Hedging Properties of Algorithmic Investment Strategies using Long Short-Term Memory and Time Series models for Equity Indices

Jakub Michańków
University of Warsaw / Dep. of Quantitative Finance and Machine Learning / QFRG / Krakow University of Economics / Dep. of Computer Science
Warsaw, Poland *jmichankow@wne.uw.edu.pl*

Paweł Sakowski
University of Warsaw / Dep. of Quantitative Finance and Machine Learning / QFRG
Warsaw, Poland *p.sakowski@uw.edu.pl*

Robert Ślepaczuk
University of Warsaw / Dep. of Quantitative Finance and Machine Learning / QFRG
Warsaw, Poland *rslepaczuk@wne.uw.edu.pl*

Abstract

This paper proposes a novel approach to hedging portfolios of risky assets when financial markets are affected by financial turmoils. We introduce a novel approach to diversification on the level of ensemble algorithmic investment strategies (AIS) built on the prices of these assets. We employ four types of diverse models (LSTM, ARIMA-GARCH, momentum, contrarian) to generate price forecasts, which are used to produce investment signals in single and complex AIS. We verify the diversification potential of different types of investment strategies consisting of various assets classes in hedging ensemble AIS built for equity indices (S&P 500). We conclude that LSTM-based strategies outperform the other models and that the best diversifier for the AIS built for the S&P 500 index is the AIS built for Bitcoin. Finally, we test the LSTM model for 1-hour frequency of data. We conclude that it outperforms the results obtained using daily data.

Keywords: Deep Learning, Neural Networks, LSTM, Algorithmic Investment Strategies

1. Introduction

The main objective of this research is to improve the decision-making process by incorporating energy commodities and other asset classes into the hedging strategy of a diversified portfolio comprised of ensemble algorithmic investment strategies (AIS) constructed for the S&P 500 index. We present novel multidimensional verification of the possibilities of constructing and combining algorithmic investment strategies developed based on 1) the Long Short-Term Memory (LSTM) model, 2) the ARIMA-GARCH class models, as well as concepts of 3) contrarian and 4) momentum strategies for various assets: equity indices, precious metals, energy and soft commodities, and cryptocurrencies. The selection of theoretical models and assets is dictated by the aim to include a set of those which is diverse enough and at the same time highly tested in the literature. We are going to achieve it by: testing the efficiency of single strategy and ensemble strategies built with: 1) various types of assets, 2) various theoretical models; introducing a walk-forward approach enabling us to test theoretical models on various training, validation, and testing periods with different characteristics of return distributions; verifying the diversification potential of various strategies built using different theoretical concepts and different types of assets in hedging investment strategies built on the S&P500 index; performing sensitivity analysis to check the robustness of final results to various frequencies of data.

Our main contribution to existing literature can be found in a completely novel approach to testing diversification and hedging potential. We focus on the combination of single and ensemble algorithmic investment strategies built for various types of assets to maximize risk-adjusted return instead of focusing on just a single combination of new assets with adequate characteristics of returns enabling us to optimize the weights of our portfolio.

In our research, we use a walk-forward procedure on a daily time series with dates ranging from 2004-01-02 to 2022-03-29. In practice, the starting point of data depends on the asset and availability of data for the tested asset and varies between 2004-01-02 and 2010-07-17. In order to accomplish the main aim we decided to formulate the following research questions (RQ): RQ1: *Which of the tested groups of assets (energy commodities, cryptocurrencies, gold, or soft commodities) has the largest diversification potential in complex AIS (built with machine learning (ML) models and ARIMA-GARCH models) for equity indices?*; RQ2: *Are ML techniques more efficient than ARIMA-GARCH models and the concepts of momentum and contrarian in the case of single and complex (ensemble model combining all tested strategies for the given assets - type I) investment strategies.*; RQ3: *Are complex (ensemble) AIS based on the aggregation of all theoretical models for the single asset (type I) or all assets for a single theoretical model (type II) more efficient than individual strategies?*; RQ4: *Are results for LSTM models on higher frequencies of data (1h) better than those on daily data.*

The problem analyzed in this research is a fundamental issue not only from the micro, but also from the macro point of view, especially if we realize how much the stability of the financial systems of individual countries, and the state of savings of their citizens, are affected by the efficient and effective asset management in mutual and pension funds, investment funds, hedge funds or insurance companies. Wrong decisions in the allocation of these assets, especially in the context of long-term investment policies and specific investment strategies in the medium-term have very important consequences in the context of financial security and the quality of life of citizens of these countries. A similar approach to the one presented in this paper could also be extended to financial risk or macroeconomic forecasting.

2. Literature review

In this short literature review, we present a historical background covering the development of (recurrent neural networks) RNN and long short-term memory (LSTM) models and the summary of various empirical papers testing the efficiency of LSTM on various types of assets, frequencies, and studies trying to ensemble it in different ways.

Authors of [13] are responsible for the introduction of LSTM. By introducing Constant Error Carousel (CEC) units, LSTM deals with the exploding and vanishing gradient problems. The forget gate was introducted into LSTM architecture in [10], enabling the LSTM to reset its own state. Then, [5] put forward a simplified variant called Gated Recurrent Unit (GRU).

The study [3] implemented the LSTM model to predict the next-day returns for China stocks. [27] presented the AT-LSTM model which is the combination of LSTM and Attention-based model. Authors of [17] compared the performance of classical techniques with the LSTM model and showed that LSTM model results are highly dependent on initial hyperparameters assumptions. The study [25] investigates whether and how newly identified deep learning time series forecasting algorithms, such as LSTM, outperform more seasoned, econometric ones.

