Erasmus University Rotterdam
Erasmus School of Economics
Bachelor Thesis Econometrics & Operations Research

Performance of Regularization Methods in Commodity Price Forecasting: A Comparative Analysis

Daan van Peer (572124)

Supervisor:	Prof.Dr. D.J.C. van Dijk
Second assessor:	T. van der Zwan, PhD candidate
Date final version:	29th June 2023

Abstract

The literature on forecasting commodity prices with the application of technical variables is growing. When using these technical variables there is less attention to selecting suitable technical indicators, which leads to a fall in general performance. This research analyses the forecasting performance of various models when forecasting commodity prices during the time period from January 1992 to December 2022. This paper is focused on the use of variable selection through the regularization methods LASSO and elastic net. Additionally, the combination elastic net method is applied to find improvement in forecasting performance. The results show that when LASSO and elastic net methods are executed with technical variables the forecasting performance increases significantly. These forecasts are evaluated by means of percentage out-of-sample R^2, Clark and West statistics and the Model Confidence Set procedure.

1 Introduction

Forecasting commodity prices can provide valuable insights for investors, businesses and policymakers. Returns on commodities can be lucrative when the price change is forecasted correctly. If the price of a certain commodity is forecasted to increase, investors can decide to allocate a bigger share of their portfolio to companies that produce this commodity. Moreover, businesses that are in need of certain commodities can make more informed decisions on their inventory when using a more precise forecast. Also, policymakers benefit from forecasting commodity prices. Due to the fact that commodity prices have a significant influence on the overall economy. Just like Gargano and Timmermann (2014) state: 'Commodity prices are widely believed to influence price levels more broadly, and thus are of interest to central banks, policymakers, firms and consumers whose decisions depend on their expectations of future inflation.' Therefore, this paper will discuss the forecasting performance of various models that forecast eight commodity price indices.

In this paper, we forecast commodity prices for the time period from January 1992 until December 2022. We use a data sample of eight commodity price indices and some macroeconomic and technical variables corresponding to the time period from January 1982 until December 2022. The methods that we used to forecast the commodity prices, like the LASSO method and elastic net method, are performed first. After that, we evaluate the forecasts made by the models with forecast evaluation methods like percentage out-of-sample R^2, Clark and West statistics and the Model Confidence Set procedure. The paper is focused on the use of regularization methods to forecast commodity prices. Regularization methods have the attractive property of variable selection. Meaning, the methods will select the most important variables for the model and this helps against overfitting. The general forecasting performance increases when the problem of overfitting is successfully challenged.

The works of Wang, Liu and Wu (2020) inspired us to use five technical variables based on five popular trading rules. These trading rules, like the momentum rule that is based on the trend of the price, consist of multiple technical indicators. These indicators provide buying and selling signals following mathematical rules for the price of the commodity. However, due to the relatively big amount of indicators, the problem of overfitting can emerge. That is why, in this research, we study if the regularization methods enhance forecasting performance.

The main goal of this research is to answer the question, how do regularization methods impact the performance of forecasting commodity prices when using both conventional variables and technical indicators? The conventional variables are mostly used in past literature. We use them to compare the forecasting performance of the technical models. We also want to know which indicators are truly significant for the making of the technical models. Moreover, we study how the models made with regularization methods perform compared to each other.

Due to the fact that the data sample stretches until December 2022. The current energy crisis is contained in the sample and it will be interesting to study this time period because of the volatile energy prices. A second captivating period in the sample is the covid-pandemic time period. The commodity prices increased remarkably during this crisis. Are the technical indicators still useful in times of volatility? Or are they even more helpful when we want to forecast commodity prices during a crisis?

The main findings of the research are that the regularization methods indeed improve the forecast performance. The percentage out-of-sample R^2 values of the models made with LASSO and elastic net are for almost all of the technical models relatively higher than the technical models made with ordinary least squares. Only the model containing all the macroeconomic variables performs relatively worse when using regularization methods. Also, the p-values of the Model Confidence Set procedure have provided the results that the technical models outperform the conventional models. We found that when we compared the models made with regularization methods, there is no significant winner.

There has been done a lot of research on the topic of commodity prices. The majority of this literature uses macroeconomic or financial variables to forecast commodity prices. For example, the works of Gargano and Timmermann (2014), Borensztein and Reinhart (1994), Chen, Rogoff and Rossi (2010), Akram (2009) and Groen and Pesenti (2011). For instance, Akram (2009) identifies that commodity prices increase when real interest rates decrease. In the paper of Borensztein and Reinhart (1994) they discuss the proportion of variance of commodity prices that is explained by macroeconomic determinants. There are also papers of the past that research the effect of technical indicators like Wang et al. (2020), Moskowitz, Ooi and Pedersen (2012) and Fuertes, Miffre and Rallis (2010). For example, Moskowitz et al. (2012) find a significant effect of 'time series momentum' on the forecast of future contracts, where this 'time series momentum' is based on the momentum of the price. This 'time series momentum' is comparable to the momentum rule indicators we use in the research. The works of Wang et al. (2020) note that multiple technical indicators outperform the conventional macroeconomic variables in forecasting commodity prices. Where Wang et al. (2020) specifically focus on one of the most conventional regression methods, namely ordinary least squares, we extend the amount and complexity of the methods used to forecast commodity prices. Namely, we apply regularization methods because they fit with the research question of Wang et al. (2020). The LASSO method, proposed by Tibshirani (1996), permits shrinkage of coefficients to exactly zero, this means the method uses variable selection (Rapach & Zhou, 2020). That is how we can further investigate the significance of certain technical indicators. Another regularization method that we use is the elastic net method, proposed by Zou and Hastie (2005). The elastic net method is similar to the LASSO method, however, it can prevent the problem of arbitrary variable selection arising when the correlation between predictors is high. The paper of Rapach and Zhou (2020) suggests the method of combination elastic net, where conventional variables and technical indicators come together in a forecast combination for the US market excess return. The combination elastic net method can provide a clear selection of the significant models. Combination elastic net also seems suitable for the forecasting of commodity prices.

