Comparing Time-Series Prediction Strategies for Automated Trading on Commodity Markets

Anonymous Author(s)

ABSTRACT

Price prediction has been the subject of increasing interest within the machine learning research community, where the majority of work has focused on stocks and shares. However, commodity markets (those concerned with base materials/ingredients like wheat, copper and oil) are equally important, but critically under-researched. Recent inflationary concerns, fuelled by rapid rises in energy prices, has emphasised the importance of reliable commodity price forecasting. This paper acts as an entry-point for researchers and practitioners interested in commodity price forecasting, by providing a structured analysis of forecasting techniques and their effectiveness when applied as part of trading strategies. In particular, we investigate three forecasting techniques of varying difficulty, namely: direction prediction; profitability prediction; and price prediction. For each, we perform a structured set of experiments with a range of supervised machine learning models over a commodity market dataset, evaluating their forecasting accuracy for varying time-horizons and how this translates into trading profitability, demonstrating that direct price prediction leads to the greatest downstream profits, despite high degrees of prediction error.

KEYWORDS

Commodity Trading, Time-Series Prediction, Machine Learning

ACM Reference Format:

Anonymous Author(s). 2022. Comparing Time-Series Prediction Strategies for Automated Trading on Commodity Markets. In *Proceedings of Applied Machine Learning Methods for Time Series Forecasting (AMLTS22).* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

In recent years, advancements in technology have markedly lowered the barriers to investment in financial markets, leading to record numbers of individuals choosing to invest (e.g. 13.5% of LSE stocks were held by individuals in 2018, up from only 10.6% in 2012.[1] However, markets and their interdependencies are complex, meanwhile for these individuals investing is not their full-time job. Hence, this growing cohort is increasingly reliant on automated solutions (referred to as robo-advisors) to analyse the markets and identify good investment opportunities.

[1]https://www.finder.com/uk/investment-statistics

AMLTS22, Oct. 24, 2022, Atlanta , GA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

The core of many robo-advisors is asset performance prediction, i.e. they use the past performance of financial assets to predict how they are likely to perform in the future – forming the basis for recommending investment options to the customer. However, asset performance prediction can be performed at different granularities, which in-order of increasing complexity we refer to as *direction prediction, profitability prediction and price prediction.* Under direction prediction, the advisor simply attempts to predict whether the asset will lose or gain value (but not by how much). Profitability prediction aims to predict the future return on investment for an asset for a fixed period. Finally, price prediction directly attempts to predict the price of an asset at a future point in time (factoring in differing price ranges and variances across asset classes). However, which strategy is the most effective from the perspective of investment? While there have been numerous past works that experiment with instances of these three strategies, to the best of our knowledge these three approaches have never been directly compared.

Additionally, the vast majority of research in this area has targeted stocks and shares markets, however there is another highly influential set of markets that are chronically under-studied, namely: *commodity markets.* Commodity markets are concerned with the trading of raw materials, such as wheat, copper or oil, and are known to strongly impact the stocks and shares markets, as the listed company's profit margins are impacted by the price of the raw materials they consume [9]. This is of particular concern at present: huge increases in food and energy commodity prices have resulted in record inflation levels [17, 18].

Hence, as a step toward more effective asset performance prediction solutions, in this paper we compare different prediction strategies over a range of commodity marketplaces. Through experimentation over a large dataset spanning 20 years of pricing data for 20 commodities, we demonstrate that directly predicting the profitability of assets is the best overall strategy in terms of profit made. Indeed, profitability prediction was 6% more effective on average than direction prediction and more than 20% better than price prediction. Moreover, we show that prediction performance metrics such as classification F_1, MAE and MAPE are not necessarily strongly correlated with resultant profit, as the downstream trading strategy is not affected by all types of error evenly.

2 RELATED WORK

The use of machine learning to attempt to improve price prediction is well established. Table 1 summarises significant prior works and groups by task and asset class.

Initial works focused on stock price prediction [25, 12]. Utilising a small dataset of historical pricing and volume data as input and framing the task as a regression on profitability, White [25] met with limited success, in part due to the hardware constraints of the time. These resource constraints may, due to the bias variance trade-off, have encouraged interest in the use of classification models. As

Table 1: Sample of papers from both the stock and commodity price prediction literature grouped by task type

Task type	Stocks	Commodities
Direction	[10, 11, 12, 22]	[6]
Profitability	[25]	
Price	[8, 16, 19, 20, 23, 24, 26]	[3, 4, 7, 21]

the model and task definition tends to be simpler for classification than regression, they are less susceptible to data noise issues when using smaller sample sizes [15]. For example Kim [11] used a support vector classifier when predicting the price movement direction of a basket of stocks on the South Korean Stock Exchange. More recently, with increasing sample sizes, regression models have become more popular. Nunno [19] compared a number of regression models including linear regression and support vector regression when predicting a selection of stocks listed on the New York Stock Exchange and Vijh et al. [23] utilised 10 years worth of daily trading data to train a neural network aiming to predict 6 different stocks on the same exchange. It is worth noting however, that the recent literature still contains examples of classification. Khaidem et al. [10] make use of a random forest classifier when predicting price direction of a mixture of stocks from American and Korean stock exchanges.