In paper [2], a portfolio of AIS built on S&P500 and Nasdaq Composite indices in the period of the last 40 years was tested. Authors argue that ensemble models can beat the benchmark in times of turbulent events as well as during very fast market growth. The study [6] analyzed the performance of three different RNN models using the price of Google. The data showed that on a five-day horizon, the LSTM outperformed other versions. Study [12] applied several ML algorithms to TA indicators showing that quantitative techniques beat passive strategies in terms of risk-adjusted returns.

Studies additionally make an effort to combine ensemble techniques with LSTM. Authors of [14] created a deep learning hybrid model using LSTM and GRU showing that the proposed network outperforms earlier neural network methodologies. The use of the LSTM model in AIS on BTC and S&P500 index on various frequencies was compared in [20]. They showed that the efficiency of LSTM in AIS strictly depends on hyperparameters, the construction of the model, and the estimation process. Additionally, they argue that proper loss functio is crucial in the model estimation process and that the results are dependent on asset classes tested and frequencies used. A good example of how an LSTM-RNN model may deliver exceptional predictions on non-stationary data is provided [24], where LSTM model yields great outcomes for daily and 7-day forecasts.

Authors of [26] compared the performance of ARIMA with the combination of ARIMA and GARCH models to construct AIS on S&P 500 index. Their main contribution was that the hybrid models outperformed ARIMA and the benchmark (Buy&Hold strategy on S&P 500 index) over the long term. In their study on high-frequency Bitcoin trading, [19] used three different machine learning (ML) models. Their findings show that artificial neural networks perform better than other types of systems in noisy signal environments. In paper [1], a walk-forward procedure is presented, which is in charge of training models and choosing the best one in order to predict future values of financial assets. Authors discovered that LSTM outperforms GRU in the vast majority of cases. In order to compare the performance of random forests and LSTM networks in predicting the directional movements of the stocks from S&P 500 index, [11] used both training methodologies.

In paper [9], LSTM signals were used to improve portfolios of pairs trading strategies. It was shown that LSTM signals contain information that goes above and beyond traditional indicators. Moreover, what is important in our study they revealed that LSTM signals allow for the disentangling of the reversal effect from the momentum effect. Another paper that applied long short-term memory networks to financial market predictions was written by [8]. LSTM was benchmarked against deep nets, random forests, and logistic regression. It occurred that Long short-term memory networks exhibit the highest predictional accuracy and returns.

Based on the above, we conclude that implementation of the forecasts from LSTM models in buy/sell signals can increase the efficiency of investment strategies. Moreover, we observe a growing number of publications on various types of ensemble models that combine frequencies or assets on the level of the given theoretical models or try to develop new investment techniques by joining many kinds of theoretical models in the process of price forecasting. Finally, we can notice that the type of input variables, the type of normalization, and specifically the architecture of the selected ML model can significantly affect the final results.

3. Methodology and Data

3.1. Terminology and Metrics

The investment strategies we use in this work are based on the forecasts obtained from 1) ARIMA-GARCH class models, 2) the Long Short-Term Memory network (LSTM), the concepts of 3) contrarian, and 4) momentum effects. In the case of ARIMA-GARCH models, we apply the concise rolling walk-forward procedure with three information criteria (AIC, SBC, and HQC). For the purpose of LSTM modeling, a custom loss function (MADL) was utilized as the network performance metric and used during the training process ([20]). Buy and sell signals that we use for single investment strategies are based on 1-period ahead forecasts of daily returns. Strategy performance metrics are calculated using the equity line constructed for each strategy separately.

3.2. ARIMA-GARCH model

Log-returns of assets are described by the ARIMA$(p, 0, q)$-GARCH$(1, 1)$ model:

$$r_t = \mu + \sum_{i=1}^{p} \phi_i r_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \quad (1)$$

$$\varepsilon_t = \sqrt{h_t} z_t, \qquad z_t \overset{\text{IID}}{\sim} N(0,1) \quad (2)$$

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1} \quad (3)$$

where μ, ω, α, β are parameters, z_t is the IID error term, and h_t is the conditional variance function.

In order to prepare forecasts based on ARIMA$(p, 0, q)$ -GARCH$(1, 1)$ model, we use the following estimation process: the parameters of the model are re-estimated every day; ARMA(p, q) orders are re-optimized every quarter with AIC, SBC, and HQC (p_{max} = 5, q_{max} = 5); AIC is used for the base case scenario; when the estimation of the model was not possible we use the last available model.

3.3. Contrarian and momentum strategies

Contrarian approach

It is one of the simplest investment strategies ([23], [7], [16]) assuming a strong mean-reverting process in the analyzed time series, which implies that our next day return forecast is exactly opposite to the previous day's return. Therefore, we define a $\text{Buy}_{\text{signal}}$ on P_t if $r_t < 0$, and a $\text{Sell}_{\text{signal}}$ on P_t if $r_t \geq 0$, where P_t is the price at time t.

Momentum approach

On the other hand, the momentum strategy ([15], [4], [9], [21], and [22]) assumes that financial returns tend to be persistent, which implies that our next-day return forecast is exactly the same with regard to the sign to the previous day's return. As a result, we define a $\text{Buy}_{\text{signal}}$ on P_t if $r_t \geq 0$ and a $\text{Sell}_{\text{signal}}$ on P_t if $r_t < 0$.

In the case of contrarian and momentum signals, their values are based on the return from the previous day.