The rest of the paper is structured as follows. First, Section 2 concisely describes the data that is used for the research. Second, Section 3 clarifies which methods will be utilized in order to perform and evaluate the forecasting. Thereafter, in Section 4 the results of the research will be presented with corresponding reflection. Lastly, Section 5 will conclude the paper.

2 Data

2.1 Commodity Prices

In this paper, we use a data sample of eight commodity price indices in the time period from January 1982 until December 2022. The data that is used for the commodity price indices, inspired by the paper of Wang et al. (2020), can be obtained from the World Bank's website[1]. The World Bank provides commodity price indices that can be split up into two categories: energy commodity prices and non-energy commodity prices. We acquired monthly data on the prices of eight commodity indices with a sample period from January 1982 to December 2022. As stated by Wang et al. (2020), 'The energy index is the weighted average of the prices of coal, crude oil and natural gas, where the weight on the price of oil is 84.6%.' The non-energy commodity index is a weighted average of commodities like food, raw materials and beverages. The actual eight commodity price indices we will research are the energy index, the non-energy index, the agriculture index, the beverage index, the food index, the raw materials index, the metals & minerals index and the precious metals index. Due to the extension of the data until December 2022, the sample contains a more volatile period. Especially the energy crisis from 2021 up until now and the covid-pandemic period are interesting to research. If we compare the energy index from June 2021 with the energy index from June 2022 this value is almost doubled. Whereas, in the same time period the non-energy index increased approximately 12.5%. To provide some insight into the commodity prices from January 1982 until December 2022, Figure 1 and Figure

[1]https://www.worldbank.org/en/research/commodity-markets.

2 respectively represent the commodity prices of the energy and non-energy indices.

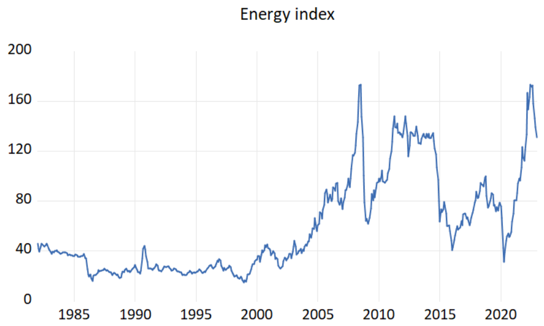

Figure 1: Commodity prices of the energy index from January 2018 until December 2022. The horizontal axis displays the time and the vertical axis the index values.

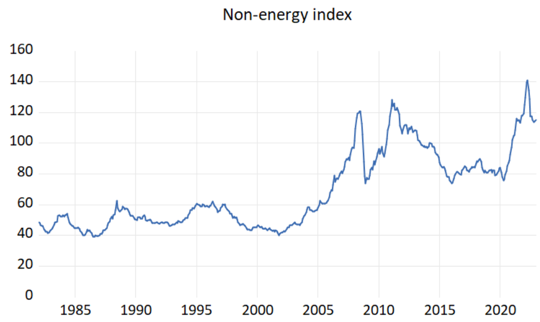

Figure 2: Commodity prices of the non-energy index from January 2018 until December 2022. The horizontal axis displays the time and the vertical axis the index values.

To talk about the returns on commodity goods we use the change in log prices as the dependent variable throughout the research. The changes in log prices are calculated as follows.

$$r_{t+1} = log(P_{t+1}) - log(P_t) \tag{1}$$

In Equation 1, P_t represents the commodity price in month t. The dependent variable r_{t+1} denotes the change in log prices in month $t+1$, also known as the return on the commodity good in month $t+1$.

2.2 Macroeconomic Variables

This research uses two different types of variables to try and forecast commodity prices. The two types are macroeconomic variables and technical indicators. For the macroeconomic variables, eight predictors have been selected based on the works of Gargano and Timmermann (2014). The eight explanatory variables are: the dividend price ratio computed as the difference between the log of the 12-month moving sum of dividends and the log of the S&P 500 index (DP); the secondary market 3-month Treasury bill rate (TBL); the long-term government bond yield (LTY); trade-weighted exchange rates of the U.S dollar (EX); stock excess returns of the S&P 500 index (RET); inflation computed as the log growth in consumer price index for all urban consumers (INFL); industrial production in the United States (IP) and global real economic activity where the Kilian's index is used as a proxy (KI).

The data was gathered from several sources. We extracted the consumer price index to create the INFL variable from the FRED-MD data base[2]. The values for industrial production in the United States are also gathered from the FRED-MD data base. Following Wang et al.

[2]https://fred.stlouisfed.org/.

(2020), we use an extra lag for the variables IP and INFL due to the delay of information. The Kilian's index is found at the website of the reserve bank of Dallas [3]. The homepage of Amit Goyal[4] provides us the variables TBL, LTY, RET and DP. The EX variable is part of two series[5] we got from the FRED-MD database. One of the indices starts in the year 1973 but stops in 2019, so to get the last years we needed the other series which starts in 2006 and proceeds until today.

To check if the macroeconomic variables we use are stationary, the augmented Dickey-Fuller test is executed (Dickey & Fuller, 1979) in Eviews. This way we test the null hypothesis that a unit root is present in a time series. Every variable caused a rejection of the null hypothesis, except the variables IP and EX. We took the log differences for the IP and EX variables to get stationary time series.

2.3 Technical Indicators

For the technical indicators Wang et al. (2020) used five popular technical trading rules to generate variables for the forecasting of commodity prices. The variables they have created are; the momentum rule (MOM), which basically indicates if the momentum of the price movement is upwards or downwards; the filtering rule (FR), which produces a buying (selling) signal when the commodity price has risen (dropped) above (below) its most recent low (high) more than a given percentage; the moving average rule (MV), a rule that gives a buying (selling) signal when the short-term moving average is higher (lower) than the long-term moving average; the oscillator trading rule (OSLT), which generates a buying or selling signal based on the works of Levy (1967), that studies the relative strength indicator; support and resistance rule (SR), this rule indicates whether to buy or sell the commodity based on the comparison of the current price and the support or resistance levels.