Though the problems of stock and commodity price prediction have a number of aspects in common, there are significant differences between the assets classes as they are commonly traded. Stocks (being shares of ownership of an entity) may be held indefinitely. Commodities on the other hand, tend to be traded as futures contracts with an expiration date [5]. These differences necessitated separate research into the problems; we cannot simply assume what works for one will work for the other. As can be seen from table 1, research into commodity prediction is relatively sparser and developed later. Price regression is most common approach of the three under examination in this paper. Herrera et al. [7] compared a number of regression techniques when predicting the prices of energy commodities finding a random forest regressor to be most effective. That classification and profitability are less well represented may be a result of this later development of commodity research, i.e. when hardware had the capacity to support datasets with large sample sizes. A further explanation may be found in the intent of commodity research: much seems aimed at assisting producers to obtain a fair price for their goods [3]. Stock price literature, on the other hand, is more likely to approach from the perspective of a speculator [12].

In addition to growing sample sizes, feature engineering has also been used by authors in stock price prediction before more recent adoption to commodities. There are two main feature engineering approaches as pertains to price prediction. Firstly, adding new data sources which contain information relating to the broader economic context. Kimoto et al. [12], used a data set of six features including interest and dollar yen exchange rates when attempting to predict Tokyo Stock Exchange prices. The second is the use of technical indicators. These are metrics derived from the historical data are used by market participants to make investment decisions. An early example of this approach is work by Kim [11]. They utilised a broad range of technical indicators, including Commodity Channel Index and Relative Strength Index when predicting price movements of the assets on the Korean Stock. Economists have questioned the theoretical underpinnings of technical indicators [14]. Regardless, it seems probable that they possess some predictive power, if only because of their widespread use by those collectively able to influence markets. The works of Naik and Mohan [16] and Yin and Yang [26] provide evidence for this in stocks and commodities respectively. Both compared the predictive strength for a large number of technical indicators, in the main finding that their inclusion improved prediction performance.

Common to both stock and commodity literature is use, by different authors, of a wide range of forecast horizons. Also referred to as prediction horizon, this is the number of trading days in advance being predicted for. For example Safari and Davallou [21] predict the oil price one month in advance while Naik and Mohan [16] predict next day stock prices.

A difference in the literature for stock and commodity pricing literature is while it is fairly common in stock price prediction to consider a wide variety of stocks; commodity researchers have tended to focus on a single commodity or sector. Vijh et al. [23] examined six stocks from sectors such finance, healthcare and clothing, whereas Herrera et al. [7] focused their research on energy commodities such as natural gas, coal and oil and Chen et al. [3] examined agricultural commodities such as chilli and tomatoes.

The main method of performance evaluation, across all of the literature, is the use of statistical measures. For classification many authors have used: recall, precision and f1 scores [22]. For regression R^2, mean absolute error and mean absolute percentage error have seen extensive use [7, 8, 12]. An additional method of evaluation used in conjugation with statistical measures by a few authors including Kimoto et al. [12] is trading simulation. This involves using the prediction output to inform a trading strategy which is then back tested on historical data, and performance estimated in terms of profit loss. Examples of simulation are significantly less common for commodity price prediction; perhaps due to complications arising from contract expiration. Such an approach however, would provide a common metric by which classification and regression could be compared.

From this review it is evident that profitability prediction is under explored. Additionally, although classification and regression have seen extensive work, this exploration appears less complete for commodities. Furthermore, although a number of authors have compared different models within these three approaches, we could find no attempts to to directly compare between approaches [7, 19, 20]. Therefore, in this work, we compare directly the effectiveness of all three techniques for commodity prediction.

3 TASK FORMULATION

As seen in the previous section, in the financial domain, three prediction strategies have been considered. Namely, *direction prediction*, *profitability prediction* and *price prediction*. However, their particular differences have made researchers consider them as separate tasks – and thus not compare them under a common setting. Due to their potential as tools for assisting investors on portfolio management decision making, this comparison represents an interesting

Table 2: Description of the three prediction strategies.

Target	Task type	Variation magnitude	Price scale
Direction	Classification	✗	✗
Profitability	Regression	✓	✗
Price	Regression	✓	✓

research perspective – which we address in this paper. As a first step towards this comparison, in this section we summarize these three tasks. Main properties of these three strategies are shown in Table 2, and we describe them in detail below.

Direction prediction represents the simplest of these tasks. It establish financial series forecasting as a classification task, where models predict the direction of the price of a commodity at some point in the future, $t + \Delta t$ – i.e. whether it grows or goes down. As we consider these models for assisting investors, we consider a third possibility, where we identify whether prices are going to keep stale, or, at least, not subject to significant changes. In order to determine whether a price difference is significant, it is possible to consider a variation threshold (for example, 1% over the original price). Besides that threshold, direction classification methods do not consider the magnitude of the changes – only their sign.

Profitability prediction adds a complexity layer for the problem, by considering the magnitude of price changes. In order to avoid problems derived from the price scaling (cocoa prices, for example, as of August 2022 are close to 2,400$, while the natural gas price is around 8$), profitability prediction just estimates the relative growth that the price might have in the future with respect to the present one. This profitability measure is known as return on investment (ROI). If we denote the price of a commodity c at time t as $p_{c,t}$, the ROI of that asset when we look Δt time periods (e.g. days) in the future is defined as

$$\mathrm{ROI}(c, t, \Delta t) = \frac{\Delta p_{c,t,\Delta_t}}{p_{c,t}} = \frac{p_{c,t+\Delta t} - p_{c,t}}{p_{c,t}} \quad (1)$$

ROI values greater than 0 indicates that the future value is greater than the present one, whereas values smaller than 0 indicate a fall on the commodity price. Differently from the direction prediction task, which could be treated as a classification task as it deals with a finite number of targets, ROI takes continuous values in the range, $[-1, \infty)$. Consequently, we apply regression algorithms to solve the task.