3.4. LSTM model

Architecture of LSTM

LSTM networks are a type of recurrent neural networks (RNNs) that can keep track of long-term dependencies in data, allowing for partial solving of vanishing gradient problems typical for classic RNNs. It's widely used to model sequential data such as text, speech, and time series. The LSTM adds a way to carry information (c_t) across many timesteps and hence preventing older signals from gradually vanishing during processing. The information c_t is combined with the input connection and the recurrent connection.

Our LSTM model consists of three layers with 512/256/128 neurons and one single neuron dense layer on the output. Each of the LSTM layers is using *tanh* activation function (to retain negative values). L2 regularization (0.000001) and dropout (0.001) are also applied to each of these layers. The first two layers return sequences with the same shape as the input sequence (full sequence), and the last LSTM layer returns only the last output.

To train the model we use the Adam optimizer, with learning rate set to 0.5 (after tuning).

Data selection, hyperparameters tuning and LSTM training

We focus primarily on logarithmic returns, using daily data for S&P500, bitcoin (BTC), gold (GLD), natural gas (UNG), and wheat (ZWF) from 2004-01-02[1] and 2022-03-29. We also use

[1] In practice, the starting point of data depends on the asset and varies between 2004-01-02 and 2010-07-17.

hourly data for SPX and UNG, from the same period[2]. Hourly data availability is restricted for extensive time periods, so proprietary data was used in this case. However, daily data for all tested assets is readily accessible.

For the training set, we use an expanding window approach, with the size of the first window set to 252 days (number of trading days in a year). The size of the validation set is 33% of the training set. The test set size is always 252 days. The input sequence size for the LSTM network is set to 10. We use the ReLU activation function on the last neuron to obtain only zero or positive values (for Long Only strategies) or inverted ReLU to obtain zero or negative values (for Short Only strategy used for UNG). The output of the model is a single number predicting the next return value. Based on the sign of the predicted return value we assign -1, 0, and 1 signals, depending on the strategy.

During our research, we conduct manual hyperparameter tuning to ensure the best possible results from our model. The hyperparameters we test are summarized in Table 1.

Table 1. Values of hyperparameters selected after network tuning.

Hyperparameter	**Description/tested range**	**Selected Value**
No. hidden layers	from 1 to 5	3
No neurons	from 5 to 512	512/256/128
Activation function		*tanh*
Dropout rate	from 0 to 0.5 and kernel regularization from 0 to 0.01	0
l2 regularizer		1e-6
Optimizer	SGD, RMSProp or Adam	Adam
Learning rate	from 0.0001 to 0.9 and momentum values from 0.1 to 0.9	0.5
Train/test size		252-exp. window/252
Batch size		exp. window
Sequence length		10
Number of Epochs	from 10 to 300 and callbacks (early stopping and model checkpoint)	300

Note: Hyperparameters used in this study for the LSTM model. Tuning process was performed on the first year of available data.

In addition, we change the following hyperparameters of the network to optimize it for high-frequency data: train and test sizes are increased to cover one calendar year of data, an additional layer with 252 neurons is added and the number of epochs is changed to 120.

For training and prediction, we use a walk-forward validation/expanding window approach. In the first iteration, the model is trained on one year of data (equal to the train set length) and then used for predictions over the next year (equal to the test set length). After that, the window is expanded by another year of data and the model was retrained. A single return value is predicted each time, using the last 10 (sequence length) values. A single iteration is trained for 300 epochs. The model checkpoint callback function is used to store the best weights (parameters) of the model based on the lowest loss function value in a specific epoch. The weights are then used for prediction.

Loss function for LSTM model

As the loss function, we use MADL, which was proposed by [20]. They appropriately evaluate the usefulness of the forecasting ability of the LSTM model in algorithmic investment strategies (AIS). The MADL is given by:

$$\mathrm{MADL} = \frac{1}{N}\sum_{i=1}^{N}(-1) \times \mathrm{sign}(R_i \times \hat{R}_i) \times \mathrm{abs}(R_i) \tag{4}$$

[2] The use of the mix of hourly and daily data was dictated by the fact that additional research on HF (hourly) dat can reveal some patters which are not visible on the daily level and could enables our model to adjust to these intraday patterns

where R_i is the observed return on interval i, $\hat{R}_i$ is the predicted return on interval i, $\text{sign}(X)$ is the function which gives the sign of X, $\text{abs}(X)$ is the function which gives the absolute value of X, and N is the number of forecasts.

In this way, the value of MADL is equal to the observed return on the investment with the predicted sign. This allows the model to inform us if its prediction will yield profit or loss and how much this profit or loss will be. MADL was designed specifically for working with AIS's instead of just verification of forecasts in point. The MADL is minimized, so that if it gives negative values, the strategy will make a profit, and for positive values, the strategy will generate a loss.

3.5. Ensemble models

We create two types of ensemble models: type I - built with various theoretical models for the selected type of asset; type II - built with various types of assets for the selected theoretical model.