Each of these technical variables consists of multiple technical indicators created via formulas based on the trading rules stated above. Further description of the calculation of these technical indicators is stated in Appendix A. For our research, it is important to understand that all trading rules have multiple indicators. The number of indicators per trading rule is different. To create the technical variables MOM, FR, MV, OSLT and SR we use 10, 20, 10, 20 and 50 technical indicators respectively. That makes a total of 840 technical indicators generated with Java code[6]. However, in our analysis, we found a few variables containing small differences in signals (too few ones[7]). Some 'extreme' cases of these indicators are therefore neglected for the research. We included a full description of which indicators are omitted in Appendix C.

[3]https://www.dallasfed.org/research/igrea.

[4]https://sites.google.com/view/agoyal145.

[5]The FRED-MD codes of the series are: TWEXAFEGSMTH and TWEXMMTH.

[6]Further explanation on the code is stated in Appendix B.

[7]The ones or zeros indicate whether a signal is provided or not. With too few ones, it is meant that there are too few signal indications.

3 Methodology

3.1 Linear Predictive Regression Model

To understand the basics of our research we first describe a univariate linear regression model. This is a model that uses only one predictor as explanatory variable. We can use various methods to obtain the coefficients in Equation 2. For example, the first method we use to estimate the coefficients is the ordinary least squares method. This is a well-known method in the world of econometrics and is therefore very common to use. Let us look at a univariate regression formulated as follows.

$$r_{t+1} = \alpha + \beta x_t + \epsilon_{t+1} \tag{2}$$

Where in Equation 2 the r_{t+1} denotes the return on commodities in period t+1, the α is the regression constant, the ϵ_{t+1} represents the error term of period $t + 1$ and the x_t denotes the explanatory variable with corresponding coefficient β. In this regression, we assume that the error terms are identically and independently distributed with corresponding variance σ^2_ϵ. The ordinary least squares method aims to estimate these α and β by minimizing the total sum of all errors squared. For more information on the ordinary least squares method, one can look at the works of Heij, de Boer, van Dijk, Kloek and Franses (2004).

3.2 Regularization Methods

3.2.1 LASSO

The first regularization method that we discuss is the Least Absolute Shrinkage and Selection Operator method, also known as the LASSO method. This machine learning method allows coefficients of certain predictors to be shrunk to zero. The method uses a penalty term (ℓ_1) to identify the relevant explanatory variables and thus provide a solution to overfitting. Especially with the models based on our technical indicators, the amount of predictors is big. The random noise contained in the data will be falsely identified by predictors that have less explanatory power. Therefore, the variance of the model is unreasonably higher and the general performance and the interpretability of the model drop.

The LASSO method is a combination of a maximum likelihood estimation and a penalty term correcting the sum of the absolute values of the regression coefficients. The following expression illustrates how the LASSO coefficients are estimated.

$$\arg \min_{\alpha,\beta_1,\ldots,\beta_j} \left(\frac{1}{2t} \sum_{i=1}^{t} \left(r_i - \alpha - \sum_{j=1}^{p} \beta_j x_{j,i-1} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right) \tag{3}$$

The left part in Equation 3 portrays the squared errors of an OLS regression with the notation from Equation 2. The amount of predictors that are included is indicated with the variable p.

The right part represents the penalty term on the coefficients. The regularization parameter λ in this expression determines the degree of shrinkage. When the value of λ increases, more predictor coefficients will be shrunk to zero.

To pick the optimal value for the regularization parameter we use K-fold cross-validation. This process divides the data set into K subsamples and then with the exception of one subsample these subsamples are used to train the model. The remaining subsample is used for testing the trained model. This process will be repeated K times to finish the K-fold cross-validation. The λ that provides the smallest cross-validation error is then selected for the LASSO regression (Parzinger et al., 2022). In our research, we use $K = 5$, due to the fact that this is a reasonable split to create a training set considering the amount of observations in our sample.

If we use the LASSO method for the technical variables we incorporate all the technical indicators that correspond to a technical variable. For instance, to get the LASSO model for the MOM variable we include the five indicators that are originally used to generate the MOM variable. After that, the LASSO method determines which of the indicators are significant for the model and we get the LASSO model for the MOM variable, the MOM-L model.

For the sake of performance comparison, we also created a multivariate model containing all macroeconomic variables. This model is estimated via OLS, which will be indicated by the MACRO model. This way we can compare the forecasting performance of the same multivariate model estimated with LASSO indicated with MACRO-L.

3.2.2 Elastic Net

A probable drawback of the LASSO method is that it does not work well with a group of explanatory variables that are highly correlated. The method will arbitrarily choose the variables due to the correlation. In our analysis, we use macroeconomic and technical variables. The macroeconomic variables are not highly correlated[8]. However, the technical indicators that build the technical variables are highly correlated. The indicators based on the same trading rule show high correlation but also when indicators from different trading rules are compared they show high correlation[9]. Therefore, we perform the elastic net method that is proposed by Zou and Hastie (2005), because of its power to select such a highly correlated group as a group itself. The expression of the elastic net is very similar to the LASSO one and is as follows.

$$\arg \min_{\alpha,\beta_1,\ldots,\beta_j} \left(\frac{1}{2t} \sum_{i=1}^{t} \left(r_i - \alpha - \sum_{j=1}^{p} \beta_j x_{j,i-1} \right)^2 + \lambda \left(\alpha \sum_{j=1}^{p} |\beta_j| + \frac{1}{2}(1-\alpha) \sum_{j=1}^{p} \beta_j^2 \right) \right) \tag{4}$$

The last part of Equation 4 where we sum the squared coefficients, is the so-called ℓ_2 penalty

[8]See Appendix D for the full correlation matrix of the macroeconomic variables.
[9]See Appendix D for an example of a correlation matrix of some of the technical indicators.

term, this term can also be found in the Ridge regularization method (Hoerl & Kennard, 1970). The variables of this elastic net expression are again based on Equation 2 with the regularization parameter λ. The new variable α determines the balance between the two penalty terms ℓ_1 and ℓ_2. In our analysis, we choose $\alpha = 0.5$, based on a similar research of Rapach and Zhou (2020).

Again the same technical models and the multivariate model with macroeconomic variables are estimated like we described in Section 3.2.1. However, now we use the elastic net method instead of LASSO, indicating these models with names like MOM-E and MACRO-E.

