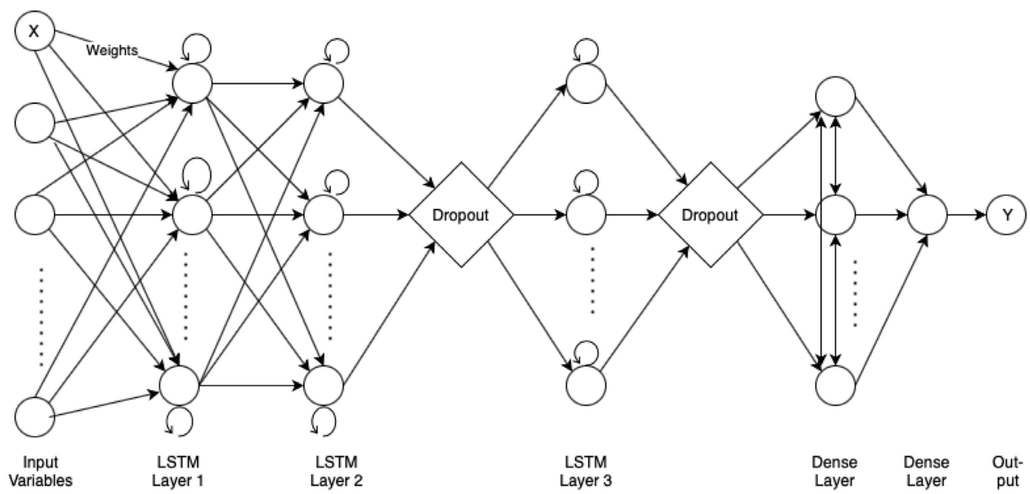

Fig. 6. LSTM model specification. **Description:** LSTM model specification, where the recurrent LSTM layer are highlighted by a round arrow pointing back to them. Furthermore, the dots present further nodes.

6.1. Summary statistics

After the specification of the hyperparameters, we split the dataset into a training dataset from May 1994 until December 2009 and introduce a forecasting dataset from January 2010 until 2019. Table 7[7] gives an overview of the descriptive statistics for the four investment factors. The average returns over the whole period are positive for all four investment factors, while RMW is the only investment factor showing a statistically significant positive return for 1994–2019 and 1994–2009, which is different from zero at a 5% level. All four investment factors report the double-digit minimum and maximum monthly returns during the training period. However, the standard deviation for all four investment factors decreased substantially during the forecasting period. Furthermore, HML and SMB show negative average returns during the forecasting period, with CMA having a flat and RMW a slightly positive average return. Fig. 18 emphasizes the break-in risk and return characteristics of investment factors after the financial crisis, as all investment factors shifted to the lower left, indicating lower variances and returns. Additionally, Table 7 provides descriptive statistics for the downside deviation of the four investment factors. The shift to a lower mean, median, and max downside deviations, as well as a lower standard deviation of the downside deviations, describe a specific change of environment. The observation of a structural break in risk and returns underlines our motivation for this paper.

6.2. Results - LSTM model

We show a different number of LSTM ANN models to calculate monthly downside deviation predictions to have an objective range of results without cherry-picking. Moreover, it helps to investigate if more available data increases the performance of the LSTM. The 1-model solution is calibrated from 1994 until 2010 and forecasts the ten following years without re-estimation. The 10-model solution also trains initially on the first fifteen years but is retrained after each new year of data, during the 10-year forecasting period. In this context, the 3- and 5-model solutions recalibrate 2 and 4 times after 2010, respectively. For illustrative purposes, we only include the figures for the 1-model solution within the text and present the other solutions in the Appendix. We overcome the stochastic nature of ANNs by calculating each LSTM model 100 times, building an ensemble, and average the results. Subsequently, we use these results as input for the investment strategy.

Fig. 7 highlights the standard deviation of predictions to show our results' stability. After each of the 100 iterations, the standard deviation of the monthly mean of all predictions is presented.[8] The 1-model solution shows the quickest convergence, while the 10-model solution needs roughly 75 iterations to reach stable results. This observation is logical since the 10-model solution is more vulnerable to stochastic

[7] The table can be found in the appendix.

[8] The solutions to the re-estimated models are stored in the Appendix.

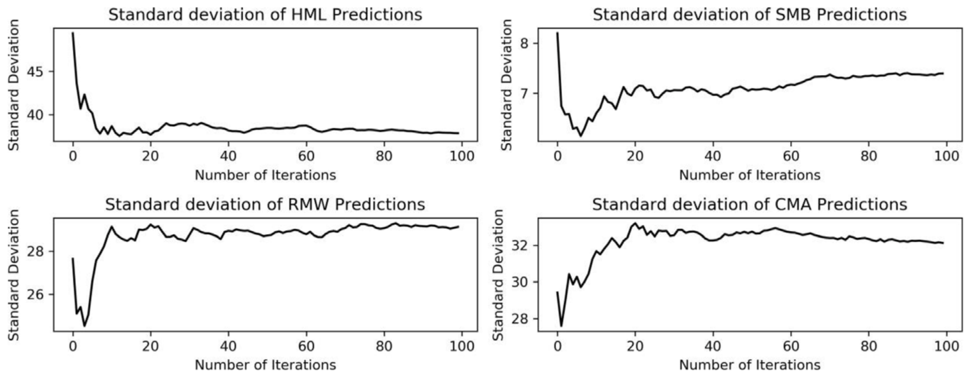

Fig. 7. Standard deviation of predictions for LSTM 1-model solution. **Description:** After every of the 100 iterations, the standard deviation of the monthly mean of all predictions is taken. One model is calibrated until 2009 and fed with data on a monthly basis to produce downside deviation forecasts between 2010 and 2019.

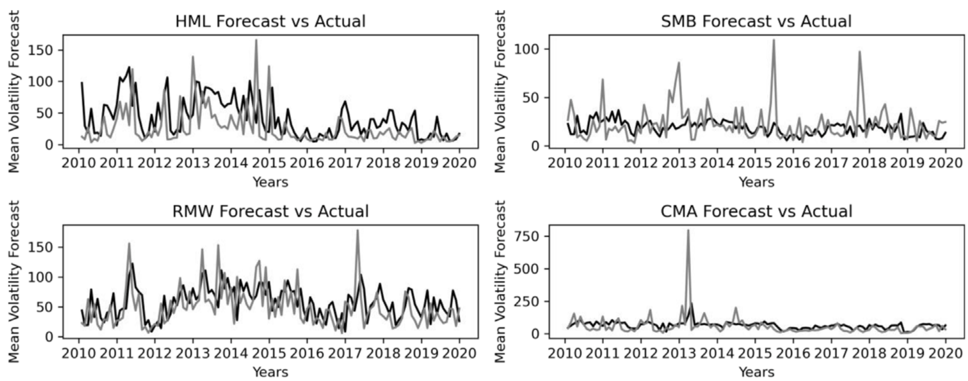

Fig. 8. Actual vs. forecasted downside deviation LSTM 1-model solution. **Description:** The graph shows the actual transformed downside deviation values in comparison to the averaged forecasts generated by the 1-model LSTM. The black and gray curve represent the actual and forecasted values, respectively.

results. However, the convergence of standard deviations shows that averaging predictions leads to more stable forecasts, which is very important for investment decisions. Fig. 8 shows the result of the forecasts of the 1-model solution plotted against the actual values.[9] We find that the LSTM 1-model solution provides good results. Especially considering the challenging nature of predicting those time series, the synchronization of actual vs. forecasted risk is satisfactory. Despite the LSTM overpredicting the forecasts for the SMB factor, we find that the application of the 1-model solution, as can be seen in Figs. 9 and 10, yields to excellent results. The excessive prediction of SMB factor risk can be explained by the changing volatility regime from the training to the testing period. The 1-model solution is calibrated based on a higher risk environment, while, e.g., the 10-model solution is adapting to the lower risk environment on a yearly basis. Fig. 9 illustrates the cumulative log-returns of an investment strategy, which invests every month into the factor with the lowest (highest) risk prediction for $T+1$.

Across all model solutions, we can see that the lowest risk strategy results at a cumulative 25% or more after ten years, while the cumulative result for the highest risk strategy is always negative. The path and the volatility of the time series differ; however, the direction is identical. We observe the most significant spread between lowest-risk and highest-risk returns for the 5-model solution, where the spread is more than 60%. Additionally, to incorporate the highest risk strategy results, Fig. 10 shows a theoretical long-short factor investment strategy, which invests long into the factor with the lowest predicted risk and short into the factor with the highest predicted risk. Furthermore, the weight of the short leg is adjusted for the difference in predicted risk. Reviewing those results, we can conclude that using LSTM ANN models to predict downside deviations to construct an investment strategy based on the low-risk anomaly leads to desirable results. Furthermore, the 1-model solution generalizes very well and achieves comparably good results. The 15 years of training data are enough for the model to understand necessary dependencies and perform reliable risk forecasts, even though the volatility environment changed during the forecasting period. Ultimately, the investment strategy shows that the LSTM model can time risk, and therefore makes factor timing possible.

7. Comparison

This section provides a comparison of the LSTM investment strategy to three benchmark models and the individual investment into the factors. Firstly, we compare the LSTM to a long investment in the factors individually. Secondly, we show the results of a classical model for volatility forecasting, the GARCH model. Thirdly, we perform a LASSO regression to forecast the downside deviation and build the investment strategy accordingly. Lastly, we show the results of a MLP NN to predict the downside deviation and the according investment strategy.

7.1. Individual investment

To compare the results of the lowest-risk strategy against the individual factors, Table 4 (Panel A) compares the results to the four investment factors individually as well as their average. All four model solutions outperform the individual, and the average investment factor returns by a wide margin. With a percentage of positive returns of at least 61%, the models also exceed the RMW investment factor, which shows the highest individual investment factor return. In terms of statistical significance, all models, except the 10-model solution, are statistically different from zero at a 10% level.

[9] The figures for the re-estimated solutions are stored in the Appendix.