Therefore, type I ensemble models for a given asset j, were created as:

$$\text{EQline}_j^{(\text{I})} = \frac{1}{n}\sum_{i=1}^{n}\text{EQline}_{i,j} \tag{5}$$

where n - the number of theoretical models, $i = \{1,\ldots,n\}$, $\text{EQline}_j^{(\text{II})}$ - the value of the ensemble equity line on day t for algorithmic investment strategy on the j-th asset (S&P 500 index, Bitcoin, Gold, Natural Gas, and Wheat) for all theoretical models (LSTM, ARIMA-GARCH, Momentum, and Contrarian models), $\text{EQline}_{i,j}$ - the value of the single equity line on the day t for algorithmic investment strategy on the j-th asset (S&P 500 index, Bitcoin, Gold, Natural Gas, and Wheat) for the i-th theoretical model (LSTM, ARIMA-GARCH, Momentum, and Contrarian),

On the other hand, type II ensemble models for a given theoretical model i, were defined as:

$$\text{EQline}_i^{(\text{II})} = \frac{1}{m}\sum_{j=1}^{m}\text{EQline}_{i,j} \tag{6}$$

where m - the number of assets, $j = \{1,\ldots,m\}$, $\text{EQline}_i^{(\text{I})}$ - the value of the ensemble equity line on day t for algorithmic investment strategy on all assets for one of the i-th theoretical model.

3.6. Performance metrics

Based on [17], we calculated the following performance metrics: annualized return compounded (aRC), annualized standard deviation (aSD), Maximum Drawdown (MD), Maximum Loss Duration (MLD), and three variants of Information Ratio (IR*, IR**, IR***). We treat the IR** as the most important in the evaluation of our final results, as this indicator combines information from two crucial risk metrics: aSD and MD.

3.7. Research description

The detailed research conducted in this study has several key components. Firstly, it involves testing two versions of investment strategies: Long Only and Short Only[3]. Additionally, a new loss function (MADL), as introduced by [20], is inc orporated. Hyperparameter tuning is conducted, following the specifics outlined in Section 3.4.

[3]Long Only means the strategy which opens only Long positions when Long signal is generated and stays out of the market when Short or Hold signal is generated. the details can be found in [18]

Walk-forward optimization is employed, which consists of both *in-sample* and *out-of-sample* procedures. During the *in-sample phase*, model parameters estimation (LSTM) or optimization of p and q orders (ARIMA-GARCH) is performed. The in-sample period involves utilizing the last $n \times 365$ actual days, where n = 1, 2, 3, 4, 5, with the base case being 3 years. Data for the last 1 to 5 years is included in this phase. The out-of-sample phase involves re-estimation and re-optimization of models, with forecast generation. The first out-of-sample period commences one year after the data start, across all five cases (1 to 5 years). Out-of-sample forecasts are generated one day ahead.

Furthermore, buy/sell signals are defined based on the next day forecasts. Equity lines and performance metrics are assessed following the methodology outlined by [28], with DFL equal to 1. The study also examines verifying the diversification potential of various asset classes and theoretical models for AIS constructed for the S&P 500 index. Moreover, ensemble investment strategies are constructed by combining signals across different asset classes and theoretical models. Finally, sensitivity analysis is conducted to assess the impact of various data frequencies used in the LSTM model.

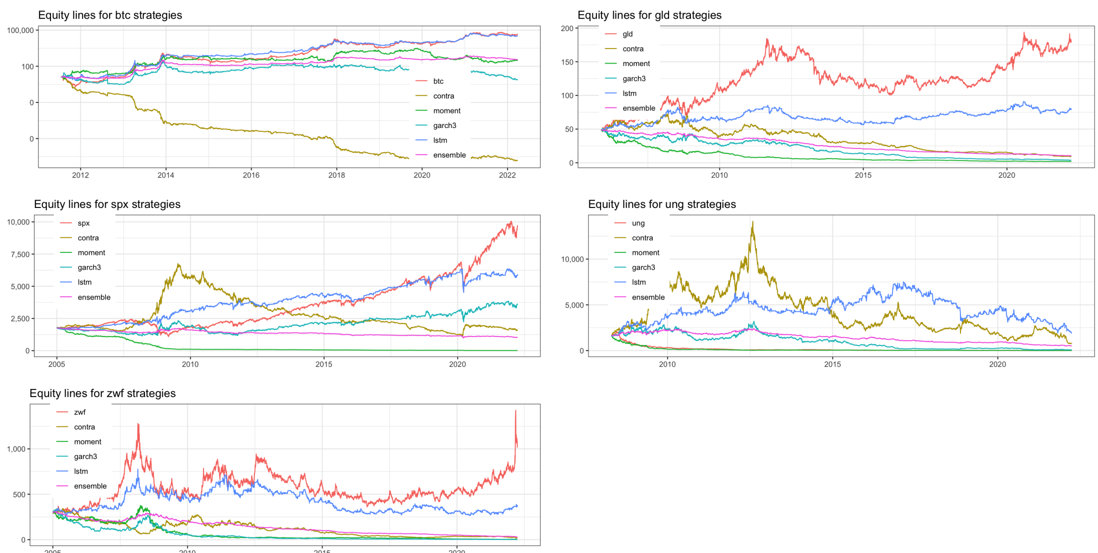
4. Results

We present results from less complex (single investment models) to more complex (ensemble investment models) while emphasizing their diversification potential. This sequence is not necessarily connected with the order of our research questions in the Introduction.

4.1. Base case scenario

In the first part of the results, we describe individual strategies and type I of the ensemble model where ensembling is defined as an equally weighted portfolio of different models/strategies for a single asset. Rebalancing is performed on the first available day of Jan, Apr, Jul, and Oct.

Figure 1 presents equity lines for every investment strategy and confirms the results described in Table 2.



Note: Each panel presents five equity lines for each tested asset (BTC, GLD, SPX, UNG, and ZWF). These equity lines represent the results for 4 individual strategies based on the model/concept of LSTM, ARIMA-GARCH, momentum, and contrarian, and one additional equity line for the ensemble model built using these four above-mentioned.