3.3 Forecasting

To generate our forecasts we need an in-sample period and an out-of-sample period. The sample of T observations is split up into the first M in-sample observations and the remaining $T - M$ out-of-sample observations. In this paper, we use a recursive estimation window to forecast the out-of-sample part of the observations. This means that the estimates of α and β are updated when we move to the next out-of-sample forecast. The in-sample part expands when the time proceeds.

After the α and β are estimated via one of the estimation methods OLS, LASSO or elastic net, the out-of-sample forecasts are given by Equation 5.

$$\hat{r}_{k+1} = \hat{\alpha_k} + \hat{\beta}_k x_k \tag{5}$$

When the estimation of the coefficients is performed for returns on time k through Equation 2, we get $\hat{\alpha_k}$ and $\hat{\beta}_k$ and use them in Equation 5 as estimates of α and β. We then repeat this whole process for $k = M, ..., T - 1$ to get our out-of-sample forecasts.

Due to the big amount of technical indicators we choose to average the forecasts per each trading rule. Resulting in five time series of forecasted returns belonging to the technical variables MOM, FR, MV, OSLT and SR.

Now we have the out-of-sample forecasts from both macroeconomic and technical variables. However, it is still risky to rely on models that consist of a single variable. Like stated in Wang et al. (2020), 'It is well-known that it is dangerous to rely on a single model's forecasts due to the problem of model uncertainty.' A solution to this problem is a technique called forecast combination. This technique gives weights to generated forecasts of certain models.

$$\hat{r}_{t,comb} = \sum_{j=1}^{p} w_{j,t}\hat{r}_{j,t} \tag{6}$$

Considering Equation 6 one can see how the weights $w_{j,t}$, belonging to the jth model, are appointed to their produced forecast $\hat{r}_{j,t}$ on time t. When we sum the weighted forecasts to p, which is the number of models we take into consideration, we get a time series of forecast combinations $\{\hat{r}_{t,comb}\}_{t=M+1}^{T}$. In our analysis we use an equal-weighted combination, hence all weights are set to $w_{j,t} = 1/p$. We utilize the equal-weighted forecast combination for the macroeconomic variables resulting in the model EW-M and also for the technical variables resulting in the model EW-T. The choice of using an equal-weighted combination is based on the works of Rapach, Strauss and Zhou (2010), who state that it performs quite well for economic and financial forecasting.

3.4 Combination Elastic Net

In Section 3.3 we described a technique called forecast combination where we combine the various forecasts made by the models. As stated by Rapach and Zhou (2020), this technique possibly neglects 'the substantive relevant information in the predictor variables'. Therefore we introduce the combination elastic net method. A variation on the forecast combination technique where we first need to pinpoint which of the models we should include in the forecast combination. We use the elastic net method in the process to find out which of the models we need to incorporate. First, we need to divide the sample into three parts, namely the initial in-sample period with size t_1, the initial holdout out-of-sample period denoted with $(s = t_1 + 1, ..., t)$ and the initial out-of-sample period from $t + 1$ until T.

The second step is to compute forecasts made from univariate OLS regressions with a recursive window over the holdout out-of-sample period, utilizing Equation 2 and Equation 5.

Then we estimate an expression proposed by Granger and Ramanathan (1984) with the elastic net method, again over the holdout out-of-sample period. Which is basically regressing the commodity returns on the forecasts of all of the models and a constant, resulting in the following equation. We use the forecasts of the previous step.

$$r_s = \eta + \sum_{j=1}^{p} \theta_j \hat{r}_{j,s} + \epsilon_s \tag{7}$$

Where η represents the constant, ϵ_t the error term on time t and the θ_j is the coefficient belonging to its corresponding forecasted return $\hat{r}_{j,t}$ on time t. In this step, the elastic net method will provide the θ_j coefficients. After that, we ensemble a set Θ containing all the univariate regressions that are given a non-negative coefficient ($\theta_j \geq 0$) by the elastic net estimation stated in Equation 7.

Lastly, we produce the combination elastic net forecasts by using the following formula.

$$\hat{r}_{t+1}^{C-ENet} = \frac{1}{|\Theta|} \sum_{j \in \Theta} \hat{r}_{j,t} \tag{8}$$

In this equation, we have $\hat{r}_{j,t}$ as the forecasted returns made with the jth model on time t. However, we use only use the univariate models that are included in Θ. So we basically use the models from the Θ set in an equal-weighted forecast combination.

3.5 Forecast Evaluation

3.5.1 Out-of-sample R^2

To evaluate the various forecasting models we generate, one of the forecast evaluation methods we use is percentage out-of-sample R^2. For the calculation of this value, we need the mean squared prediction error (MSPE) of the forecasting model and the benchmark model. The benchmark model that is originally chosen by Wang et al. (2020) is the historical mean model with $\bar{r}_{t+1} = \frac{1}{t}\sum_{j=1}^{t} r_j$, due to similarity in research we also use this model as the benchmark model. To compute the MSPE of a model we use the following formula.

$$MSPE = \frac{1}{T-M}\sum_{t=M+1}^{T}(\hat{r}_t - r_t)^2 \tag{9}$$

In Equation 9 we have $T - M$ observations representing the out-of-sample period, $\hat{r}_t$ as the forecasted return by the model and r_t as the actual return value, both on time t. This way we create the mean squared prediction error for the models we analyze. After that, we use $MSPE_{model}$ and $MSPE_{bench}$ to represent respectively the MSPE of one of the forecast models and the MSPE of the benchmark model. With the following expression, we calculate the percentage out-of-sample R^2.

$$R^2_{OoS} = 100 \times \left(1 - \frac{MSPE_{model}}{MSPE_{bench}}\right) \tag{10}$$

Looking at Equation 10, we can see that a positive value of percentage out-of-sample R^2 (R^2_{OoS}) means that the model outperformed the benchmark model in the given out-of-sample period.