Price prediction represents the last approach we consider for the task and the most complex one. As the name indicates, price prediction applies regression methods to approximate the value of the future price. It represents the most complex task. Algorithms addressing this task not only have to deal with the magnitude of pricing variations, but also need to consider the price ranges on which the different commodities move.

4 EXPERIMENTAL SETUP

In order to compare the three price forecasting strategies introduced in section 3, we run several offline experiments aimed at determining the effectiveness of the different approaches at generating profits for potential commodities investors. In this section, we describe the experimental setup we use.

4.1 Dataset

We build a dataset from pricing and volume data downloaded from Yahoo! Finance[2] , combining data for 20 commodities. For each commodity, we extract daily prices, including the opening, closing, highest and lowest prices, and we also consider the volume, i.e. the number of units of the commodity which are sold within each day. Time series data for each commodity extends 20 years, from January 2001 to December 2020. Table 3 shows a 5 day sample from the natural gas data.

Table 3: Sample of 5 days of Natural Gas Futures price and volume data

Date	Open	High	Low	Close	Volume
2016-01-04	2.366	2.375	2.239	2.334	135238.0
2016-01-05	2.290	2.346	2.255	2.325	123923.0
2016-01-06	2.345	2.360	2.241	2.267	166278.0
2016-01-07	2.290	2.429	2.271	2.382	214462.0
2016-01-08	2.397	2.495	2.377	2.472	163907.0

As noted previously, many of the prior works on commodities tend to examine a single sector. For example, Herrera et al. [7] focus on energy commodities. We depart from this, instead selecting a basket of 20 commodities from 5 different sectors: energy, crops, livestock, metals and construction. Table 4 contains a listing of commodities used, grouped by sector. As assets from within a single sector tend to show strong price correlation, e.g. if the oil price rise so too does the price of petrol, having a wide range of commodities from different sectors allows for checking the robustness of our methods [1].

4.2 Methods

For the three different tasks (price prediction, trend detection and direction classification), we have considered variations of the same machine learning algorithm: random forest. Random forest models have previously been applied successfully to financial series forecasting [7, 10, 16]. In the case of direction classification, we have considered a random forest classifier, whereas for the other two

[2]https://finance.yahoo.com/commodities

Table 4: Commodities in the dataset, grouped by sector

Sector	Commodities
Energy	Brent Crude, Gasoline, Heating Oil, Natural Gas, West Texas Intermediate (WTI) Crude
Crops	Corn, Coffee, Cocoa, Oats, Wheat, Sugar
Livestock	Feeder Cattle, Lean Hogs, Live Cattle
Metals	Copper, Palladium, Platinum, Gold, Silver
Construction	Lumber

tasks, we apply a regression algorithm based on random forests. For both regression and classification, random forest models are resistant to data scaling issues [2]. In all of our models, we use 100 trees.

4.2.1 Feature generation. As input to our models, we build feature vectors from the time series of each commodity. We consider four types of features: pricing and volume data, technical indicators, temporal information and commodity identifier.

- **Pricing and volume data** include the raw opening, closing, highest and lowest prices of the commodity at prediction time, as well as the volume data.
- **Technical indicators** represent heuristic signals based on historical pricing and volume of financial assets, and which have been deemed to have predictive power when it comes to financial assets prices – and, in particular, commodities [26]. Considering the huge number of technical indicators to choose from, we make a selection considering those which have been successfully applied by other works – in particularly, the works by Kim [11], Naik and Mohan [16] and Yin and Yang [26]. In total, 14 different technical indicators are included. As these technical indicators rely on past pricing data, some of them might use different time horizons (for instance, we might compute the volatility at 1 month in the past, or at 3 months). Table 5 contains the full list of indicators, along with the number of trading days[3] we look into the past for each of them.
- **Temporal information** extracts information about (a) prediction and (b) target times. Although not common in the literature, there is an apparent periodicity on the volume and prices of some commodities (for example, the supply of some agricultural products are seasonal, and prices might react accordingly) indicating they may be helpful. Because of this, we include, as our features, month, day of the week and day of the year, as well as a one-hot encoding if the year for both the current date and the (future) prediction target.
- **Commodity information**: Finally, as our models are trained for all the commodities, we include a one-hot encoding of the commodity ticker, so models can learn to distinguish between the different assets.

Although our chosen models are resistant to data scaling issues, to further mitigate against this problem, we normalize the technical indicators and temporal features using a min-max strategy [13].

4.2.2 Targets. For training and evaluation purposes, we establish a target for each of the feature vectors. Due to the unique characteristics of every machine learning task, these targets differ between models:

- **Direction prediction:** We consider three different labels: if the future price is more than 1% higher than the current price, a '*buy*' label is assigned; if it is more than 1% lower than the current price, a '*sell*' label is set; otherwise, we assign the neutral '*hold*' label.
- **Profitability prediction:** The target of each feature vector is the future return on investment of the asset c, i.e. $\mathrm{ROI}(c, t, \Delta t)$ as defined in Equation 1.
- **Price prediction:** The target is the price of the commodity c at prediction time $t + \Delta t$: $p_{c,t+\Delta t}$

[3]Prices are only considered on days where the market is open. For instance, there is no pricing data for weekends and some holidays. We refer to these days where the market is open as financial or trading days. The financial year has 252 days, as opposed to the 365 of the natural year, and months have (roughly) 21 days.