Fig. 1. Equity lines for individual and ensemble strategies for single assets

Table 2 shows the performance metrics for tested strategies (individual and ensemble - type I) and the benchmark Buy&Hold strategy. Based on these results we can notice that the LSTM model-based strategy is characterized by the highest IR (IR*, IR**, and IR***) in most cases.

Table 2. Base case scenario results for individual and ensemble strategies, for a single asset

		aRC	aSD	MD	MLD	IR*	IR**	IR***	nObs	nTrades
BTC	B&H	**114.78**	88.29	86.67	3.24	1.30	1.722	0.610	3909	2
	contra	-77.59	87.78	100.00	10.67	-0.88	-0.686	-0.050	3909	1980
	moment	34.66	88.97	94.09	4.01	0.39	0.144	0.012	3909	1984
	garch3	-5.10	88.58	97.41	8.30	-0.06	-0.003	0.000	3909	1351
	lstm	109.32	64.76	67.19	2.89	**1.69**	**2.747**	**1.040**	3909	525
	ensemble	36.57	44.83	68.13	3.99	0.82	0.438	0.040	3909	6012
GLD	B&H	**8.33**	18.28	45.56	8.92	**0.46**	**0.083**	0.001	4117	2
	contra	-9.68	17.74	87.46	14.05	-0.55	-0.060	0.000	4117	2133
	moment	-17.83	18.85	96.20	16.21	-0.95	-0.175	-0.002	4117	2143
	garch3	-14.29	18.28	92.88	16.27	-0.78	-0.120	-0.001	4117	1593
	lstm	3.03	12.63	34.14	8.53	0.24	0.021	0.000	4117	720
	ensemble	-8.95	6.10	78.95	16.27	-1.47	-0.166	-0.001	4117	6853
SPX	B&H	**10.35**	19.48	55.25	4.48	**0.53**	0.100	0.002	4340	2
	contra	-0.75	18.99	83.62	12.70	-0.04	0.000	0.000	4340	2269
	moment	-25.37	19.99	99.40	17.21	-1.27	-0.324	-0.005	4340	2269
	garch3	4.24	19.33	49.95	6.96	0.22	0.019	0.000	4340	1274
	lstm	7.23	14.92	28.43	1.99	0.48	**0.123**	**0.004**	4340	722
	ensemble	-3.03	7.45	44.52	17.22	-0.41	-0.028	0.000	4340	6810
UNG	B&H	-27.52	44.44	99.58	13.73	-0.62	-0.171	-0.003	3512	2
	contra	-5.47	43.86	94.69	9.64	-0.12	-0.007	0.000	3512	1803
	moment	-30.88	45.05	99.70	13.93	-0.69	-0.212	-0.005	3512	1797
	garch3	-19.90	44.40	97.75	9.61	-0.45	-0.091	-0.002	3512	1416
	lstm	**1.09**	31.15	74.79	5.10	**0.04**	**0.001**	0.000	3512	612
	ensemble	-8.05	14.98	79.28	9.61	-0.54	-0.055	0.000	3512	5852
ZWF	B&H	**7.24**	33.22	71.80	14.01	**0.22**	**0.022**	0.000	4362	2
	contra	-14.61	32.65	95.09	16.94	-0.45	-0.069	-0.001	4362	2240
	moment	-21.97	33.82	99.17	13.99	-0.65	-0.144	-0.002	4362	2264
	garch3	-29.25	33.21	99.78	17.18	-0.88	-0.258	-0.004	4362	1818
	lstm	1.04	22.98	65.99	14.08	0.05	0.001	0.000	4362	778
	ensemble	-12.68	10.61	90.69	17.18	-1.20	-0.167	-0.001	4362	7376

Note: Results cover the performance metrics for 4 individual strategies and 1 ensemble model for 5 various assets (BTC, GLD, SPX, UNG, and ZWF). The ensemble model stands for the combination of all theoretical models for the given asset.

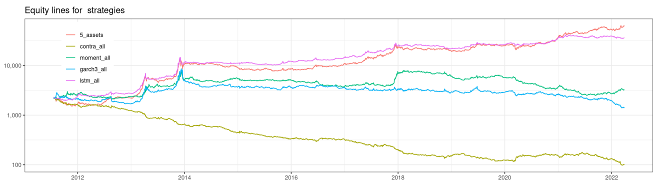
Table 3 contains the performance metrics for all types of ensemble models (type I - ensemble model combining all tested strategies for the given assets: SPX_all, BTC_all, GLD_all, UNG_all, ZWF_all, and type II - ensemble model combining all assets for the given tested strategy: contr_all, momentall, garch3_all, lstm_all) and compare it with Buy&Hold strategy for all 5 assets. The important conclusion from this table is that lstm_all outperforms other strategies and Buy&Hold and that BTC_all outperforms other assets. The former could be attributed to the distinctive architecture of LSTM networks, which provides them with the capability to more effectively capture intricate temporal patterns within the data, while the latter to the availability and continuity of BTC data, which is quoted 24/7.