3.5.2 Clark and West Method

To assess if a predictive model is significantly better than the benchmark model we use a method proposed by Clark and West (2007). This method uses an alternative for loss differences to test the null hypothesis that the MSPE of the benchmark model is less than or equal to the MSPE of the model ($H_0 : MSPE_{bench} \leq MSPE_{model}$). The reason that we choose the Clark and West method is that we want to test nested models[10]. Therefore the more intuitive test proposed by Diebold and Shin (2019) with loss differences is not suitable. We first need to calculate the alternative metrics for the loss differences, which are computed as follows.

$$L_t = (r_t - \bar{r}_t)^2 - (r_t - \hat{r}_t)^2 + (\bar{r}_t - \hat{r}_t)^2 \tag{11}$$

Regarding Equation 11 one can see the variables r_t , $\hat{r}_t$ and $\bar{r}_t$. They respectively represent

[10]The definition of a nested model is a model that contains a subset of the variables from the other model we want to test.

the actual returns, the forecasted returns from a chosen model and the returns forecasted by the benchmark model. The variables on time t are used to generate a time series $\{L_t\}_{M+1}^{T}$. Once this time series is created for a certain model we can regress $\{L_t\}_{M+1}^{T}$ on a constant. The t-statistic of the estimated constant portrays the CW statistic, as described by Clark and West (2007). After that, we test the null hypothesis ($H_0 : MSPE_{bench} \leq MSPE_{model}$) at significance levels 10%, 5% and 1%. The p-value is computed for the upper-tail test with the Student t-distribution (Wang et al., 2020).

3.5.3 Model Confidence Set

Another way to evaluate the forecasts is the use of the Model Confidence Set (MCS) procedure described by Hansen, Lunde and Nason (2011). This method uses loss functions to determine whether or not a certain model can fit in a so-called Superior Set Model (SSM). The benefit of using MCS as forecast evaluation is the fact that the approach accounts for the limitations of the provided information. MCS does so by selecting a set of models, unlike other model criteria that single out one superior model. At the start, we have a set of forecasts that correspond to the models we want to analyze. Then the MCS process eliminates the models that are significantly inferior to the other models contained in the set at a given confidence level $1 - \alpha$. After the eliminations, we get the Superior Set Model, which is the set of models that are not significantly inferior to each other. The confidence level $1-\alpha$ implies that the created SSM contains the real superior set with probability $(1 - \alpha)$%. A higher α, therefore, means that the MCS process is more strict in terms of elimination, usually more models will be eliminated.

In our research, the squared forecast errors made by the models are representing the loss functions for the MCS process. In our analysis, we choose to use the test statistic $T_{max,M}$, this way we take the maximum of the t-statistic calculated with the loss functions. The full computation is described in the works of Hansen et al. (2011). Additionally, we chose to perform 5000 bootstraps in the MCS procedure. Lastly, the chosen confidence levels that are used to create the SSM are 90% and 75%, where the sets will be indicated respectively with $\hat{M}_{0.90}$ and $\hat{M}_{0.75}$.

On top of that, the MCS procedure provides p-values that can be interpreted as the probability that a model is a member of the real superior set of models. When looking at the works of Hansen, Lunde and Nason (2003), the p-value for model i denoted as $\hat{p}_i$ is as follows.

$$i \in \hat{M}_\alpha \quad for\ \alpha \leq \hat{p}_i \qquad i \notin \hat{M}_\alpha \quad for\ \alpha > \hat{p}_i \tag{12}$$

Essentially, Equation 12 implies that the p-value of model i is such that the model is still contained in the SSM denoted by $\hat{M}_\alpha$. If the p-value drops below α the model is eliminated. Once a SSM is generated with a certain confidence level, the p-values of the models are stated as well. However, the p-values differ as the α of the MCS procedure changes. That is why we only mention the p-values that are generated for the superior set models with confidence levels of 90% and 75%.

4 Results

Due to the fact that this research is inspired by the paper of Wang et al. (2020), we first forecasted an out-of-sample period from January 1992 until December 2017 as replication. For full details on the forecasting results of this time frame, one can take a look at Appendix E.

4.1 General Forecasting Results

We take a look at the forecasting results of one month ahead forecasts made with a recursive window, while we use ordinary least squares as estimation method. We use the full sample, namely a sample from January 1982 until December 2022. The period from January 1982 until December 1991 is used as in-sample period. We extend the out-of-sample period to a time interval from January 1992 until December 2022. The five years that are added to the sample contain some interesting periods like the covid-pandemic and the energy crisis we currently experience. The numbers in bold in Table 1 represent a positive R^2_{OoS}. When we test the null hypothesis described in Section 3.5.2 ($H_0 : MSPE_{bench} \leq MSPE_{model}$) at significance levels of 10%, 5% and 1%, we indicate rejections of the null hypothesis respectively with one, two or three asterisks in Table 1. For instance, the null hypothesis is rejected at a 10% significance level for the MV model when forecasting the precious metals index. Although, when the MV model is used to forecast non-energy returns the null hypothesis is rejected at a 1% significance level. When looking at the CW statistic one can see that almost all the technical indicator models significantly outperform the benchmark model when forecasting commodity returns. Only the FR model does not notably outperform the benchmark model. Looking solely at the R^2_{OoS} values one can say that the OSLT and SR models perform best regarding all commodities. The models based on the variables DP, TBL and LTY do not exceed the predictive performance of the benchmark model.

The results for the combination of forecasts made by both macroeconomic and technical models are denoted in Table 1 by EW-M and EW-T respectively. The EW-M model provides for all commodities positive R^2_{OoS} values. However, for three out of eight commodities we do not reject the null hypothesis at a 10% confidence level. For the forecast combination of the technical variables, we see that for seven out of eight commodities, the forecasts are significantly better at a 1% confidence level than the ones made by the benchmark model. Merely the forecasts for the precious metals returns cause the null hypothesis to be rejected at a 5% level.

To take up the issue of whether technical variables can still forecast well during volatile periods, we changed the out-of-sample forecast period to January 2018 until December 2022. The computed R^2_{OoS} values for the technical models are included in Appendix F. We can see that some of the values increased in comparison with the full sample period, indicating that the models perform relatively better in volatile periods than the benchmark. Only a few values dropped to a negative R^2_{OoS}, like some of the technical models forecasting the beverage index. Based on the R^2_{OoS} values, one can say that the technical models are still usable in volatile periods.