4.3 Experimental procedure

Finally, we describe the procedures we follow in our experiments. We distinguish between two different strategies: an statistical evaluation, where we check the effectiveness of our models at minimizing the classification or regression errors, and a trading simulation-based evaluation, where we analyze the effectiveness of these methods at providing useful recommendations on how to manage a commodity portfolio.

4.3.1 Statistical Evaluation. For the statistical evaluation, we perform a temporal split into train and test segments, to ensure information does not leak between the training and test sets: data from 2001 to 2016 is used as the training set, while data from 2016 to 2021 is used forms the test set. Feature vectors in the training set are given as input to the models for training their parameters. Afterwards, predictions for every example in the test set are computed and compared with their targets to compute the prediction error. Under this evaluation perspective, the three different tasks (direction, profitability and price prediction) are not comparable, so we use three different metrics to evaluate it:

- **Direction prediction:** We use F_1 score for ternary classification.
- **Profitability prediction:** We use mean average error for this regression problem.
- **Price prediction:** Differently from profitability prediction, due to the varying scaling on the price of the different commodities, we use here the mean absolute percentage error.

4.3.2 Simulator Evaluation. The three prediction strategies produce output for which there is no single statistical measure; complicating direct comparison. We test them using a simulator, so we can compare them under a unified setting, following a similar approach to Kimoto et al. [12]. We first divide the period between January 2016 and December 2019 in 20 smaller periods (of around a months in length). We take the data before the period (from January 2001) to train our models and the data during that period as test.

At the beginning of the simulation (corresponding to day t), we consider an investor which changes its commodity portfolio daily. She has an initial capital (10,000$ in our experiments) which she equally invests among the set of commodities in our dataset. Then, each day from t, the trained model provides predictions for each commodity, which the investor uses to modify her portfolio. First, the investor sells those commodities which the model has considered not profitable, i.e. those classified as *sell* by the classifier, or those assets which are predicted to lose more than a 1% of their price according to regression algorithms; then, the investor uses that recovered cash to invest on those assets which the algorithm predicts as profitable, i.e. those classified as *buy* by the classifier, or those predicted to increase their price by more than a 1% of their

Table 5: Selection of technical indicators, and the period of time they look into the past. When ∞ is indicated, it means that they take the whole historical series (up to prediction time t) to compute the indicator.

Technical indicator	Time period (financial days)
Accumulation distribution index	∞
Average directional index	14
Commodity channel index	14
Detrended close oscillator	28
Force index	1
Momentum	1,3,5,7,14,21,28
Money flow index	∞
Moving average convergence divergence	26
On balance volume index	∞
Rate of change	1,3,5,7,14,21,28
Relative strength index	14
Return on investment	1,3,5,7,14,21,28
Volatility	3,5,7,14,21,28
Vortex indicator	14

original price. The investor does not modify the investment on the rest of commodities. This is designed to prevent the simulator simply holding cash. We repeat this process for h days (with h equal to 28 in our experiments, as it roughly corresponds to the shortest common commodities futures contract length). Finally, at the end of the simulation, we find how much money the investor has on its portfolio, considering the prices at $t + h$. All results are reported relative to the baseline of maintaining the initial equal value spread of commodities to the end of the trading period. Pseudocode for this simulation procedure is provided in Algorithm 1.

We used these 20 starting points, as averaging over them mitigates the issue of differences between training and test segments – just training over one of these periods might result into a biased model. This is of particular concern with this dataset as it contains both the 2008 financial crash as well as the Covid-19 pandemic in the train and test segments respectively.

A limitation of this approach is that it fails to model the expiration dates within this period, and therefore does not capture the potential costs of contract swaps or forced sales to avoid having to take delivery. It does however enormously simplify having to account for all 20 different commodities having a variety of different expiration dates, an issue complicated by the uncertainty surrounding precisely which contracts are included in the Yahoo's data.

5 RESULTS

This paper address the following research questions:

- **RQ1:** Which Commodity Price Forecasting Strategy is the Most Effective? (Section 5.1)
- **RQ2:** Does Prediction Effectiveness Correlate with Profit? (Section 5.2)
- **RQ3:** When do Prediction Errors Matter? (Section 5.3)

Algorithm 1: Simulation procedure

Data: m: algorithm to use
C: the set of commodities
t: initial time,
initialMoney: initial amount of cash to invest.
h: period length

for c *in* C **do**
  portfolio[c] = initialMoney / ($C \cdot p_{c,t}$)
end
for t' *in 1 to* h **do**
  buy = []
  cash = 0
  for c *in* C **do**
    pred ← m.predict(c)
    if *pred = "sell"* **then**
      cash ← cash + portfolio[c] $\cdot p_{c,t+t'}$
      portfolio[c] = 0
    end
    else if *pred == "buy"* **then**
      buy.add(c)
    end
  end
  for c *in buy* **do**
    portfolio[c] += cash / (|buy|$\cdot p_{c,t+t'}$)
  end
  return sum(portfolio[c] $\cdot p_{c,t+h}$ for c in C)
end

5.1 RQ1: Which Commodity Price Forecasting Strategy is the Most Effective?

As discussed in the related work (Section 2) commodity trading is rarely analysed, and no works have attempted a cross comparison of prediction strategies previously, so it is unclear how best to build robo-advisors or automated traders for commodity markets. Hence, we first ask whether it is more effective to predict price change direction, asset profitability, or the future price of an asset when selecting assets to trade in.