Table 3. Ensemble strategies for single assets and theoretical models

	aRC	aSD	MD	MLD	IR*	IR**	IR***	nObs	nTrades
B&H all assets									
B&H_all	**23.947**	23.564	44.433	5.024	1.016	0.548	0.026	3908	10
ensembles for single assets									
SPX_all	-3.035	7.454	44.519	17.218	-0.407	-0.028	0.000	4340	6810
BTC_all	36.568	44.834	68.130	3.989	0.816	0.438	0.040	3909	6012
GLD_all	-8.949	6.099	78.950	16.274	-1.467	-0.166	-0.001	4117	6853
UNG_all	-8.053	14.976	79.276	9.611	-0.538	-0.055	0.000	3512	5852
ZWF_all	-12.684	10.608	90.690	17.179	-1.196	-0.167	-0.001	4362	7376
models for all assets									
contra_all	-18.166	15.448	95.657	15.472	-1.176	-0.223	-0.003	3908	10728
moment_all	2.567	22.708	68.705	5.980	0.113	0.004	0.000	3908	10760
garch3_all	-2.862	23.202	83.842	12.028	-0.123	-0.004	0.000	3908	7755
lstm_all	19.674	18.072	30.274	4.460	**1.089**	**0.707**	**0.031**	3908	3660

Note: Each panel presents performance metrics for the Buy&Hold strategy for all assets (5_assets), for ensemble models for single assets (SPX_all, BTC_all, GLD_all, UNG_all, ZWF_all), ensemble models for theoretical concepts (contra_all, moment_all, garch3_all, lstm_all).

Figure 2 visualizes fluctuations of equity lines for ensemble models and Buy&Hold and confirms the high performance of the LSTM model-based strategy.



Note: 5_assets stands for Buy&Hold strategy for all assets. Contra_all, moment_all, garch3_all, lstm_all stand for ensemble models for all assets within one theoretical concept.

Fig. 2. Ensemble strategies for all theoretical models

Table 4 contains a summary of the research which enables us to refer to research questions.

Table 4. Ensemble strategies for all assets within one theoretical model.

IR**	BTC	GLD	SPX	UNG	ZWF	positive IR**	beat B&H	winner
B&H	1.722	**0.083**	0.100	-0.171	0.022	80%	0%	40%
contrarian	-0.686	-0.060	0.000	-0.007	-0.069	0%	20%	0%
momentum	0.144	-0.175	-0.324	-0.212	-0.144	20%	0%	0%
garch3	-0.003	-0.120	0.019	-0.091	-0.258	20%	20%	0%
lstm	**2.747**	0.021	**0.123**	**0.001**	0.001	100%	60%	60%
ensemble	0.438	-0.166	-0.028	-0.055	-0.167	20%	20%	0%

Note: B&H stands for Buy&Hold strategy for all assets. Contrarian, momentum, garch3, lstm stand for ensemble models for all assets within one theoretical concept. The ensemble stands for ensemble model for all assets and all theoretical models.

Based on the results for the base case scenario, presented in Tables 2, 3, and 4 and Figure 1, and 2, we can refer to RQ2 and RQ3. Referring to RQ2, we can confirm that ML models are more efficient than classical models. In the case of single investment strategies because LSTM was the best strategy in 60% of the cases (3 out of 5 asset classes tested). Moreover, in the case of complex investment strategies (type II) based on the aggregation of all assets for a single theoretical model lstm_all was the best strategy in comparison to contrarian_all, momentum_all, and garch_all.

Regarding RQ3, the ensemble AIS based on the aggregation of all theoretical models for the single asset (type I) were never better than the LSTM model or the B&H strategy. Moreover, none of the ensemble strategies based on the aggregation of all assets for the single theoretical model (type II): lstm_all, contrarian_all, momentum_all, garch_all, and B&H_all were better than the single strategies for the given class of asset.

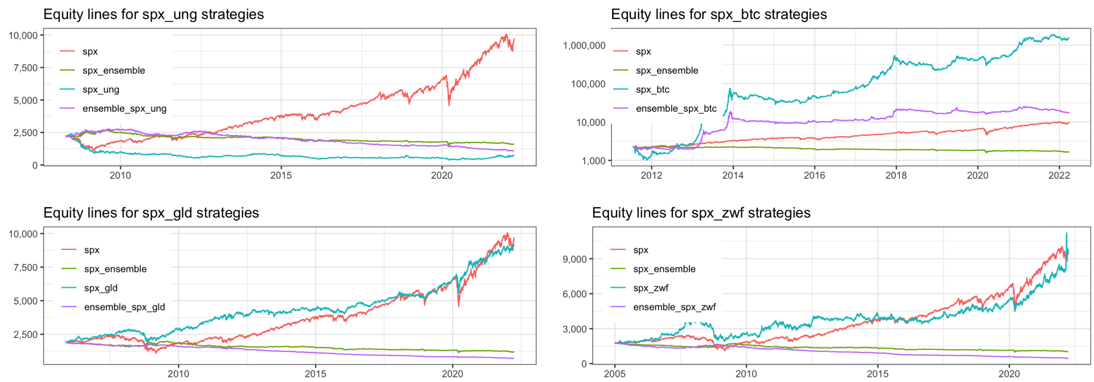
4.2. Base Case Scenario. Ensemble models based on two assets - diversification potential.

Based on the results presented in Table 5 and Figure 3 for the ensemble models of two assets and their diversification potential with regard to strategies based on SPX, we can refer to the RQ1. Looking at the IR**, we can state that the only diversification potential can be noticed after adding the ensemble model based on BTC to the ensemble model based on SPX, where the IR** for the ensemble_spx_btc increases.