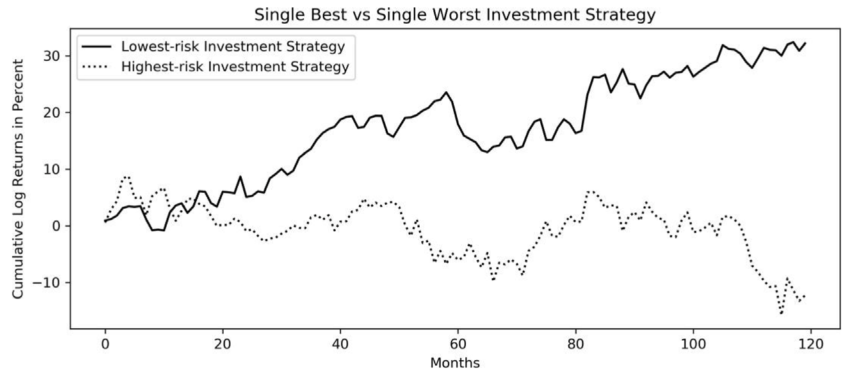

Fig. 9. Lowest- and highest-risk investment factor returns for the LSTM 1-model solution. **Description:** The strategy is based on one model, calibrated until 2009 and fed with data on a monthly basis to produce downside deviation forecasts between 01/2010 and 12/2019. The lowest-risk investment strategy resembles the investment into the factor, which has the lowest forecasted risk for month T + 1, with monthly rebalancing. The highest-risk investment strategy invests into the factor with the highest forecasted risk.

Fig. 10. Long-short investment strategy results for the LSTM 1-model solution. **Description:** The strategy is based on one model, calibrated until 2009 and fed with data on a monthly basis to produce downside deviation forecasts between 01/2010 and 12/2019. The investment strategy invests long the lowest-risk factor and short the highest-risk factor, with monthly rebalancing based on monthly risk forecasts. The exposure of the short leg is adjusted for the difference in forecasted risk.

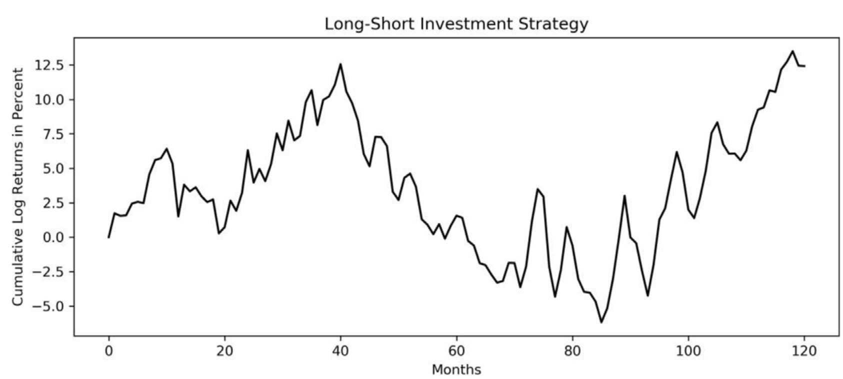

Fig. 11. GARCH(1,1)-based investment strategy. **Description:** Cumulative log returns of a mean GARCH model-based investment strategy. The strategy is based on one model, calibrated until 2009 and fed with data on a monthly basis to produce downside deviation forecasts between 01/2010 and 12/2019. The lowest-risk investment strategy resembles the investment into the factor, which has the lowest forecasted risk for month T + 1, with monthly rebalancing. The highest-risk investment strategy invests into the factor with the highest forecasted risk.

7.2. GARCH model

We train the GARCH model similar to the 1-model solution of the LSTM model, using the period from 1994 to 2010 as training data and 2010 to 2020 as the forecasting period. Using model selection metrics such as the Akaike and Bayesian information criterion to find a suited number of lags, we end up using a GARCH(1,1) with normally distributed errors. Since we include five lags in the LSTM, we also estimate a GARCH(5,5) model.

Utilizing the GARCH models for the investment strategy, we find a mixed return behavior. The results are presented in Figs. 11 and 12, while Table 5 summarizes the results further.[10] All four LSTM model long-short solutions yield average monthly returns between 0.25% and 0.35%, while the GARCH(1,1) model only achieves 0.10%. The standard deviation of returns for the four model solutions is slightly higher than for the GARCH(1,1) model. However, the 1-model and 5-model solutions show results statistically different from zero at a

10 We use a simple GARCH Model as a benchmark since our objective is to present a better performing self-learning LSTM ANN as opposed to fitting a model perfectly to the data.

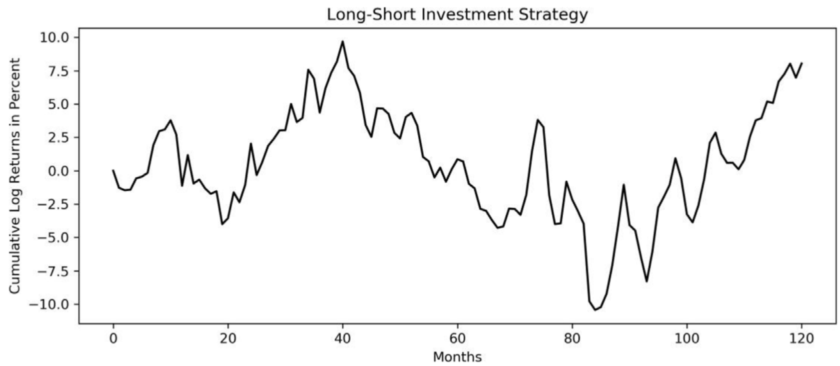

Fig. 12. GARCH(5,5)-based investment strategy. **Description:** Cumulative log returns of a mean GARCH model-based investment strategy. The strategy is based on one model, calibrated until 2009 and fed with data on a monthly basis to produce downside deviation forecasts between 01/2010 and 12/2019. The lowest-risk investment strategy resembles the investment into the factor, which has the lowest forecasted risk for month T + 1, with monthly rebalancing. The highest-risk investment strategy invests into the factor with the highest forecasted risk.

Table 4
Model comparison to individual investment factor returns.

Panel A: Comparison long-only strategy					
	SMB	HML	RMW	CMA	
Avg % return	−0.10	−0.26	0.15	0.01	
SD in %	2.04	2.31	1.37	1.48	
z - score	−0.53	−1.25	1.20	0.10	
Positive returns %	52.50	44.17	58.33	50.83	
	Average of investment factors	1-model solution	3-model solution	5-model solution	10-model solution
Avg % return	−0.05	0.26	0.25	0.29	0.22
SD in %	1.00	1.60	1.52	1.61	1.62
z - score	−0.54	1.79	1.80	1.98	1.51
Positive returns %	49.77	62.18	60.83	60.83	61.67

Description: Avg abbreviates Average and SD abbreviates Standard Deviation. SMB, HML, RMW and CMA refer to buy-and-hold investments into the factors.

Table 5
Model comparison to the GARCH(1,1).

Statistics	GARCH Model	1-model solution	3-model solution	5-model solution	10-model solution
Avg return %	0.10	0.32	0.28	0.35	0.25
SD in %	1.57	1.65	1.61	1.76	1.76
z - score	1.13	2.15	1.87	2.15	1.53
Positive returns %	55.00	60.83	60.83	60.00	59.17

Description: Avg abbreviates Average and SD abbreviates Standard Deviation. SMB, HML, RMW and CMA refer to buy-and-hold investments into the factors.

Table 6
MSE forecast vs. actual values.

Panel A: MSE LSTM forecast				
	1 model	2 models	3 models	4 models
HML	1391.66	1269.85	1119.76	1151.21
SMB	360.88	356.54	446.00	442.44
RMW	1194.02	1091.75	1176.70	1055.73
CMA	5323.62	5649.49	4959.51	5466.87
Panel B: MSE MLP forecast				
	1 model	2 models	3 models	4 models
HML	929.23	951.60	951.60	1562.27
SMB	455.56	440.14	449.61	453.43
RMW	1391.48	440.14	998.99	1013.05
CMA	7561.61	7724.52	5355.83	5771.96

Description: The table summarizes the mean squared error achieved for the LSTM and MLP.

5% level. Furthermore, all four LSTM solutions have a percentage of positive returns of at least 60%, while the GARCH model achieves onlu 55%. However, the comparison is naturally unfair, as the LSTM builds solely on predictive power, where the GARCH tries to provide inference. Therefore, we extend the comparison to a powerful regression predictor, the LASSO.

7.3. The LASSO

We feed the LASSO with the same input variables as the LSTM. Additionally, we tune the regularization parameter α with a 3-fold cross-validation on the same training data as the LSTM.

As presented in Fig. 13, the LASSO based long-short investment strategy results in a cumulative return of 25%, outperforming the GARCH based strategy, but still underperforming in comparison to the LSTM model. Moreover, it stands out that the four LSTM model solutions yield more sustainable and less volatile results compares to the LASSO model (see Fig. 14). Therefore, we increasingly support the necessity and ability of neural networks to model these complex relationships. However, to further investigate the LSTM investment strategy's performance, we ultimately build a multi-layer perceptron neural network for comparison.

7.4. Multilayer perceptron

To compare the LSTM with another neural network, we build a multi-layer perceptron with the same architecture as the LSTM model. More specifically, we produced the risk forecasts using an MLP with three fully connected dense layers and summarize the network with one dense layer with 200 nodes and one node to project the network on one output value. Additionally, the same hyperparameter specifications as the LSTM are used. Like the LSTM, we estimate different versions for the MLP to observe if additional observations increase or decrease performance.[11] Fig. 15 shows the forecasting results of the MLP ensemble for the 1-model solution. One the one hand, there is an evident lag between actual and predicted risk values. On the other hand, the MLP tends to overestimate the actual data throughout the forecasting period. This observation emphasizes that the most recent observation plays a large role for the MLP forecasts. While the results of the investment strategy are still good, the LSTM models outperform the MLP models. The LSTM provides long- and short-term memory and non-autoregressive dynamics, helping to time the factors with historical data better. Table 6 supports this finding because the LSTM forecasts are more accurate according to the MSE.

We find that the MLP produces slightly worse and less smooth returns than the LSTM, peaking on a cumulative return of 35% for

[11] The solutions to the additional models can be found in the Appendix.