To answer this question, we train models for each prediction strategy and then apply those models for automated trading. In this way, we can compare strategy effectiveness based on the average profit each strategy provides. We evaluate profit returned when averaged over 20 trading periods, each 28 days in length, since profit will be influenced by changing market conditions. When performing this comparison, there are two notable variables that might impact performance, namely: 1) whether we provide the models with only past pricing data or provide pricing and trading volume data; and 2) how far into the future the prediction strategy is considering. Intuitively, the amount of a commodity that is being traded might be a useful indicator for predicting how that commodity will perform, however for some commodities volume data is sparse. Meanwhile, the further into the future we ask the model to predict, the more uncertain we expect it to become.

Table 6 reports the average profit returned by each prediction strategy after 28 days. Columns marked vol[✓] indicate models that use both pricing and volume data, whilst columns marked vol[✗]

only consider previous commodity prices. Each row in Table 6 reports profit for models producing predictions for a certain number of days in the future. The row header is of the form 'trading days (actual days)'.[4] The automated trader was provided $10,000 in seed money evenly distributed amongst the 20 commodities forming its starting portfolio. Reported trading profit is the difference in value between the traders final portfolio and the value of its starting portfolio both on day 28 (in this way we normalize out the affect of the overall market trend in the profit figure, i.e. the profit value is how much it 'beat-the-market' by).

Which Strategy is Best?: From Table 6, we observe the following. First, comparing the different prediction strategies based on their returned profit, we see that profitability prediction (with volume data) was the most effective, returning $283.53 on average across the 20 time periods and 9 prediction horizons tested. This strategy also produced the most profitable model, returning $529.94 when predicting for 20 trading days in the future. However, we also observe that the less granular direction predictor was also quite effective, returning $266.50 on average over 28 days. The price-based predictor was slightly worse again, with an average profit of $224.78. Hence, we can conclude that if building a robo-advisor or automated trader, it is best to directly predict commodity profitability, rather than direction of change or the future price of the asset.

Should we include Volume Data? From Table 6, comparing each trading strategy when volume data is included or omitted, we observe differences between the strategies. Specifically, both direction prediction ($266.50 vs. $241.15) and profitability prediction ($283.53 vs. $251.94) benefit from the provision of information about trading volumes. However, this trend is reversed when performing price prediction ($93.56 vs. $224.78), indicating that while volume is a useful indicator of the pricing direction or profitability for a commodity, it is not directly correlated with the magnitude of commodity price movements.

What should we set the Prediction Horizon To? From the results in Table 6, contrasting the different horizon settings, we observe that for all models, performance is maximised when predicting between 20 and 40 trading days into the future. In particular, the best direction-based predictor achieved peak profitability predicting for 40 trading days in the future with $431.13, while the profitability and price predictors performed best when predicting for 20 trading days in the future ($529.94 and $394.22, respectively). It might initially appear intuitive that the best performance will be achieved when the prediction horizon and number of days in the trading period (28 days) are closely aligned, as these results suggest. However, we note that predictions are not simply generated once, but on each of the 28 days, and portfolio changes are made each day. Therefore, we hypothesise that the observed peak returns around the 20-40 trading day horizons are likely more due to accuracy of the direction/profitability/price prediction models, which we analyse in the next section.

[4]Trading days are days the markets were open, while actual days includes intervening days where the markets were closed.

Table 6: Average profit ($) relative to baseline for each method for each prediction length (in trading days). The largest returns for each variable pair (prediction horizon & volume inclusion) are highlighted in bold. are highlighted in bold.

Simulated Investment Returns (Actual Profit, USD)						
Prediction Horizon	Direction		Profitability		Price	
	vol[✓]	vol[✗]	vol[✓]	vol[✗]	vol[✓]	vol[✗]
1 (~1 days)	$132.10	$68.15	$7.85	$-47.93	$-16.90	$8.84
3 (~3 days)	$130.18	$123.77	$135.95	$72.55	$-5.53	$-22.64
5 (~7 days)	$146.93	$156.46	$125.83	$161.44	$-1.89	$72.33
10 (~14 days)	$332.63	$313.61	$410.11	$325.81	$54.55	$258.10
20 (~28 days)	$397.06	$328.92	**$529.94**	**$438.82**	$155.44	**$394.22**
40 (~57 days)	**$431.13**	**$357.90**	$424.69	$406.30	**$182.87**	$360.43
60 (~91 days)	$329.41	$304.31	$331.48	$343.48	$170.18	$358.18
80 (~120 days)	$250.05	$287.99	$292.17	$292.26	$165.27	$338.84
100 (~148 days)	$249.02	$229.21	$293.77	$274.74	$138.05	$254.69
Averages	$266.50	$241.15	$283.53	$251.94	$93.56	$224.78

5.2 RQ2: Does Prediction Effectiveness Correlate with Profit?

In the previous section, we reported which of the three strategies was the most effective in terms of making profit for the customer when used for trading. However, when training these prediction models a proxy metric is used instead, such as classification F_1, mean average error (MAE) against the profitability prediction, or mean average percentage error (MAPE) against the price prediction. However, how similar are these metrics? Just because prediction effectiveness increases under one of these proxies does not necessarily imply that a trader using this signal will make more money (since this will depend upon the trading strategy). Hence, we ask whether trading profit and these prediction proxy metrics correlate.