	Energy	Non-energy	Agriculture	Beverage	Food	Raw materials	Metals & minerals	Precious metals
DP	-0.252	-0.120	-0.489	-1.261	-0.519	-1.307	-0.269	**0.361**
TBL	-0.275	-1.311	-0.857	-0.027	-1.427	-1.268	-1.535	-0.527
LTY	-0.310	-0.448	-0.399	-0.958	-0.344	-2.018	-0.694	-0.095
INFL	-0.521	-0.925	-1.097	-1.441	-0.970	-0.909	-0.674	**1.822****
IP	-3.094	-4.788	-2.969	**0.035***	-2.985	-1.455	-4.102	-0.726
RET	-0.825	**4.311*****	**1.224****	-0.131	**0.976****	**0.031**	**7.619*****	-0.034
KI	**0.411**	**1.672****	**2.166*****	**0.346***	**1.330*****	**0.802****	-0.870	-4.106
EX	**1.910*****	**6.059*****	**3.780*****	-0.190	**2.694*****	**0.738****	**4.697*****	**1.567*****
EW-M	**0.229**	**2.600*****	**1.893*****	**0.265**	**1.579*****	**0.047**	**1.862*****	**1.005****
MOM	**1.948*****	**5.870*****	**5.734*****	**3.167*****	**4.346*****	**6.976*****	**4.239*****	**1.587*****
FR	**2.517*****	**5.257*****	**3.446*****	**2.858*****	**3.349*****	**3.646*****	**5.580*****	**0.646***
MV	**0.968*****	**5.099*****	**3.734*****	**2.150*****	**1.826*****	**4.836*****	**3.169*****	**0.761***
OSLT	**1.799*****	**6.645*****	**6.448*****	**3.522*****	**4.664*****	**7.212*****	**4.272*****	**1.831*****
SR	**2.353*****	**7.190*****	**6.007*****	**3.790*****	**5.780*****	**8.525*****	**5.766*****	**0.996****
EW-T	**1.982*****	**6.368*****	**5.337*****	**3.301*****	**4.206*****	**6.638*****	**4.821*****	**1.241****

Table 1: Percentage Out-of-sample R^2 for the models estimated with ordinary least squares, while evaluating the period from January 1992 until December 2022.

The Model Confidence Set procedure helps choosing a set of models that have equal predictive ability. We analyze the forecasts made for the out-of-sample period from January 1992 until December 2022. As we performed the MCS process for the confidence levels 90% and 75%, the p-values corresponding to the models are generated and registered in Table 2. For example, the SSM that is created for the forecasting of the energy index with a confidence level of 90% contains all forecast models. Some have a p-value of one, meaning they are more likely to stay in the SSM when the process gets more strict. The lowest p-value regarding the energy index is generated for the TBL model, which implies the TBL model is less likely to be in the real superior set of models. The crosses in Table 2 indicate the elimination of a model from the SSM. So the SSM created for the forecast models of the raw materials index contains only the models SR, OSLT and MOM for both confidence levels. All the other models are significantly inferior and thus do not belong to the SSM.

	Energy		Non-energy		Agriculture		Beverage		Food		Raw materials		Metals & minerals		Precious metals	
	$\hat{M}_{0.90}$	$\hat{M}_{0.75}$	$\hat{M}_{0.90}$	$\hat{M}_{0.75}$	$\hat{M}_{0.90}$	$\hat{M}_{0.75}$	$\hat{M}_{0.90}$	$\hat{M}_{0.75}$	$\hat{M}_{0.90}$	$\hat{M}_{0.75}$	$\hat{M}_{0.90}$	$\hat{M}_{0.75}$	$\hat{M}_{0.90}$	$\hat{M}_{0.75}$	$\hat{M}_{0.90}$	$\hat{M}_{0.75}$
DP	0.869	0.854	×	×	×	×	0.119	×	0.147	×	×	×	0.147	×	1.000	0.979
TBL	0.619	0.596	0.328	0.317	0.292	0.266	0.610	×	0.236	×	×	×	0.291	0.265	0.997	0.965
LTY	0.672	0.658	0.199	×	0.132	×	×	×	0.629	0.352	×	×	0.190	×	1.000	0.990
INFL	0.872	0.864	×	×	×	×	×	×	0.196	×	×	×	×	×	1.000	1.000
IP	0.762	0.741	0.659	0.581	0.479	0.406	0.647	0.258	0.654	0.463	×	×	0.692	0.582	0.537	0.301
RET	0.945	0.938	1.000	1.000	0.749	0.607	0.127	×	0.992	0.845	×	×	1.000	1.000	0.971	0.724
KI	1.000	1.000	0.985	0.961	1.000	1.000	0.822	0.340	1.000	0.999	×	×	0.310	0.267	0.106	×
EX	1.000	1.000	1.000	1.000	1.000	1.000	0.814	0.330	1.000	1.000	×	×	1.000	1.000	1.000	1.000
EW-M	0.706	0.712	0.530	0.389	0.419	0.266	0.271	×	1.000	0.699	×	×	0.688	0.375	1.000	1.000
MOM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.266	0.287	1.000	1.000	1.000	1.000
FR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	×	×	1.000	1.000	1.000	1.000
MV	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.882	×	×	1.000	1.000	1.000	1.000
OSLT	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.517	0.517	1.000	1.000	1.000	1.000
SR	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
EW-T	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	×	×	1.000	1.000	1.000	1.000

Table 2: P-values per forecast model generated with the MCS procedure for the confidence levels 90% and 75%, while evaluating the period from January 1992 until December 2022.

Where in Table 1 one can see a clear difference in forecasting performance when comparing the macroeconomic model and technical models, we see that the MCS procedure provides a more nuanced result. For seven out of eight commodities, almost all of the macroeconomic models are included in the SSM, meaning they are not significantly inferior. Only for the raw materials index the SSM consists of three technical models. Nevertheless, the p-values of the macroeconomic models are in general lower than the ones of the technical models. Still, one can say that the statement of Wang et al. (2020), that technical indicators are stronger predictors than economic indicators, is not fully backed by the MCS procedure.