Table 5. Diversification potential of investment models for hedging equity index investment model

		aRC	aSD	MD	MLD	IR*	IR**	IR***	nObs	nTrades
UNG	spx	**11.27**	20.59	51.52	2.75	**0.55**	**0.120**	0.005	3512	2
	spx_ensemble	-2.14	7.90	42.82	12.69	-0.27	-0.014	0.000	3512	5357
	spx_ung	-7.39	25.30	84.11	13.77	-0.29	-0.026	0.000	3512	224
	ensemble_spx_ung	-4.91	8.25	61.09	12.59	-0.59	-0.048	0.000	3512	11209
BTC	spx	10.01	14.44	33.79	1.08	0.69	0.205	0.019	3908	2
	spx_ensemble	-1.76	5.64	30.19	14.37	-0.31	-0.018	0.000	3908	4003
	spx_btc	**52.72**	42.33	61.87	4.42	**1.25**	**1.061**	**0.127**	3908	172
	ensemble_spx_btc	14.36	23.01	52.71	5.82	0.62	0.170	0.004	3908	10012
GLD	spx	**10.58**	19.85	55.25	4.48	0.53	0.102	0.002	4117	2
	spx_ensemble	-2.72	7.61	42.82	12.69	-0.36	-0.023	0.000	4117	6417
	spx_gld	10.17	13.58	34.05	1.72	**0.75**	**0.224**	**0.013**	4117	264
	ensemble_spx_gld	-5.80	5.00	63.21	16.31	-1.16	-0.107	0.000	4117	13270
ZWF	spx	**10.30**	19.43	55.25	4.48	**0.53**	**0.099**	**0.002**	4362	2
	spx_ensemble	-3.02	7.44	44.52	17.31	-0.41	-0.028	0.000	4362	6802
	spx_zwf	10.06	20.45	56.04	6.81	0.49	0.088	0.001	4362	276
	ensemble_spx_zwf	-7.80	6.54	75.55	17.18	-1.19	-0.123	-0.001	4362	14178

Note: Each of the 4 panels contains the results for 4 strategies: SPX - B&H for S&P 500 index, spx_ensemble - the ensemble models combining all theoretical models for S&P 500 index, spx_asset - combined B&H for S&P 500 index and the given asset, ensemble_spx_asset - the combination of two ensemble models built for all theoretical models for SPX and the given asset.



Note: Each panel presents the equity lines for 4 different strategies: SPX - B&H for the S&P 500 index, spx_ensemble - the ensemble models combining all theoretical models for the S&P 500 index, spx_asset - combined B&H for S&P 500 index and the given asset, ensemble_spx_asset - the combination of two ensemble models built for all theoretical models for SPX and the given asset.

Fig. 3. Equity lines for hedging strategies for equity index

4.3. Daily versus hourly results for selected assets

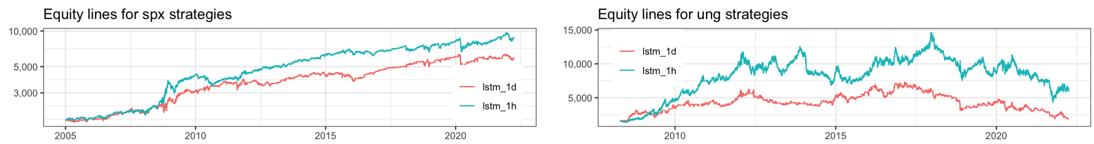
To answer RQ4, we repeat training and estimation of the LSTM model for SPX and UNG assets on hourly data in the same period as for the daily data, i.e. from 2008-04-17 to 2022-03-29. The selection of these two assets was dictated by the following reasons. The S&P 500 index was chosen for its wide usage in financial literature, ensuring comparability with other research. UNG represents a distinct dynamic with decreasing asset prices over time and potential diversification benefits during geopolitical stress, such as the Russian-Ukrainian conflict.

Table 6. LSTM model results for S&P 500 index and UNG on daily and hourly data

		aRC	aSD	MD	MLD	IR*	IR**	IR***	nObs	nTrades
S&P 500	lstm_1d	7.23	14.92	28.43	1.99	0.48	0.123	0.004	4340	722
	lstm_1h	**9.72**	12.34	24.25	1.72	**0.79**	**0.315**	**0.018**	34702	5364
UNG	lstm_1d	1.09	31.15	74.79	5.1	0.04	0.001	0.000	3512	612
	lstm_1h	**6.80**	24.38	70.54	6.17	**0.28**	**0.027**	0.000	122318	16310

Note: lstm_1d stands for LSTM model-based investment strategy trained and estimated for the S&P 500 index (first panel) and UNG on daily data. lstm_1h denotes the same strategy test on 1h data.

Table 6 and Figure 4 shows that in each tested case LSTM models on hourly data outperform the ones on daily data in each case of risk-adjusted measures (IR*, IR**, and IR***).



Note: Each panel presents equity lines for SPX and UNG for two different frequencies daily and hourly.

Fig. 4. Equity lines for LSTM model for S&P 500 index and UNG on daily versus hourly data

5. Conclusions

The novelty and the main contribution of this paper is an attempt to focus on the problem of diversification from a different perspective than what is usually presented in state-of-the-art research. We verify the diversification potential of investment strategies for the S&P 500 index, based on various theoretical concepts against other investment strategies. Therefore, referring to RQ1, based on the results in Table 5 and Figure 3, we can state that only ensemble_BTC has the diversification potential that increases the efficiency of ensemble models for the equity index. Moreover, since the distribution of returns for other equity indices is similar to that of the S&P 500, our conclusions can be extended to them as well.

Based on Table 4 and Figure 2, we can affirmatively address RQ2. After analyzing the results in Table 2 and Table 4, we can assert an unfavorable response to RQ3. Finally, based on the results presented in Table 5 and Figure3, we can provide a positive response to RQ4.