To answer this question, Table 7 reports prediction model performance of the three proxy metrics for the same experimental settings used for evaluating trading profitability. Prediction effectiveness is calculated per-commodity per-day in the test period and then averaged. Since a trading profit value (as reported in Table 6) represents an average of twenty 28-day periods, while one of these proxy metrics represents an accuracy or error value averaged over every day in the test period, we cannot apply a standard correlation metric. Hence, we instead contrast the trends observed as we vary the prediction horizon. Contrasting Table 7 with Table 6, we observe the following:

Profitability vs. Direction Prediction: First, considering direction prediction (F_1) effectiveness, we see that as the prediction horizon increases F_1 performance also increases. This appears counter-intuitive, until you consider that the directional classes are defined in terms of %-change for the horizon period. As the horizon period extends, less commodities will remain in the +/-1% hold bound, and as such the task becomes easier as it trends towards binary buy/sell classification rather than buy/hold/sell classification. Comparing direction prediction against trading profitability, we see that initially as direction prediction F_1 increases the trading profitability also increases. However, after the 40 day horizon these metrics diverge, with F_1 continuing to increase, while trading profitability decreases.

Table 7: F1, MAPE and MAE scores for the corresponding model, for each prediction length in trading days. Results for the best performing model for each variable pair (prediction horizon & volume inclusion) are highlighted in bold.

Prediction Effectiveness						
Prediction Horizon	Direction (F_1)		Profitability (MAE)		Price (MAPE)	
	vol[✓]	vol[✗]	vol[✓]	vol[✗]	vol[✓]	vol[✗]
1 (~1 days)	0.4498	0.4483	**0.0146**	**0.0142**	**0.0161**	**0.0157**
3 (~3 days)	0.3762	0.3800	0.0285	0.0259	0.0298	0.0294
5 (~7 days)	0.3817	0.3769	0.0380	0.0341	0.0399	0.0389
10 (~14 days)	0.4091	0.4074	0.0515	0.0493	0.0565	0.0553
20 (~28 days)	0.4467	0.4523	0.0727	0.0708	0.0845	0.0810
40 (~57 days)	0.4698	0.4690	0.1027	0.1022	0.1322	0.1185
60 (~91 days)	0.4883	0.5007	0.1278	0.1222	0.1490	0.1431
80 (~120 days)	0.5005	**0.5244**	0.1501	0.1312	0.1693	0.1620
100 (~148 days)	**0.5049**	0.5218	0.1638	0.1442	0.1896	0.1684

Profitability vs. Profitability Prediction: Second, if we examine profitability prediction error (MAE), as we might expect, when the model tries to predict further into the future the prediction error increases. In particular, profitability MAE increases approximately linearly from 0.0146 to 0.1638 as we transition from predicting from 1 to 100 days in the future. However, if we contrast this against trading profitability, this increasing error does not appear to negatively impact the trader until predicting more than 20 working days into the future.

Profitability vs. Price Prediction: Finally, examining price prediction error (MAPE), we see a similar pattern to profitability prediction, where as the model tries to predict further into the future the prediction error increases (from 0.0161 to 0.1896). Contrasting this against trading profitability, again we see that this increasing error does not negatively impact profits until predicting more than 20 working days into the future.

From these results, we can conclude that these proxy prediction effectiveness metrics don't follow the same trend as trading profitability as the horizon increases. Most notably, the profitability and price predictors error margins increase dramatically as the prediction horizon increases, but this does not negatively impact profit made until predicting over 20 days into the future (approximately 0.07 MAE and 0.085 MAPE, respectively).

5.3 RQ3: When do Prediction Errors Matter?

Above we have shown that there are cases where increases in prediction error does not result in reduced downstream performance, so why is this? There are multiple factors that explain this unusual behaviour, which we discuss below:

Trader Error Sensitivity: First, the trading simulator by design largely ignores prediction magnitudes when trading (this was to avoid the trading simulator becoming a confounding variable when comparing prediction strategies). As a result, not all errors are equal. Incorrectly predicting a commodity will lose value is far more damaging than incorrectly predicting a commodity will gain value, since the former will trigger the selloff of all held assets of that type, while the latter will trigger a small increase in holdings of that asset.[5] Meanwhile, the magnitude of the prediction is only considered when distinguishing between hold and either buy or sell positions, e.g. a predicted profit of 1.5% when actual profit was only 0.5% would matter, since the error causes the 1% threshold to be crossed, where predicted profit of 8% when actual profit was only 2% does not impact the trading strategy (as both trigger a buy action).

Ignored Commodities: Second, the proxy metrics consider all commodities as equal. However, a trader will only trade in an asset if it is believed profitable or to limit a future loss. This means that any errors associated with assets predicted to be stable will not influence the trader, since those assets will never be exchanged and thus their value in the final result will be neutral relative to the baseline strategy. To illustrate this, Table 8 reports the model error and average dollar value of each commodity bought and sold by the profitability prediction strategy (vol[✗], prediction horizon=1,20,100) over the 20 time periods.[6] Due to the marked disparity in trader exposure to different commodities, performance is skewed towards those most heavily traded. For example for prediction horizon=20, it is clear that the accuracy of natural gas buy predictions are more influential on final performance than the buy predictions of sugar. This is a particular issue where there model outputs predictions which result in only buy or sells. Prediction horizon=1 is dominated by sell actions, hence the accuracy of positive profitability predictions for most commodities are not influencing the downstream profit made. This is likely a consequence of the phenomenon noted above, price movements of +/-1% are much less likely for shorter prediction horizons than longer ones, therefore the number of days where the trader does not alter their portfolio is higher for shorter prediction lengths.