Fig. 13. LASSO long-short investment strategy. **Description:** Cumulative log returns of a LASSO model-based investment strategy. The strategy is based on one model, calibrated until 2009 and fed with data on a monthly basis to produce downside deviation forecasts between 01/2010 and 12/2019. The lowest-risk investment strategy resembles the investment into the factor, which has the lowest forecasted risk for month T + 1, with monthly rebalancing. The highest-risk investment strategy invests into the factor with the highest forecasted risk.

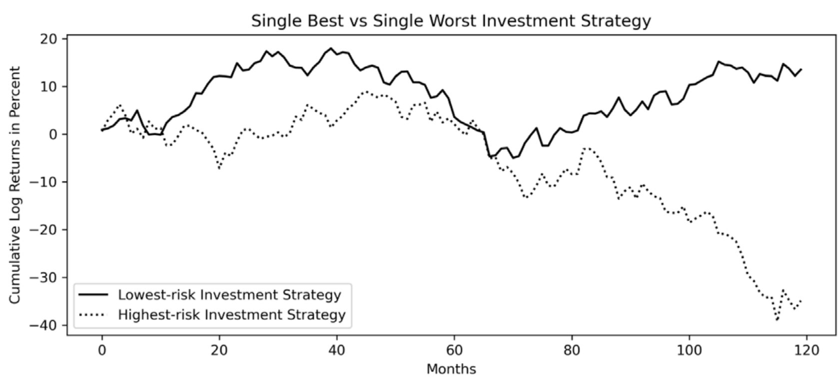

Fig. 14. LASSO best vs. worst investment. **Description:** Cumulative log returns of a LASSO model-based investment strategy. The strategy is based on one model, calibrated until 2009 and fed with data on a monthly basis to produce downside deviation forecasts between 01/2010 and 12/2019. The lowest-risk investment strategy resembles the investment into the factor, which has the lowest forecasted risk for month T + 1, with monthly rebalancing. The highest-risk investment strategy invests into the factor with the highest forecasted risk.

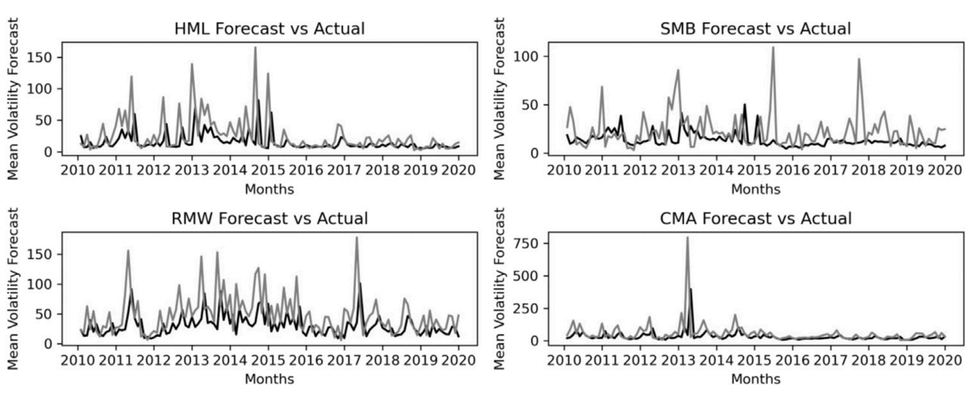

Fig. 15. Actual vs. forecasted downside deviation MLP 1-model solution. **Description:** The graph shows the actual transformed downside deviation values in comparison to the averaged forecasts generated by the 1-model MLP. The black and gray curve represent the actual and forecasted values, respectively.

the long-short strategy (see Fig. 16). However, as shown in Fig. 17 the MLP is also able to differentiate between high and low factor risk. In summary, the LSTM has slight advantages in contrast to the MLP.

8. Conclusion

Investment factor returns have been challenging over the last decade, which is why we conclude that selective exposure to investment factors is favorable. The smaller range of outcomes and the autocorrelation of risk metrics compared to stochastic returns make risk forecasting the preferred choice. Additionally, leveraging on the low-risk anomaly to perform investment factor timing shows promising results. Compared to variance and semi-variance, downside deviation is most effective in differentiating high- and low-performing investment factors. LSTM ANN models are very suitable for this forecasting task due to their ability to model complex, non-linear, and unknown long-term dependencies between investment factors and market- as well as macroeconomic variables. We achieve stable risk forecasts by averaging the prediction results over 100 iterations and use the average to select the least-risky and riskiest investment factor every month. To further objectify the results, we calculate various model solutions ranging from one model for the whole 10-year forecasting period up to recalibrating

Fig. 16. MLP long-short investment strategy. **Description:** Cumulative log returns of a MLP model-based investment strategy. The strategy is based on one model, calibrated until 2009 and fed with data on a monthly basis to produce downside deviation forecasts between 01/2010 and 12/2019. The lowest-risk investment strategy resembles the investment into the factor, which has the lowest forecasted risk for month T + 1, with monthly rebalancing. The highest-risk investment strategy invests into the factor with the highest forecasted risk.

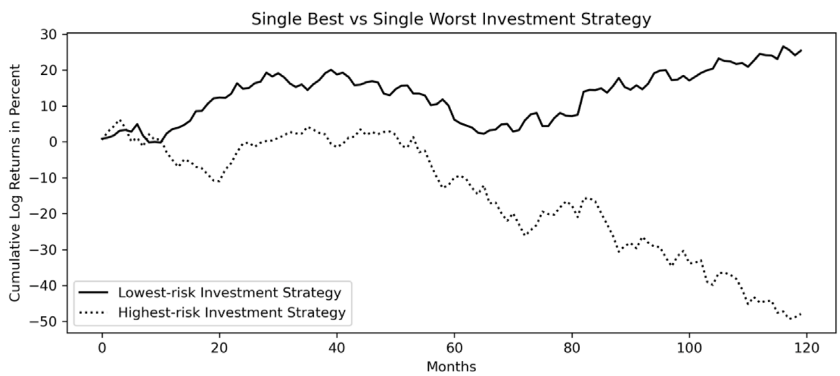

Fig. 17. MLP best vs. worst investment. **Description:** Cumulative log returns of a MLP model-based investment strategy. The strategy is based on one model, calibrated until 2009 and fed with data on a monthly basis to produce downside deviation forecasts between 01/2010 and 12/2019. The lowest-risk investment strategy resembles the investment into the factor, which has the lowest forecasted risk for month T + 1, with monthly rebalancing. The highest-risk investment strategy invests into the factor with the highest forecasted risk.

a new LSTM model every year in the ten year forecasting period. All model solutions achieve convincing results. An investment strategy investing in the lowest risk investment factor achieves at least 25% cumulative return, and an investment strategy investing in the highest risk investment factor produces negative results for all model solutions. Those results strongly outperform the average investment factor performance, which is negative during the forecasting period. Furthermore, a long-short investment strategy, investing long into the lowest risk factor and short into the highest risk factor, achieves even better results. Additionally, those results also compare well to GARCH, LASSO and MLP benchmark models. We underline the generalization power of LSTM models by the fact that the 1-model solution achieves outstanding results. Therefore, neural networks provide sufficient predictive power to perform investment factor timing.

CRediT authorship contribution statement

Philipp A. Dirkx: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Supervision, Project administration. **Thomas L.A. Heil:** Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Terminology

Legend

Terminology	Explanation
HML	Value Factor (High book-to-price stocks minus low book-to-price stocks)
SMB	Size Factor (Small stocks minus big stocks)
RMW	Profitability Factor (Highly profitable stocks minus meagerly profitable stocks)
CMA	Investment Factor (Stocks with low investment minus stocks with high investment)
VIX	Chicago Board Options Exchange's Implied Volatility Index based on S&P 500 Index Options
Momentum	Refers to the rate of change on price movements
Insider transactions	Any employee of a company, who owns more than 10% of voting rights and engages in transactions on the stock
Valuation	Analysis of the current worth of a company, often done by e.g. putting the size of the company in relation to its revenues, profits or cash flows
Cheap companies	Companies with a valuation, which is lower than the majority of the investment universe

Appendix B. Summary statistics

See Fig. 18 and Table 7.

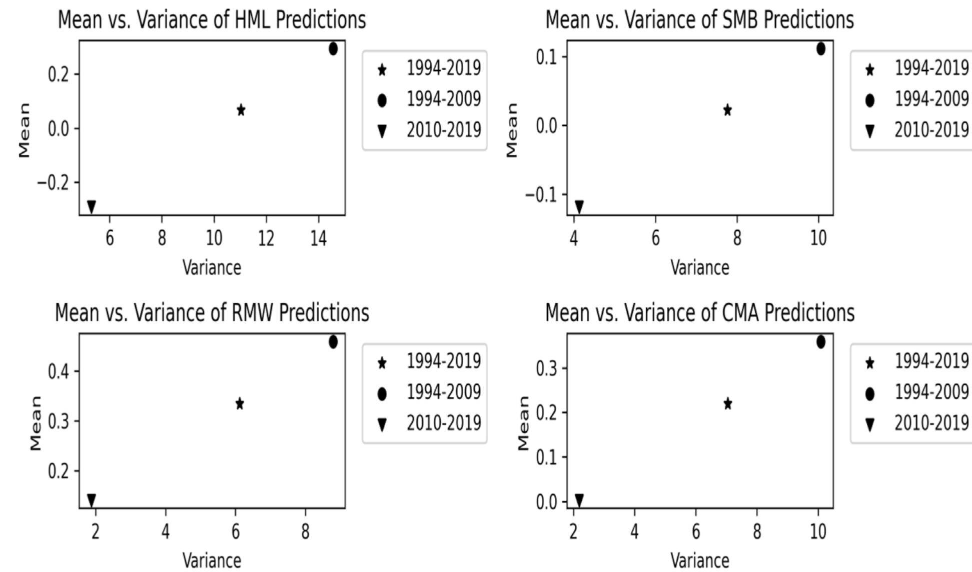

Fig. 18. Risk/return characteristics of investment factors over time. **Description:** Illustration of the risk and return relationship over three different periods: Whole period (05/1994-12/2019), training period (05/1994-12/2009) and test period (01/2010-12/2019). Calculations based on average monthly returns and monthly variances.

Table 7
Investment factor summary statistics.