4.2 LASSO

As described in Section 2.3, the technical variables consist of multiple technical indicators. To understand which of these technical indicators are relatively more relevant we put these indicators separately in the LASSO method. We forecasted the same out-of-sample period from January 1992 until December 2022 and marked the R^2_{OoS} values in Table 3, stated in Section 4.3. The same benchmark model is retained for the analysis. Once again the positive values are in bold and the asterisks represent the rejections of the null hypothesis: $H_0 : MSPE_{bench} \leq MSPE_{model}$. However, for the support-resistance rule, we choose to omit the last two technical indicators of both the buying signal and the selling signal[11]. These indicators were too similar at the start of the recursive window period which caused problems when running the LASSO method.

Due to the length of the computation time when running the code for the LASSO method, we chose to have a recursive window that expands every year instead of every month. So where before we estimated a new regression every month, we now only execute the estimation every year to make forecasts for the upcoming year.

In Table 3, MOM-L represents the model that is generated with the LASSO method considering all the technical indicators corresponding to the momentum rule. As an illustration, the non-energy index forecasts created by the OSTL-L model have a 7.961% out-of-sample R^2 when compared with the benchmark model. To assess the forecasting performance of the LASSO models one can compare the results of Table 1 with the results of Table 3[12]. Interesting to see is that almost all R^2_{OoS} values increase when we regress with LASSO before the forecasting instead of OLS. However, not all values improve. For example, the forecasting performance of the FR-L model when forecasting the raw materials index declines to a negative R^2_{OoS}. Once again, we used an equal-weighted forecast combination for all technical LASSO models to create the model EWT-L.

To discuss the significance of the indicators, we observe the number of times the LASSO method sets the coefficient of an indicator to zero. The big amount of indicators makes it hard to state the exact numbers but we do see a pattern in the variable selection. For the MOM models, we see that the indicators with the bigger lags are mostly set to zero. The one lag momentum indicator almost never has the value zero. The variable selection for the buying and selling indicators based on the filtering rule shows two different patterns. One pattern is where only the variables with low lag are selected for both buying and selling signals and the other pattern is where the indicators with medium length lag are preferred. For the second option, the indicators with $\eta = 10$ are almost never selected. These two patterns differ per commodity index. The indicators of the MV model are often set to zero, except for two which are almost never set to zero. The first is the MV indicator with $s = 1$ and $l = 3$ and the second is the MV indicator with $s = 9$ and $l = 12$[13]. However, the variable selection for the MV indicators differs a lot per commodity index. Considering the indicators of the OSLT rule, we see that usually the indicators with small k are selected, especially for the case where $\eta = 10$. The variable selection for the SR indicators seems to be the most random. The only pattern one could observe is that mostly more sell indicators with larger η are set to zero.

In the research, we also tried to use the LASSO method to generate a model with both macroeconomic and technical variables. However, the LASSO regression of the model with all

[11]Note that the missing indicators that are included in Appendix C are still also omitted.

[12]Note that the evaluation period is the same for both analyses and all the models are compared to the same benchmark model to compute the R^2_{OoS} values.

[13]See Appendix A for the calculation of the technical indicators.

variables included was not valid. A feasible explanation for this could be that the indicators belonging to the FR and SR are too similar at the beginning of the sample. This caused problems for the LASSO method because it could not select which variables were significant for the model. Moreover, looking at the results of the MACRO-L model we can see that incorporating the macroeconomic variables in a LASSO model does not enhance forecasting performance. Thus a combination of both macroeconomic and technical variables may also be inferior to a model with only technical indicators.

4.3 Elastic Net

With the same setting as described in Section 4.2, we performed the elastic net method to forecast commodity prices. Table 3 shows the R^2_{OoS} values of the elastic net models indicated with names like MOM-E and SR-E. Again the recursive window is updated every year due to the computation time. The models that are generated with the elastic net method show similar R^2_{OoS} values as the models that are estimated with LASSO. Some perform better, like the FR-E model for raw materials compared to the FR-L model and some perform worse like the MV-E model for precious metals compared with the MV-L model. Again the MACRO-E model has a relatively bad forecasting performance. The EWT-E model is the equal-weighted forecast combination model for all technical elastic net models.

While calculating the CW test statistics to discover if we should reject the null hypothesis ($H_0 : MSPE_{bench} \leq MSPE_{model}$) for the elastic net models, there was an interesting difference with the LASSO models. The CW statistic was relatively higher for almost all direct comparisons. We will use the MCS procedure to test if these elastic net models significantly outperform the LASSO models.

For the discussion on variable selection via elastic net, we refer to Section 4.2. There, we describe the patterns of the variable selection by LASSO that can be observed. The selection of variables via elastic net shows similar patterns. However, the variables are less likely to be set to exactly zero. The coefficients are shrunk to very small values. Also, for the MV indicators, we see the indicator with values $s = 9$ and $l = 12$ being shrunk to zero more often.

In Section 3.2.2 we stated that the elastic net method was a viable method to solve the shortcoming of the LASSO method, occurring when a group of variables is highly correlated. The technical indicators are highly correlated but we do not see improvement in terms of percentage out-of-sample R^2 for many models when applying the elastic net method instead of the LASSO method.

We used the MACRO model as a multivariate linear model that includes all macroeconomic variables. If we compare the R^2_{OoS} values of MACRO, MACRO-L and MACRO-E, one can conclude that regularization methods decrease the forecasting performance for models including all the macroeconomic variables.