6 CONCLUSIONS & FURTHER WORK

Robo-advisors and automated trading programs that use models to predict how assets will change in value over time are becoming increasingly prevalent in financial marketplaces. There are a range of prediction strategies these systems use, namely: direction prediction; profitability prediction; and price prediction. However, these strategies are rarely compared, and this has never been done for the influential class of commodity markets. Hence, to tackle this knowledge gap, we examined the effectiveness of these three financial asset prediction strategies when applied to commodity trading marketplaces, both from downstream trading profitability and model accuracy perspectives. Through experiments over a large dataset spanning almost 20 years of data for 20 diverse commodities we demonstrated that prediction of profitability is 6% more effective on average than direction prediction and more than 20% better than price prediction. We also unexpectedly observed that optimal prediction horizons were between 20 and 40 trading days, rather than shorter horizons where the prediction error margins were lower. Indeed, our experiments highlighted a distinct lack of correlation between traditional statistical measures (F_1, MAE and MAPE) and trading profitability.

[5] The amount bought is dependent on the number of other predicted profitable assets that day and the amount sold that day.

[6] Note that this data contains a slight sell bias - different commodity exchanges have different trading days. Where a commodity has no price data on the final day of the simulation it is instead sold on the last day within the trading period for which there is data.

Table 8: Profitability prediction error (MAE) and average dollar value of each commodity bought and sold by the price prediction strategy (vol[✗], prediction horizons=1,20,100) over the 20 time periods. ▲ denotes commodities where the model purchased $50 more than it sold on average, while ▼ denotes the reverse.

Commodity	Horizon = 1 day		Horizon = 20 days		Horizon = 100 days	
	MAE	Bought/Sold	MAE	Bought/Sold	MAE	Bought/Sold
Brent Crude	0.0175	0.00/47.99▼	0.0856	489.56/462.31	0.1973	289.87/258.17
Gasoline	0.0195	23.20/100.80▼	0.0979	358.51/396.77	0.1965	229.10/281.36▼
Heating Oil	0.0158	26.90/25.00	0.0795	463.54/337.80▲	0.2536	399.52/179.69▲
Natural Gas	0.0258	248.94/52.66▲	0.1064	819.21/373.69▲	0.219	344.20/259.51▲
WTI Crude	0.0205	0.00/75.93▼	0.1031	560.98/433.80▲	0.2389	319.86/310.20
Corn	0.0113	0.00/24.11▼	0.0539	460.28/493.24	0.1328	300.00/402.22▼
Coffee	0.0147	0.00/24.49▼	0.0657	428.33/702.30▼	0.1652	215.25/424.97▼
Cocoa	0.0148	28.86/50.88▼	0.0683	393.89/584.74▼	0.126	238.88/389.82▼
Oats	0.0151	0.00/50.31▼	0.0749	622.09/489.81▲	0.0861	524.73/408.39▲
Wheat	0.0137	0.00/74.24▼	0.0642	498.43/621.69▼	0.1529	299.40/426.70▼
Sugar	0.0165	0.00/27.96▼	0.0781	288.67/718.06▼	0.2005	204.69/426.52▼
Feeder Cattle	0.0089	0.00/0.00	0.0428	352.40/434.42▼	0.0858	246.92/309.68▼
Lean Hogs	0.0182	26.09/150.76▼	0.1166	641.07/510.07▼	0.1668	324.24/384.16▼
Live Cattle	0.0098	49.51/27.45▲	0.0489	440.54/401.37	0.1314	160.50/378.13▼
Copper	0.0094	0.00/27.62▼	0.0505	446.13/360.33▲	0.1831	259.78/182.06▲
Palladium	0.0147	0.00/24.33▼	0.074	344.93/479.55▼	0.2228	567.65/142.49▲
Platinum	0.0116	0.00/0.00	0.0532	415.75/307.18▲	0.0941	331.79/199.88▲
Gold	0.0068	0.00/0.00	0.0431	394.80/376.56	0.0854	476.44/248.37▲
Silver	0.0109	0.00/25.08▼	0.0591	366.63/598.04▼	0.1192	158.49/401.72▼
Lumber	0.0169	0.00/51.13▼	0.089	486.40/553.89▼	0.2187	356.39/257.97▲
Average	0.0146	20.18/43.04	0.0727	463.61/481.78	0.1638	312.38/313.60

More generally, our experiments highlight the importance of using simulations to evaluate the practical application of time-series forecasting models rather than just using statistical metrics, but also demonstrates some of the challenges when doing so. Firstly, using a simulated trader allows for comparison between otherwise incomparable model outputs and secondly, provides greater context for the interpretation of the statistical performance indicators. On the other hand, the design of the simulator can strongly influence the experimental results. For instance, the trader reported in our experiments only buys or sells commodities predicted to change in value by at least 1%. This means that the trader is insensitive to prediction errors within the 1 to -1% range. Additionally, as this trader was designed to be applicable to all three trading strategies, it lacks the capacity to utilise profit or price change margins if such predictions are available, which might make such strategies more attractive in practice. Furthermore, as we demonstrated, errors across commodities are not equally important, since the trader will frequently interact with only those assets that are predicted to be profitable.

Given these results we believe that the higher performance observed by the profitability prediction strategy, incentives further exploration of this technique, which is less studied in the literature than the more popular price regression and direction classification strategies. Meanwhile, in the future we aim to expand our analysis to include a thorough comparison of different trading strategies and examine the impact they have on profitability.