Panels A: Investment factor returns					Panels B: Investment factor downside volatility				
	HML	SMB	RMW	CMA		HML	SMB	RMW	CMA
1994–2019					1994–2019				
count	308.00	308.00	308.00	308.00	count	308.00	308.00	308.00	308.00
mean	0.12	0.06	0.37	0.26	mean	0.30	0.30	0.21	0.24
std	3.36	2.80	2.48	2.70	std	0.23	0.14	0.15	0.22
z-score	0.35	0.14	2.37	1.45	z-score	23.5	36.62	25.05	19.34
min	−14.01	−13.85	−15.36	−10.71	min	0.04	0.08	0.04	0.04
25%	−1.60	−1.77	−0.69	−1.11	25%	0.17	0.20	0.12	0.13
50%	0.06	−0.02	0.34	−0.01	50%	0.24	0.27	0.16	0.18
75%	1.54	1.89	1.25	1.28	75%	0.35	0.35	0.24	0.26
max	17.28	16.68	13.95	14.58	max	1.76	1.10	1.11	1.88
1994–2009					1994–2009				
count	188.00	188.00	188.00	188.00	count	188.00	188.00	188.00	188.00
mean	0.37	0.16	0.50	0.41	mean	0.32	0.32	0.24	0.28
std	3.86	3.19	2.98	3.24	std	0.28	0.16	0.18	0.26
z-score	1.06	0.48	2.11	1.55	z-score	15.92	27.24	18.21	14.7
min	−14.01	−13.85	−15.36	−10.71	min	0.04	0.08	0.04	0.04
25%	−1.30	−1.79	−0.67	−1.19	25%	0.16	0.22	0.11	0.14
50%	0.26	−0.09	0.52	−0.04	50%	0.23	0.28	0.17	0.21
75%	1.73	2.21	1.42	1.78	75%	0.41	0.39	0.29	0.32
max	17.28	16.68	13.95	14.58	max	1.76	1.10	1.11	1.88
2010–2019					2010–2019				
count	120.00	120.00	120.00	120.00	count	120.00	120.00	120.00	120.00
mean	−0.26	−0.10	0.15	0.01	mean	0.26	0.25	0.17	0.17
std	2.31	2.04	1.37	1.48	std	0.11	0.09	0.06	0.06
z-score	−1.38	−0.64	1.12	0.02	z-score	27.25	30.8	30.89	28.51
min	−5.76	−4.91	−3.61	−3.56	min	0.08	0.10	0.07	0.04
25%	−1.85	−1.59	−0.72	−0.99	25%	0.19	0.19	0.13	0.13
50%	−0.17	0.13	0.20	0.06	50%	0.25	0.23	0.15	0.16
75%	0.88	1.31	1.05	0.89	75%	0.33	0.30	0.19	0.20
max	6.71	5.33	2.94	3.55	max	0.63	0.58	0.41	0.45

Appendix C. Stability graphs

C.1. LSTM models

See Figs. 19–21.

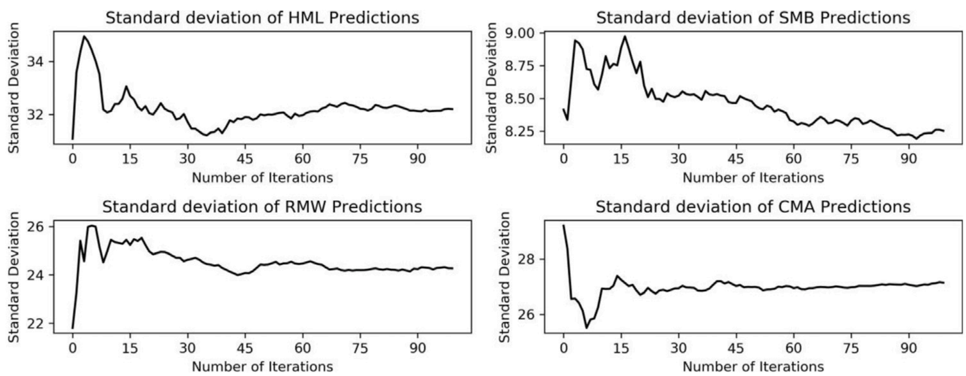

Fig. 19. Standard deviation of predictions for LSTM 3-model solution. **Description:** After every of the 100 iterations, the standard deviation of the monthly mean of all predictions is taken. The test period is split into three equal time frames, whereby the model recalibrates three times after 2009. The models are fed with data on a monthly basis to produce downside deviation forecasts between 2010 and 2019.

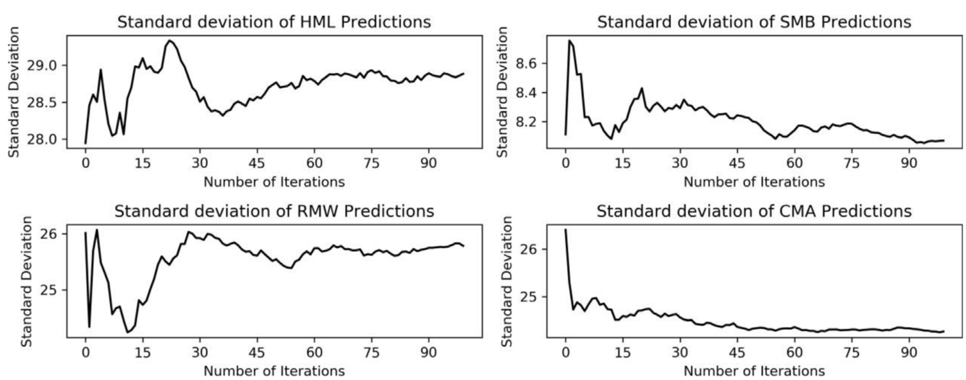

Fig. 20. Standard deviation of predictions for LSTM 5-model solution. **Description:** After every of the 100 iterations, the standard deviation of the monthly mean of all predictions is taken. The test period is split into five equal time frames, whereby the model recalibrates four times after 2009. The models are fed with data on a monthly basis to produce downside deviation forecasts between 2010 and 2019.

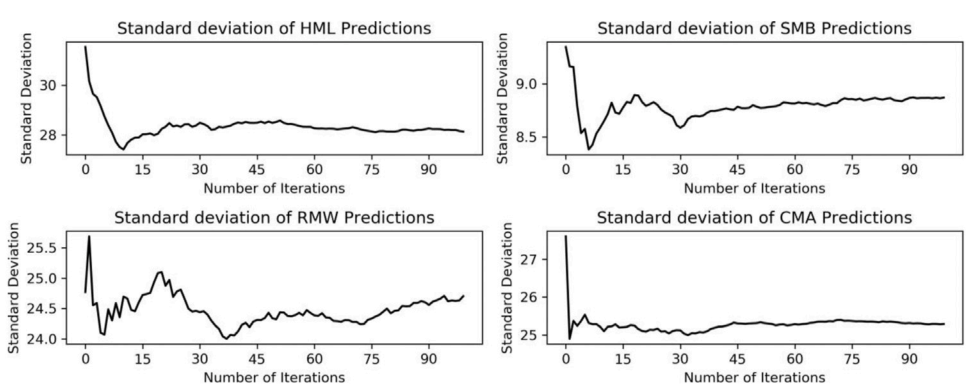

Fig. 21. Standard deviation of predictions for LSTM 10-model solution. **Description:** After every of the 100 iterations, the standard deviation of the monthly mean of all predictions is taken. The test period is split into ten equal time frames, whereby the model recalibrates every year after 2009. The models are fed with data on a monthly basis to produce downside deviation forecasts between 2010 and 2019.

C.2. MLP models

See Figs. 22–25.

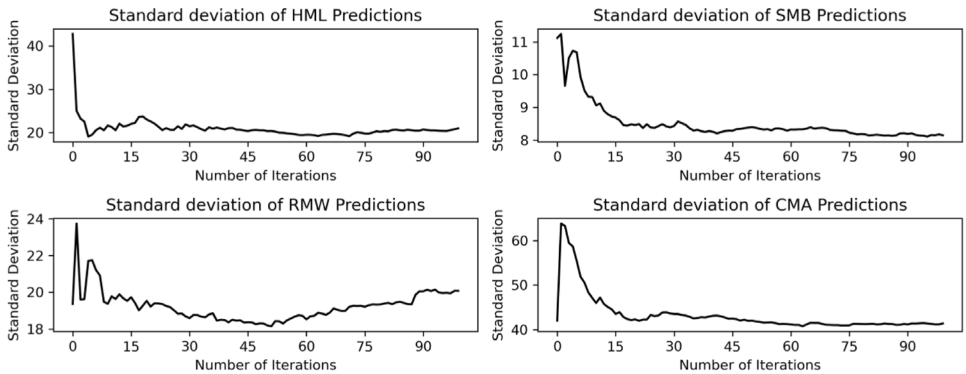

Fig. 22. Standard deviation of predictions for MLP 1-model solution. **Description:** After every of the 100 iterations, the standard deviation of the monthly mean of all predictions is taken. The test period is split into three equal time frames, whereby the model recalibrates three times after 2009. The models are fed with data on a monthly basis to produce downside deviation forecasts between 2010 and 2019.

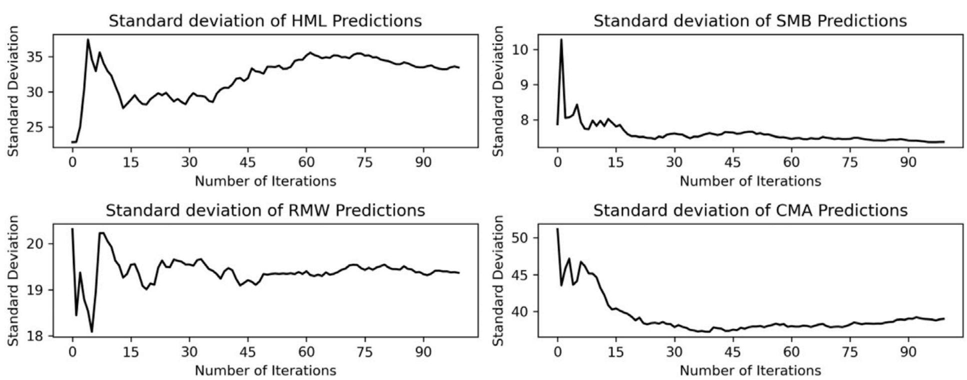

Fig. 23. Standard deviation of predictions for MLP 3-model solution. **Description:** After every of the 100 iterations, the standard deviation of the monthly mean of all predictions is taken. The models are fed with data on a monthly basis to produce downside deviation forecasts between 2010 and 2019.