At the end, we have to stress that due to various kinds of returns distributions among different asset classes and the impact of current market conditions, we see some potential limitations on the performance of the proposed strategies.

Acknowledgements

This document is the results of the research project funded by IDUB programs: BOB-IDUB-622-233/2022 and BOB-661-1057/2024 at the University of Warsaw

References

1. Baranochnikov, I., Ślepaczuk, R.: A comparison of lstm and gru architectures with novel walk-forward approach to algorithmic investment strategy WP WNE UW 21/2022(397) (2022)
2. Castellano Gomez, S., Ślepaczuk, R.: Robust optimisation in algorithmic investment strategies WP WNE UW 27/2021(375) (2021)
3. Chen, K., Zhou, Y., Dai, F.: A lstm-based method for stock returns prediction: A case study of china stock market. In: 2015 IEEE Int. Conf. on Big Data. IEEE (2015)
4. Chu, J., Chan, S., Zhang, Y.: High frequency momentum trading with cryptocurrencies. Research in international business and finance 52, pp. 101176 (2020)
5. Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. arXiv:1412.3555 [cs] (Dec 2014)
6. Di Persio, L., Honchar, O.: Rnns approach to the financial forecast of google assets. Int. Journal of Mathematics and Computers in simulation 11, pp. 7–13 (2017)
7. Dobrynskaya, V.: Avoiding momentum crashes: Dynamic momentum and contrarian trading. J. of Int. Fin. Markets, Institutions and Money 63, pp. 101141 (2019)
8. Fischer, T., Krauss, C.: Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research 270(2), pp. 654–669 (2018)
9. Flori, A., Regoli, D.: Revealing pairs-trading opportunities with long short-term memory networks. European Journal of Operational Research 295(2), pp. 772–791 (2021)
10. Gers, F.A., Schmidhuber, J., Cummins, F.: Learning to Forget: Continual Prediction with LSTM. Neural Computation 12(10), pp. 2451–2471 (Oct 2000)

11. Ghosh, P., Neufeld, A., Sahoo, J.K.: Forecasting directional movements of stock prices for intraday trading using lstm and random forests. Finance Research Letters 46, pp. 102280 (2022)
12. Grudniewicz, J., Ślepaczuk, R.: Application of machine learning in algorithmic investment strategies on global stock markets. Research in International Business and Finance 66, pp. 102052 (2023)
13. Hochreiter, S., Schmidhuber, J.: Long Short-Term Memory. Neural Computation 9(8), pp. 1735–1780 (Nov 1997)
14. Hossain, M.A., Karim, R., Thulasiram, R.K., Bruce, N.D.B., Wang, Y.: Hybrid deep learning model for stock price prediction. 2018 IEEE Symposium Series on Computational Intelligence (SSCI) pp. 1837–1844 (2018)
15. Jegadeesh, N., Titman, S.: Momentum. Annu. Rev. Financ. Econ. 3(1), pp. 493–509 (2011)
16. Kadoya, S., Kuroko, T., Namatame, T.: Contrarian investment strategy with data envelopment analysis concept. European Journal of Operational Research 189(1), pp. 120–131 (2008)
17. Kijewski, M., Ślepaczuk, R.: Predicting prices of S&P 500 index using classical methods and recurrent neural networks WP WNE UW 27/2020(333) (2020)
18. Kosc, K., Sakowski, P., Ślepaczuk, R.: Momentum and contrarian effects on the cryptocurrency market. Physica A: Statistical Mechanics and its Applications 523, pp. 691–701 (2019)
19. Lahmiri, S., Bekiros, S.: Intelligent forecasting with machine learning trading systems in chaotic intraday bitcoin market. Chaos, Solitons & Fractals 133, pp. 109641 (2020)
20. Michańków, J., Sakowski, P., Ślepaczuk, R.: LSTM in algorithmic investment strategies on BTC and S&P 500 index. Sensors 22(3) (2022)
21. Ong, J., Herremans, D.: Constructing time-series momentum portfolios with deep multi-task learning. Expert Systems with Applications 230, pp. 120587 (2023)
22. Pal, A., Singh, K.P.: Adamr-grus: Adaptive momentum-based regularized gru for hmer problems. Applied Soft Computing 143, pp. 110457 (2023)
23. Park, A., Sabourian, H.: Herding and contrarian behavior in financial markets. Econometrica 79(4), pp. 973–1026 (2011)
24. Shah, D., Campbell, W., Zulkernine, F.H.: A comparative study of lstm and dnn for stock market forecasting. In: 2018 IEEE Int. Conf. on Big Data. IEEE Computer Society (2018)
25. Siami-Namini, S., Tavakoli, N., Namin, A.S.: A comparison of arima and lstm in forecasting time series. In: 2018 17th IEEE international conference on machine learning and applications (ICMLA). pp. 1394–1401. IEEE (2018)
26. Vo, N., Ślepaczuk, R.: Applying hybrid arima-sgarch in algorithmic investment strategies on S&P 500 index. Entropy 24(2) (2022)
27. Zhang, X., Liang, X., Li, A., Zhang, S., Xu, R., Wu, B.: AT-LSTM: An Attention-based LSTM Model for Financial Time Series Prediction. IOP Conference Series: Materials Science and Engineering (Aug 2019)
28. Ślepaczuk, R., Sakowski, P., Zakrzewski, G.: Investment strategies that beat the market. what can we squeeze from the market? Financial Internet Quarterly 14(4), pp. 36–55 (2018)