REFERENCES

[1] Nathan S Balke, Stephen PA Brown, and Mine K Yucel. 1998. Crude oil and gasoline prices: an asymmetric relationship? *Economic Review-Federal Reserve Bank of Dallas*, 2.

[2] Leo Breiman. 2001. Random forests. *Machine learning*, 45, 1, 5–32.

[3] Zhiyuan Chen, Howe Seng Goh, Kai Ling Sin, Kelly Lim, Nicole Ka Hei Chung, and Xin Yu Liew. 2021. Automated agriculture commodity price prediction system with machine learning techniques. *arXiv preprint arXiv:2106.12747*.

[4] Antonio Gargano and Allan Timmermann. 2014. Forecasting commodity price indexes using macroeconomic and financial predictors. *International Journal of Forecasting*, 30, 3, 825–843.

[5] Gary Gorton and K Geert Rouwenhorst. 2006. Facts and fantasies about commodity futures. *Financial Analysts Journal*, 62, 2, 47–68.

[6] Hasraddin Guliyev and Eldayag Mustafayev. 2022. Predicting the changes in the WTI crude oil price dynamics using machine learning models. *Resources Policy*, 77, 102664.

[7] Gabriel Paes Herrera, Michel Constantino, Benjamin Miranda Tabak, Hemerson Pistori, Jen-Je Su, and Athula Naranpanawa. 2019. Long-term forecast of energy commodities price using machine learning. *Energy*, 179, 214–221. DOI: https://doi.org/10.1016/j.energy.2019.04.077.

[8] Chih-Ming Hsu. 2011. A hybrid procedure for stock price prediction by integrating self-organizing map and genetic programming. *Expert Systems with Applications*, 38, 11, 14026–14036.

[9] Deniz Igan, Emanuel Kohlscheen, Gabriela Nodari, Daniel Rees, et al. 2022. Commodity market disruptions, growth and inflation. Tech. rep. Bank for International Settlements.

[10] Luckyson Khaidem, Snehanshu Saha, and Sudeepa Roy Dey. 2016. Predicting the direction of stock market prices using random forest. *arXiv preprint arXiv:1605.00003*.

[11] Kyoung-jae Kim. 2003. Financial time series forecasting using support vector machines. *Neurocomputing*, 55, 1-2, 307–319.

[12] Takashi Kimoto, Kazuo Asakawa, Morio Yoda, and Masakazu Takeoka. 1990. Stock market prediction system with modular neural networks. In *1990 IJCNN international joint conference on neural networks*. IEEE, 1–6.

[13] Joon Ho Lee. 1997. Analyses of Multiple Evidence Combination. In *Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 1997)*. ACM, Philadelphia, PA, USA, (July 1997), 267–276. DOI: 10.1145/258525.258587.

[14] Burton G Malkiel. 1989. Efficient market hypothesis. In *Finance*. Springer, 127–134.

[15] Pankaj Mehta, Marin Bukov, Ching-Hao Wang, Alexandre GR Day, Clint Richardson, Charles K Fisher, and David J Schwab. 2019. A high-bias, low-variance introduction to machine learning for physicists. *Physics reports*, 810, 1–124.

[16] Nagaraj Naik and Biju R Mohan. 2019. Stock price movements classification using machine and deep learning techniques-the case study of indian stock market. In *International Conference on Engineering Applications of Neural Networks*. Springer, 445–452.

[17] BBC News. 2022. Bank governor in 'apocalyptic' warning over rising food prices. (May 2022). https://www.bbc.co.uk/news/business-61469532.

[18] BBC News. 2022. What is the uk inflation rate and why is the cost of living rising? (Aug. 2022). https://www.bbc.co.uk/news/business-12196322.

[19] Lucas Nunno. 2014. Stock market price prediction using linear and polynomial regression models. *Computer Science Department, University of New Mexico: Albuquerque, NM, USA*.

[20] Venkata Vara Prasad, Srinivas Gumparthi, Lokeswari Y. Venkataramana, S. Srinethe, R. M. Sruthi Sree, and K. Nishanthi. 2022. Prediction of stock prices using statistical and machine learning models: A comparative analysis. *Comput. J.*, 65, 5, 1338–1351. DOI: 10.1093/comjnl/bxab008.

[21] Ali Safari and Maryam Davallou. 2018. Oil price forecasting using a hybrid model. *Energy*, 148, 49–58. DOI: https://doi.org/10.1016/j.energy.2018.01.007.

[22] Chih F Tsai and Sammy P Wang. 2009. Stock price forecasting by hybrid machine learning techniques. In *Proceedings of the international multiconference of engineers and computer scientists* number 755. Vol. 1, 60.

[23] Mehar Vijh, Deeksha Chandola, Vinay Anand Tikkiwal, and Arun Kumar. 2020. Stock closing price prediction using machine learning techniques. *Procedia computer science*, 167, 599–606.

[24] Haiyao Wang, Jianxuan Wang, Lihui Cao, Yifan Li, Qiuhong Sun, and Jingyang Wang. 2021. A stock closing price prediction model based on cnn-bislstm. *Complex.*, 2021, 5360828:1–5360828:12. DOI: 10.1155/2021/5360828.

[25] Halbert White. 1988. Economic prediction using neural networks: the case of ibm daily stock returns. In *ICNN*. Vol. 2, 451–458.

[26] Libo Yin and Qingyuan Yang. 2016. Predicting the oil prices: do technical indicators help? *Energy Economics*, 56, 338–350.