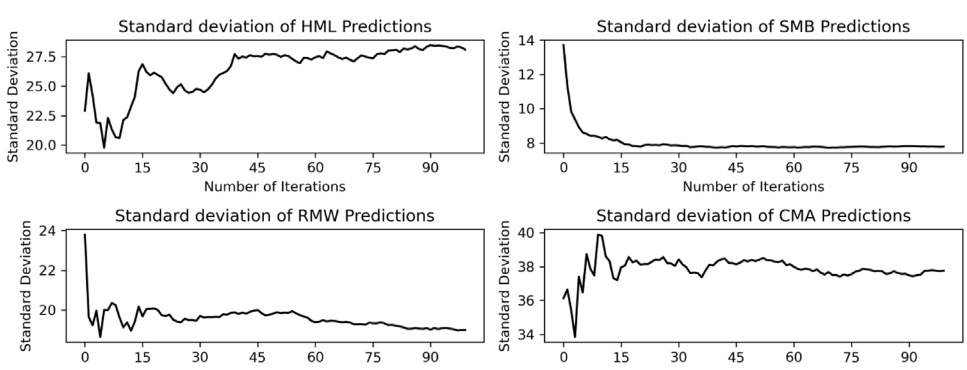

Fig. 24. Standard deviation of predictions for MLP 5-model solution. **Description:** After every of the 100 iterations, the standard deviation of the monthly mean of all predictions is taken. The test period is split into five equal time frames, whereby the model recalibrates four times after 2009. The models are fed with data on a monthly basis to produce downside deviation forecasts between 2010 and 2019.

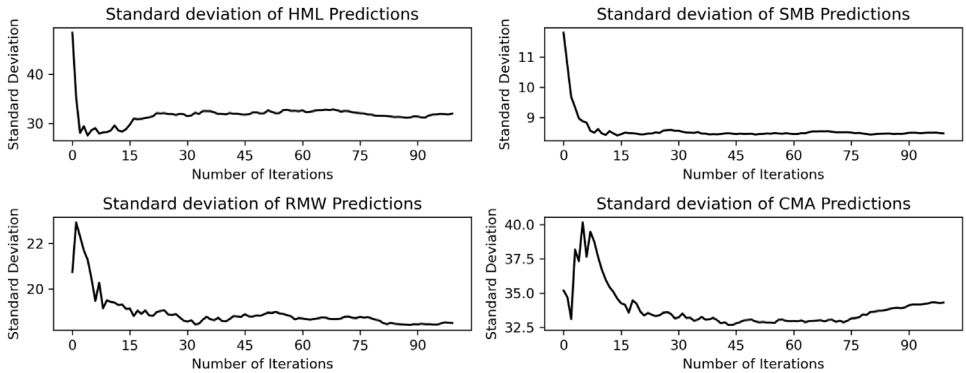

Fig. 25. Standard deviation of predictions for MLP 10-model solution. **Description:** After every of the 100 iterations, the standard deviation of the monthly mean of all predictions is taken. The test period is split into ten equal time frames, whereby the model recalibrates every year after 2009. The models are fed with data on a monthly basis to produce downside deviation forecasts between 2010 and 2019.

Appendix D. Investment strategy results

D.1. LSTM models

See Figs. 26–34.

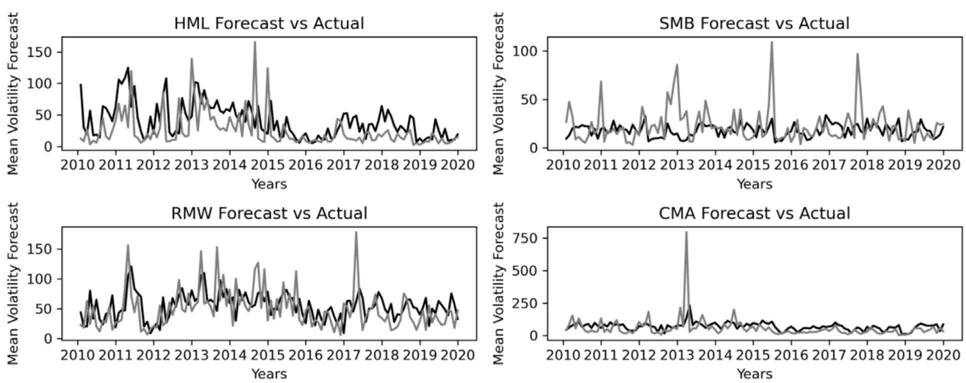

Fig. 26. Actual vs. forecasted downside deviation with LSTM 3-model solution. **Description:** The graph shows the actual transformed downside deviation values in comparison to the averaged forecasts generated by the 3-model LSTM. The black and gray curve represent the actual and forecasted values, respectively.

Fig. 27. Long-short investment strategy results for LSTM 3-model solution. **Description:** The strategy is based on three models, where the test period is split into three equal time frames, giving the model the opportunity to recalibrate two times after 2009. The models are fed with data on a monthly basis to produce downside deviation forecasts between 01/2010 and 12/2019. The investment strategy invests long the lowest-risk factor and short the highest-risk factor, with monthly rebalancing based on monthly risk forecasts. The exposure of the short leg is adjusted for the difference in forecasted risk.

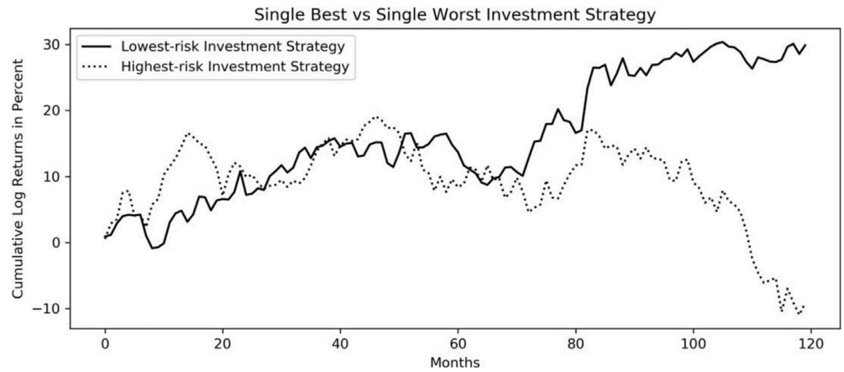

Fig. 28. Lowest- and highest-risk investment factor returns for LSTM 3-model solution. **Description:** The strategy is based on three models, where the test period is split into three equal time frames, giving the model the opportunity to recalibrate two times. The models are fed with data on a monthly basis to produce downside deviation forecasts between 01/2010 and 12/2019. The lowest-risk investment strategy resembles the investment into the factor, which has the lowest forecasted risk for month T + 1, with monthly rebalancing. The highest-risk investment strategy invests into the factor with the highest forecasted risk.

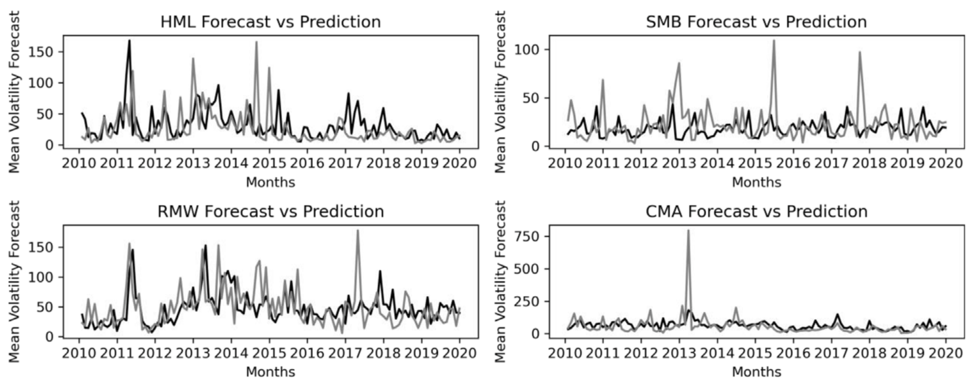

Fig. 29. Actual vs. forecasted downside deviation with LSTM 5-model solution. **Description:** The graph shows the actual transformed downside deviation values in comparison to the averaged forecasts generated by the 5-model LSTM. The black and gray curve represent the actual and forecasted values, respectively.

Fig. 30. Long-short investment strategy results for LSTM 5-model solution. **Description:** The strategy is based on five models, where the test period is split into five equal time frames, giving the model the opportunity to recalibrate four times after 2009. The models are fed with data on a monthly basis to produce downside deviation forecasts between 01/2010 and 12/2019. The investment strategy invests long the lowest-risk factor and short the highest-risk factor, with monthly rebalancing based on monthly risk forecasts. The exposure of the short leg is adjusted for the difference in forecasted risk.

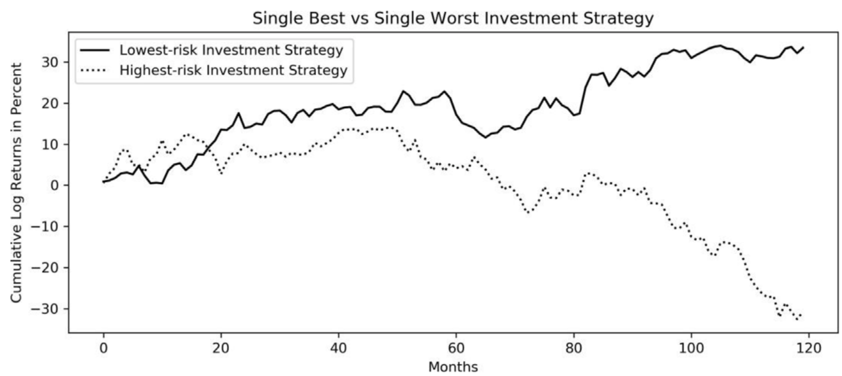

Fig. 31. Lowest- and highest-risk investment factor returns for LSTM 5-model solution. **Description:** The strategy is based on five models, where the test period is split into five equal time frames, giving the model the opportunity to recalibrate four times. The models are fed with data on a monthly basis to produce downside deviation forecasts between 01/2010 and 12/2019. The lowest-risk investment strategy resembles the investment into the factor, which has the lowest forecasted risk for month T + 1, with monthly rebalancing. The highest-risk investment strategy invests into the factor with the highest forecasted risk.

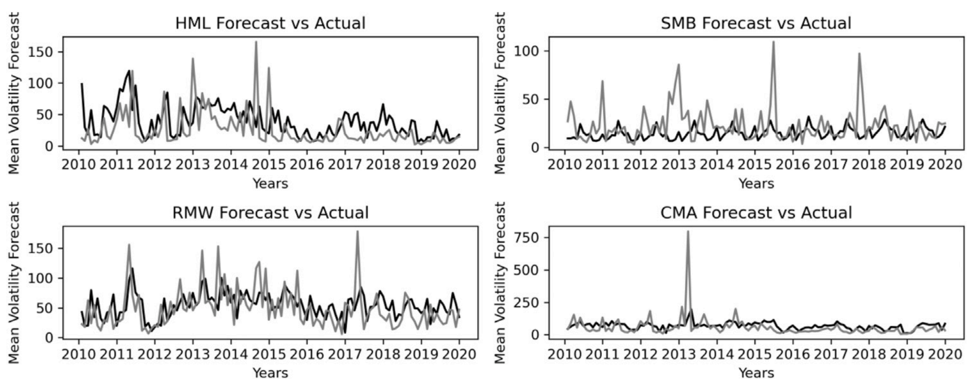

Fig. 32. Actual vs. forecasted downside deviation with LSTM 10-model solution. **Description:** The graph shows the actual transformed downside deviation values in comparison to the averaged forecasts generated by the 10-model LSTM. The black and gray curve represent the actual and forecasted values, respectively.

Fig. 33. Long-short investment strategy results for LSTM 10-model solution. **Description:** The strategy is based on ten models, where the test period is split into ten equal time frames, giving the model the opportunity to recalibrate every year. The models are fed with data on a monthly basis to produce downside deviation forecasts between 01/2010 and 12/2019. The investment strategy invests long the lowest-risk factor and short the highest-risk factor, with monthly rebalancing based on monthly risk forecasts. The exposure of the short leg is adjusted for the difference in forecasted risk.

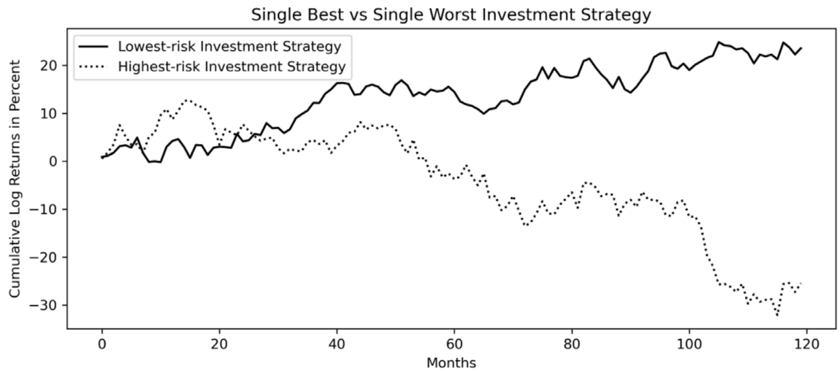

Fig. 34. Lowest- and highest-risk investment factor returns for LSTM 10-model solution. **Description:** The strategy is based on ten models, where the test period is split into ten equal time frames, giving the model the opportunity to recalibrate every year. The models are fed with data on a monthly basis to produce downside deviation forecasts between 01/2010 and 12/2019. The lowest-risk investment strategy resembles the investment into the factor, which has the lowest forecasted risk for month T + 1, with monthly rebalancing. The highest-risk investment strategy invests into the factor with the highest forecasted risk.

D.2. MLP models

See Figs. 35–43.

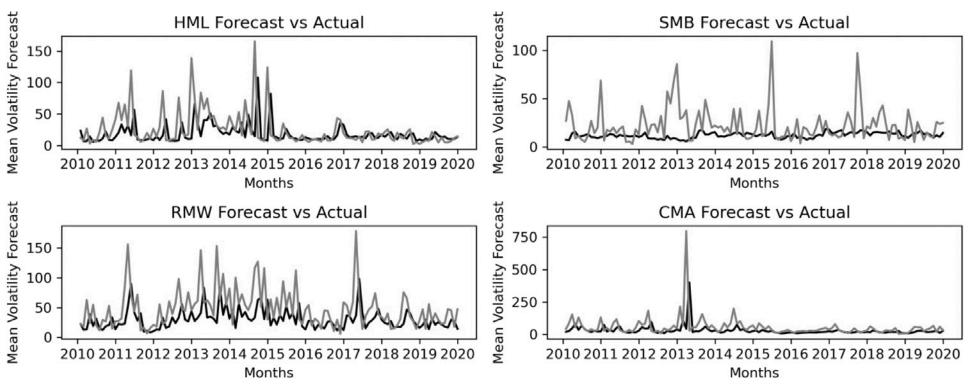

Fig. 35. Actual vs. forecasted downside deviation with MLP 3-model solution. **Description:** The graph shows the actual transformed downside deviation values in comparison to the averaged forecasts generated by the 3-model MLP. The black and gray curve represent the actual and forecasted values, respectively.

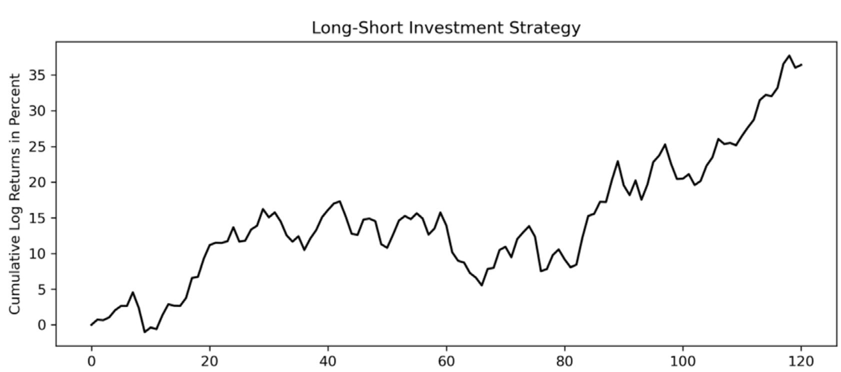

Fig. 36. Long-short investment strategy results for MLP 3-model solution. **Description:** The strategy is based on three models, where the test period is split into three equal time frames, giving the model the opportunity to recalibrate two times after 2009. The models are fed with data on a monthly basis to produce downside deviation forecasts between 01/2010 and 12/2019. The investment strategy invests long the lowest-risk factor and short the highest-risk factor, with monthly rebalancing based on monthly risk forecasts. The exposure of the short leg is adjusted for the difference in forecasted risk.

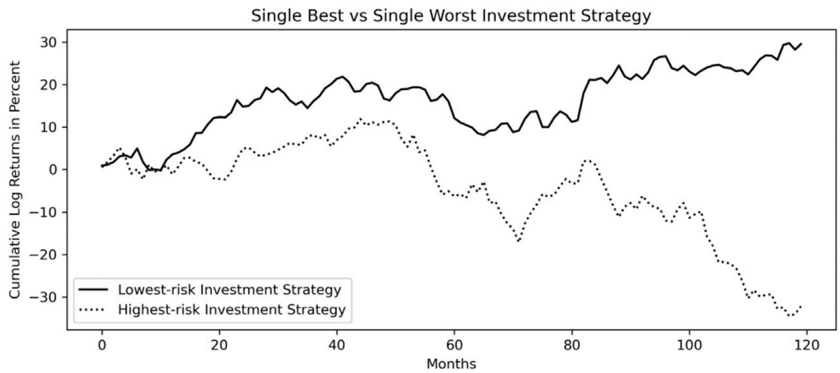

Fig. 37. Lowest- and highest-risk investment factor returns for MLP 3-model solution. **Description:** The strategy is based on three models, where the test period is split into three equal time frames, giving the model the opportunity to recalibrate two times. The models are fed with data on a monthly basis to produce downside deviation forecasts between 01/2010 and 12/2019. The lowest-risk investment strategy resembles the investment into the factor, which has the lowest forecasted risk for month T + 1, with monthly rebalancing. The highest-risk investment strategy invests into the factor with the highest forecasted risk.

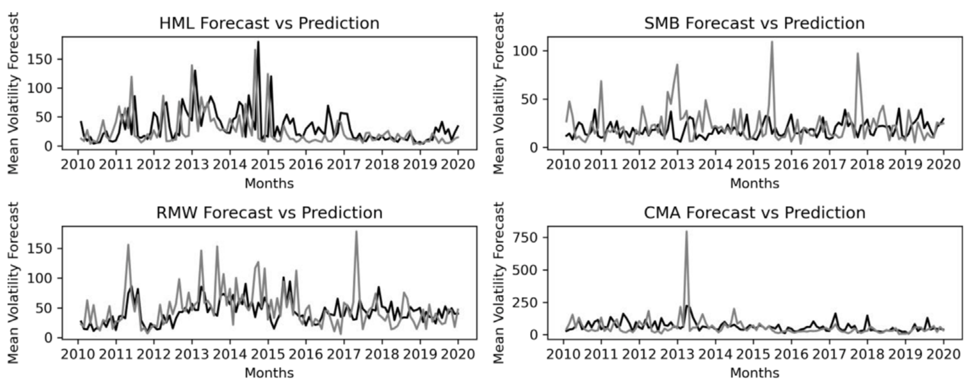

Fig. 38. Actual vs. forecasted downside deviation with MLP 5-model solution. **Description:** The graph shows the actual transformed downside deviation values in comparison to the averaged forecasts generated by the 5-model MLP. The black and gray curve represent the actual and forecasted values, respectively.

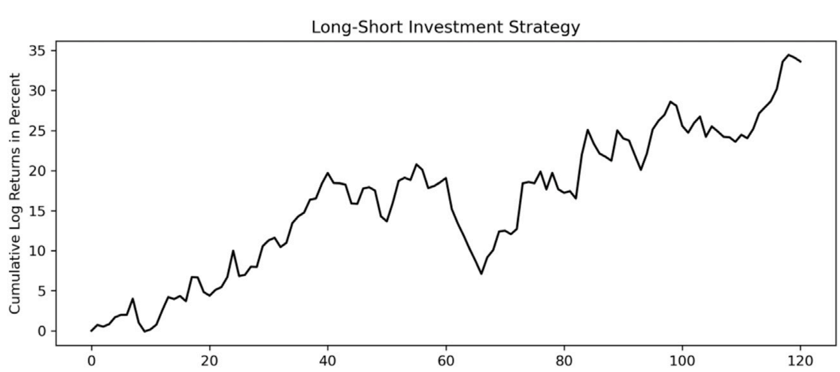

Fig. 39. Long-short investment strategy results for MLP 5-model solution. **Description:** The strategy is based on five models, where the test period is split into five equal time frames, giving the model the opportunity to recalibrate four times after 2009. The models are fed with data on a monthly basis to produce downside deviation forecasts between 01/2010 and 12/2019. The investment strategy invests long the lowest-risk factor and short the highest-risk factor, with monthly rebalancing based on monthly risk forecasts. The exposure of the short leg is adjusted for the difference in forecasted risk.

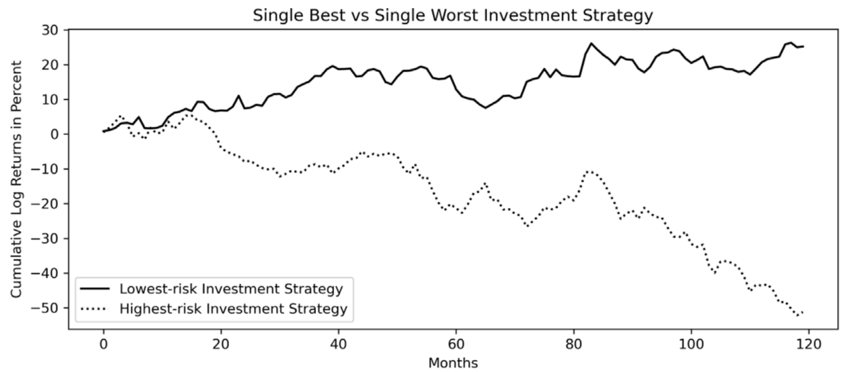

Fig. 40. Lowest- and highest-risk investment factor returns for MLP 5-model solution. **Description:** The strategy is based on five models, where the test period is split into five equal time frames, giving the model the opportunity to recalibrate four times. The models are fed with data on a monthly basis to produce downside deviation forecasts between 01/2010 and 12/2019. The lowest-risk investment strategy resembles the investment into the factor, which has the lowest forecasted risk for month T + 1, with monthly rebalancing. The highest-risk investment strategy invests into the factor with the highest forecasted risk.

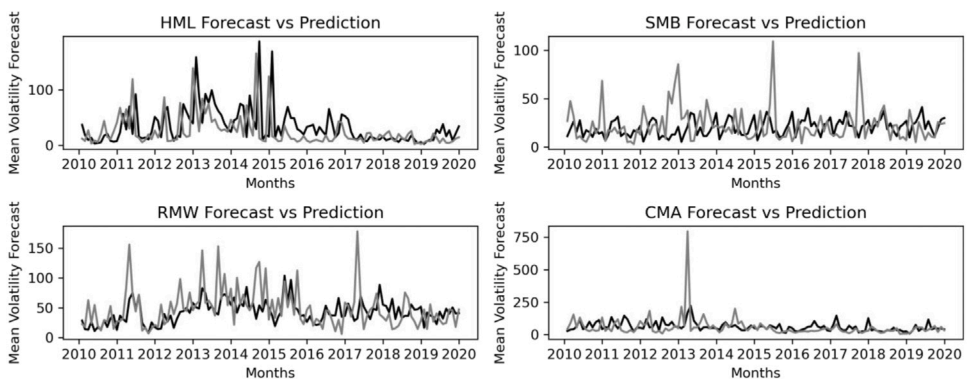

Fig. 41. Actual vs. forecasted downside deviation with MLP 10-model solution **Description:** The graph shows the actual transformed downside deviation values in comparison to the averaged forecasts generated by the 10-model MLP. The black and gray curve represent the actual and forecasted values, respectively.

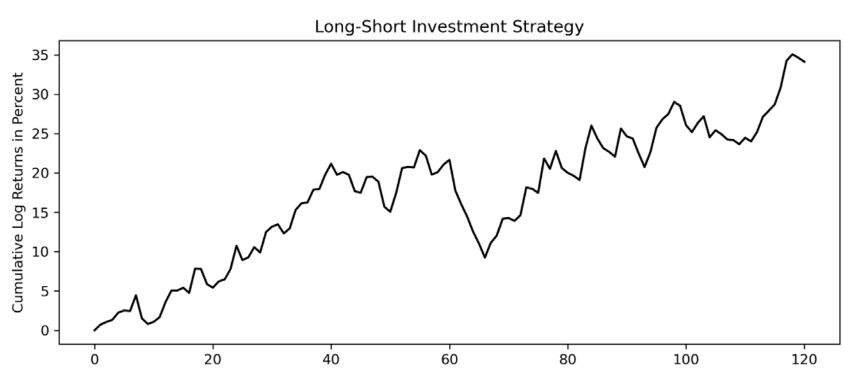

Fig. 42. Long-short investment strategy results for MLP 10-model solution. **Description:** The strategy is based on ten models, where the test period is split into ten equal time frames, giving the model the opportunity to recalibrate four times after 2009. The models are fed with data on a monthly basis to produce downside deviation forecasts between 01/2010 and 12/2019. The investment strategy invests long the lowest-risk factor and short the highest-risk factor, with monthly rebalancing based on monthly risk forecasts. The exposure of the short leg is adjusted for the difference in forecasted risk.

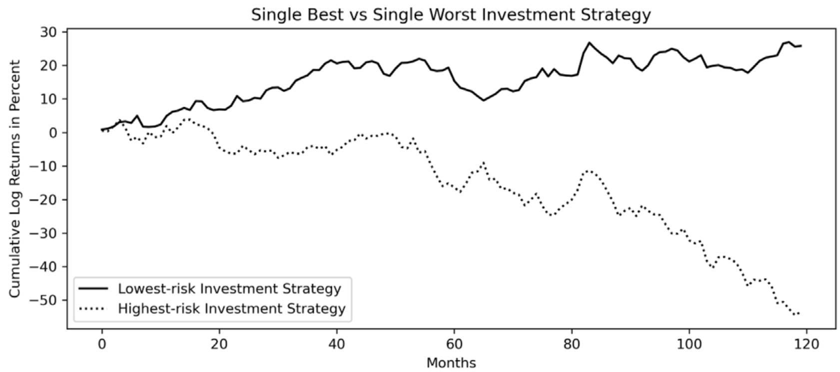

Fig. 43. Lowest- and highest-risk investment factor returns for MLP 10-model solution. **Description:** The strategy is based on ten models, where the test period is split into ten equal time frames, giving the model the opportunity to recalibrate every year. The models are fed with data on a monthly basis to produce downside deviation forecasts between 01/2010 and 12/2019. The lowest-risk investment strategy resembles the investment into the factor, which has the lowest forecasted risk for month T + 1, with monthly rebalancing. The highest-risk investment strategy invests into the factor with the highest forecasted risk.

References

Andersen, Torben, & Bollerslev, Tim (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, *39*(4), 885–905.

Andersen, Torben G., Bollerslev, Tim, Diebold, Francis X., & Ebens, Heiko (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, *61*(1), 43–76.

Asness, Clifford S., Friedman, Jacques A., Krail, Robert J., & Liew, John M. (2000). Style timing. *The Journal of Portfolio Management*, *26*(3), 50–60.

Atsalakis, George S., & Valavanis, Kimon P. (2009). Surveying stock market forecasting techniques – Part II: Soft computing methods. *Expert Systems with Applications*, *36*(3, Part 2), 5932–5941.

Bachelier, Louis (1900). Théorie de la spéculation. *Annales Scientifiques de l'École Normale Supérieure*, *3e série, 17*, 21–86.

Baker, Nardin, & Haugen, Robert (2012). Low risk stocks outperform within all observable markets of the world. *SSRN Electronic Journal*.

Bender, Jennifer, Sun, Xiaole, Thomas, Ric, & Zdorovtsov, Volodymyr (2018). The promises and pitfalls of factor timing. *The Journal of Portfolio Management*, *44*(4), 79–92.

Bird, Ron, & Casavecchia, Lorenzo (2008). Conditional style rotation model on enhanced value and growth portfolios: The European experience. *Journal of Asset Management*, *11*(6), 375–390.

Black, Michael C. Jansen, & Scholes, Myron (1972). *The capital asset pricing model: Some empirical tests*. New York: Praeger.

Blin, Olivier, Ielpo, Florian, Lee, Joan, & Teiletche, Jerome (2018). Factor timing revisited: Alternative risk premia allocation based on nowcasting and valuation signals. *SSRN Electronic Journal*.

Blitz, David C., & van Vliet, Pim (2007). The volatility effect. *The Journal of Portfolio Management*, *34*(1), 102–113.

Bollerslev, Tim (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Celikoglu, Hilmi Berk (2006). Application of radial basis function and generalized regression neural networks in non-linear utility function specification for travel mode choice modelling. *Mathematical and Computer Modelling*, *44*(7), 640–658.

Celikoglu, Hilmi Berk (2007a). A dynamic network loading model for traffic dynamics modeling. *IEEE Transactions on Intelligent Transportation Systems*, *8*(4), 575–583.

Celikoglu, Hilmi Berk (2007b). A dynamic network loading process with explicit delay modelling. *Transportation Research Part C (Emerging Technologies)*, *15*, 279–299.

Celikoglu, Hilmi Berk (2011). Travel time measure specification by functional approximation: application of radial basis function neural networks. *Procedia - Social and Behavioral Sciences*, *20*, 613–620.

Celikoglu, Hilmi B., & Cigizoglu, Hikmet K. (2007). Public transportation trip flow modeling with generalized regression neural networks. *Advances in Engineering Software*, *38*(2), 71–79.

Chen, Hsiu-Lang, & De Bondt, Werner (2004). Style momentum within the S&P-500 index. *Journal of Empirical Finance*, *11*(4), 483–507.

Claeskens, Gerda, & Hjort, Nils Lid (2008). *Cambridge books, Model selection and model averaging*. Cambridge University Press.

Cochrane, John H. (2011). Presidential address: Discount rates. *The Journal of Finance*, *66*(4), 1047–1108.

Copeland, Maggie M., & Copeland, Thomas E. (1999). Market timing: Style and size rotation using the VIX. *Financial Analysts Journal*, *55*(2), 73–81.

de Carvalho, Raul Leote, Dugnolle, Patrick, Lu, Xiao, & Moulin, Pierre (2014). Low-risk anomalies in global fixed income: Evidence from major broad markets. *The Journal of Fixed Income*, *23*(4), 51–70.

Desrosiers, Stéphanie, L'Her, Jean-François, & Plante, Jean-François (2004). Style management in equity country allocation. *Financial Analysts Journal*, *60*(6), 40–54.

Donaldson, R. Glen, & Kamstra, Mark (1997). An artificial neural network-GARCH model for international stock return volatility. *Journal of Empirical Finance*, *4*(1), 17–46.

Engle, Robert (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*(4), 987–1007.

Fama, E. F., & French, K. (1992). The cross-section of expected stock returns. *The Journal of Finance*, *47*(2), 427–465.

Fama, Eugene F., & French, Kenneth (2006). Profitability, investment and average returns. *Journal of Financial Economics*, *82*(3), 491–518.

Fama, Eugene F., & French, Kenneth R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1–22.

Fischer, Thomas, & Krauss, Christopher (2017). *Deep learning with long short-term memory networks for financial market predictions*: *FAU Discussion Papers in Economics 11/2017*, Friedrich-Alexander-Universität Erlangen-Nürnberg, Institute for Economics.

Frazzini, Andrea, & Pedersen, Lasse Heje (2014). Betting against beta. *Journal of Financial Economics*, *111*(1), 1–25.

Ghiassi, M., Saidane, H., & Zimbra, D. K. (2005). A dynamic artificial neural network model for forecasting time series events. *International Journal of Forecasting*, *21*(2), 341–362.

Goodfellow, Ian J., Bengio, Yoshua, & Courville, Aaron (2016). *Deep learning*. Cambridge, MA, USA: MIT Press, http://www.deeplearningbook.org.

Granger, Clive, & Poon, Ser-Huang (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, *41*, 478–539.

Gregory, Alan, Harris, Richard D. F., & Michou, Maria (2001). An analysis of contrarian investment strategies in the UK. *Journal of Business Finance & Accounting*, *28*(9–10), 1192–1228.

Guan, Peng, Huang, De-Sheng, & Zhou, Bao-Sen (2004). Forecasting model for the incidence of hepatitis A based on artificial neural network. *World Journal of Gastroenterology*, *10*(24), 3579–3582.

Hahn, Jaehoon, & Lee, Hangyong (2006). Yield spreads as alternative risk factors for size and book-to-market. *Journal of Financial and Quantitative Analysis*, *41*(2), 245–269.

Hamid, Shaikh A., & Iqbal, Zahid (2004). Using neural networks for forecasting volatility of S&P 500 Index futures prices. *Journal of Business Research*, *57*(10), 1116–1125.

Harvey, C. R., & Lui, Y. (2019). A census of the factor zoo. In *SSRN working paper*.

Haugen, Robert A., & Heins, A. James (1975). Risk and the rate of return on financial assets: Some old wine in new bottles. *Journal of Financial and Quantitative Analysis*, *10*(5), 775–784.

He, Kaiming, Zhang, Xiangyu, Ren, Shaoqing, & Sun, Jian (2015). Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. arXiv.org - CoRR abs/1502.01852.

Hochreiter, Sepp (1998). The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *6*(2), 107–116.

Hochreiter, Sepp, & Schmidhuber, Jürgen (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Hodges, Philip, Hogan, Ked, Peterson, Justin R., & Ang, Andrew (2017). Factor timing with cross-sectional and time-series predictors. *The Journal of Portfolio Management*, *44*(1), 30–43.

Houweling, Patrick, & van Zundert, Jeroen (2017). Factor investing in the corporate bond market. *Financial Analysts Journal*, *73*(2), 100–115.

Ibrahim, Dogan (2016). An overview of soft computing. *Procedia Computer Science*, *102*, 34–38, 12th International Conference on Application of Fuzzy Systems and Soft Computing, ICAFS 2016, 29-30 August 2016, Vienna, Austria.

Israel, R., Palhares, D., & Richardson, S. A. (2018). Common factors in corporate bond returns. *Journal of Investment Management*, *16*(2).

Jordan, Bradford D., & Riley, Timothy B. (2015). Volatility and mutual fund manager skill. *Journal of Financial Economics*, *118*(2), 289–298.

Kelly, P. J. (2003). Real and inflationary macroeconomic risk and the fama and french size and book-to-market portfolios. In *EFMA 2003 Helsinki meetings*.

Khandelwal, Ina, Satija, Udit, & Adhikari, Ratnadip (2015). Efficient financial time series forecasting model using DWT decomposition. In *CONNECCT-2015, conference* (pp. 1–5).

Khashei, Mehdi, & Bijari, Mehdi (2010). An artificial neural network (p,d,q) model for timeseries forecasting. *Expert Systems with Applications*, *37*(1), 479–489.

Kingma, Diederik P., & Ba, Jimmy (2014). Adam: A method for stochastic optimization. arXiv.org - CoRR.

Kisser, Michael (2014). What explains the gross profitability premium. In *SSRN working paper*.

Knewtson, Heather S., Sias, Richard W., & Whidbee, David A. (2010). Style timing with insiders. *Financial Analysts Journal*, *66*(4), 46–66.

Kuremoto, Takashi, Kimura, Shinsuke, Kobayashi, Kunikazu, & Obayashi, Masanao (2014). Time series forecasting using a deep belief network with restricted Boltzmann machines. *Neurocomputing*, *137*, 47–56, Advanced Intelligent Computing Theories and Methodologies.

Lam, F. E., Wang, S., & Wei, C. J. (2014). The profitability premium: Macroeconomic rrisk or expectation errors? In *Financial Management Association (FMA) 2014 conference*.

Levin, A. U. (1995). Stock selection via nonlinear multi-factor models. In *Conference on neural information processing systems November 1995*.

Lochstoer, & Muir (2019). Volatility expectation and returns. In *SSRN working paper*.

Markowitz, Harry M. (1959). *Portfolio selection: Efficient diversification of investments*. Yale University Press.

Markowitz, H. M. (1991). Foundations of portfolio theory. *The Journal of Finance*, *46*(2), 469–477.

Miller, Keith L., Li, Hong, Zhou, Tiffany G., & Giamouridis, Daniel (2015). A risk-oriented model for factor timing decisions. *The Journal of Portfolio Management*, *41*(3), 46–58.

Nalbantov, Georgi, Bauer, Rob, & Sprinkhuizen-Kuyper, Ida (2006). Equity style timing using support vector regressions. *Applied Financial Economics*, *16*(15), 1095–1111.

Nielsen, C. Y. (2015). *Hidden in the factors? The effect of credit risk on the cross-section of equity returns*: *Scandinavian Working Papers in Economics No 2001-38*.

Novy-Marx, Robert (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics*, *108*(1), 1–28.

Penman, Stephen H., Reggiani, Francesco, Richardson, Scott A., & Tuna, İrem (2018). A framework for identifying accounting characteristics for asset pricing models, with an evaluation of book-to-price. *European Financial Management*, *24*(4), 488–520.

Rey, Alberto, Arcay, Bernardino, & Castro, Alfonso (2021). A hybrid CAD system for lung nodule detection using CT studies based in soft computing. *Expert Systems with Applications*, *168*, Article 114259.

Sarwar, G., Mateus, C., & N., Todorovic (2015). Macroeconomic determinants of cyclical variations in value, size, and momentum premium in the UK. In *SSRN working paper*.

Schwert, G. W. (1989). Why does stock market volatility change over time? *The Journal of Finance*, *44*(5), 1115–1153.

Sharpe, William F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, *19*(3), 425–442.

Srivastava, Nitish, Hinton, Geoffrey, Krizhevsky, Alex, Sutskever, Ilya, & Salakhutdinov, Ruslan (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*(56), 1929–1958.

Steiner, Michael (2009). Predicting premiums for the market, size, value, and momentum factors. *Financial Markets and Portfolio Management*, *23*(2), 137–155.

Titman, Sheridan, Wei, K. C. John, & Xie, Feixue (2004). Capital investments and stock returns. *Journal of Financial and Quantitative Analysis*, *39*(4), 677–700.

Varol, Yasin, Koca, Ahmet, Oztop, Hakan F., & Avci, Engin (2010). Forecasting of thermal energy storage performance of Phase Change Material in a solar collector using soft computing techniques. *Expert Systems with Applications*, *37*(4), 2724–2732.

Wang, Yi-Hsien (2009). Nonlinear neural network forecasting model for stock index option price: Hybrid GJR–GARCH approach. *Expert Systems with Applications*, *36*(1), 564–570.

Xiong, Ruoxuan, Nichols, Eric P., & Shen, Yuan (2016). *Deep learning stock volatility with google domestic trends*: *Papers*, arXiv.org.

Yao, Jingtao, Li, Yili, & Tan, Chew Lim (2000). Option price forecasting using neural networks. *Omega*, *28*(4), 455–466.

Zadeh, Lotfi A. (1994). *Soft computing and fuzzy logic, Vol. 11*.

Zhang, Qi, Hopkins, Peter, Satchell, Stephen, & Schwob, Robert (2009). The link between macro-economic factors and style returns. *Journal of Asset Management*, *10*.