	Energy	Non-energy	Agriculture	Beverage	Food	Raw materials	Metals & minerals	Precious metals
MOM-L	**2.250*** **	**7.632*** **	**9.562*** **	**2.449*** **	**6.453*** **	**9.766*** **	**5.050*** **	**2.283*** **
FR-L	**2.704** **	**0.929** **	**0.746** **	**4.410*** **	**2.838*** **	-2.320	**4.262*** **	**0.021**
MV-L	**1.805*** **	**8.418*** **	**7.767*** **	**3.404*** **	**5.083*** **	**9.355*** **	**4.215*** **	**0.474* **
OSLT-L	**2.467*** **	**7.961*** **	**10.168*** **	**4.198*** **	**7.523*** **	**9.319*** **	**6.072*** **	**1.435*** **
SR-L	**2.465*** **	**13.289*** **	**4.827*** **	**4.568*** **	**10.071*** **	**12.422*** **	**6.094*** **	**1.347** **
EWT-L	**3.288*** **	**9.163*** **	**9.891*** **	**4.851*** **	**7.643*** **	**11.875*** **	**5.747*** **	**1.834*** **
MOM-E	**2.772*** **	**7.626*** **	**9.823*** **	**2.402*** **	**6.418*** **	**9.779*** **	**4.932*** **	**2.092*** **
FR-E	**0.562** **	**0.479*** **	**0.546** **	**5.219*** **	**1.790*** **	**3.973*** **	**2.839*** **	**0.535**
MV-E	**1.822*** **	**8.961*** **	**7.443*** **	**3.064*** **	**5.051*** **	**9.376*** **	**3.453*** **	-0.435
OSLT-E	**2.287*** **	**8.107*** **	**10.059*** **	**3.066*** **	**8.018*** **	**9.205*** **	**4.841*** **	**2.223*** **
SR-E	**1.764*** **	**12.060*** **	-0.275	**4.091*** **	**10.554*** **	**15.170*** **	**6.409*** **	**1.853*** **
EWT-E	**3.925*** **	**8.820*** **	**9.140*** **	**4.424*** **	**7.819*** **	**11.858*** **	**5.924*** **	**1.735*** **
MACRO	-4.293	**3.003*** **	**0.223*** **	-4.096	-1.772	-6.247	**4.094*** **	-3.884
MACRO-L	-9.869	-3.252	-5.286	-4.763	-8.690	-13.020	-2.087	-4.155
MACRO-E	-9.907	-1.975	-13.672	-5.939	-7.514	-11.831	-3.189	-6.946

Table 3: Percentage Out-of-sample R^2 for the LASSO technical models, elastic net technical models and the standard multivariate macroeconomic model, while evaluating the period from January 1992 until December 2022.

4.4 Combination Elastic Net

To assess if the forecasting performance increases when the macroeconomic and technical variables are put together in one model we used the combination elastic net method. The initial in-sample period starts in January 1982 and goes until December 1991, the initial holdout out-of-sample period is from January 1992 until December 2001 and the initial out-of-sample period is from January 2002 until December 2022. Therefore, the forecast evaluation is performed over the last time frame. Table 4 portrays the R^2_{OoS} values when comparing the models with the benchmark model. For comparative purposes we included the C-Net model, the EW-M model and the EW-T model because they are all forecast combinations, utilizing the estimation method OLS. Also, we incorporated one of the best-performing regularization methods and one of the best-performing standard methods in Table 4 for comparison, namely SR-E and SR. An overview of which variables are included in the combination elastic net (C-Net) model's Θ sets is included in Appendix G.

One can see in Table 4 that the C-Net R^2_{OoS} values are positive for all commodities and therefore outperform the benchmark model. Also the null hypothesis ($H_0 : MSPE_{bench} \leq MSPE_{model}$) is rejected on a 1% significance level for all commodities. Nevertheless, we can see that for the same time frame, the other models all perform better for almost all commodities. Only the EW-M model performs relatively worse than the C-Net regarding this forecasting period. Note that the R^2_{OoS} values for the models differ if the evaluation time frame changes.

	Energy	Non-energy	Agriculture	Beverage	Food	Raw materials	Metals & minerals	Precious metals
C-Net	**1.633*****	**4.525*****	**2.962*****	**3.227*****	**3.722*****	**4.933*****	**4.657*****	**1.803*****
EW-M	**0.350**	**2.981*****	**2.140*****	**1.078*****	**1.885*****	**0.523**	**2.419*****	**1.347****
EW-T	**2.124*****	**6.775*****	**5.892*****	**3.335*****	**5.106*****	**7.138*****	**5.051*****	**1.705*****
SR-E	**4.032*****	**15.433*****	**15.316*****	**2.424*****	**12.324*****	**12.844*****	**6.943*****	**2.842*****
SR	**2.701*****	**8.044*****	**7.026*****	**3.535*****	**7.204*****	**7.961*****	**6.307*****	**1.410****

Table 4: Percentage out-of-sample R^2 for the combination elastic net model, the equal-weighted forecast combinations, the best-performing regularization model and the best-performing standard model, while evaluating the period from January 2002 until December 2022.

The forecast results of the EW-T model outperform the C-Net for all commodities except the precious metals index. One can say the elastic net combination of the macroeconomic variables and technical variables does not increase forecasting performance relative to the other methods. To check the latter we also executed the MCS procedure including the LASSO models, elastic net models and the combination elastic net model. Once again we evaluate the forecasting performance in the time period from January 2002 to December 2022.

We choose to include only the LASSO models, elastic net models and the combination elastic net model. The reason for this is that when we performed the MCS procedure for all the models we generated, we see a pattern. The models with all macroeconomic variables (MACRO, MACRO-L, MACRO-E) were usually eliminated first. Then the worst performing linear macroeconomic models, in terms of R^2_{OoS} values, were eliminated. The LASSO models and elastic net models are not eliminated for all commodities. Combined with the R^2_{OoS} value results, showing that the regularization method models outperform the standard models, we decided that considering all models is redundant. Thus, to really asses the difference in forecasting performance between the LASSO models, elastic net models and the combination elastic net model we only included these in the MCS procedure.

The results of the MCS procedure for the regularization methods can be seen in Table 5. What is remarkable to see is that some of the models that provide relatively high R^2_{OoS} values are not always included in the SSM. For instance, the SR-E model has a 15.433% out-of-sample R^2 for the same evaluation period but is not incorporated in the SSM. Some have a SSM containing almost all regularization models, so the models are not significantly different in forecasting performance. Nonetheless, the SSM created for the precious metals index contains only the models OSLT-L and SR-L, which clearly states their superiority.

We want to judge if the elastic net combination of both macroeconomic and technical variables beats the forecasting performance of the technical models with selected indicators[14]. In Table 5 we can see that the C-Net model is included in the SSM for five out of eight commodities. Consequently, the forecasting performance of the C-Net model is not overall better than the

[14]The most important indicators are chosen via the LASSO or elastic net method, as described in the corresponding sections.

models using regularization methods. However, none of the models is included in the SSM for every commodity, which indicates there is no absolute winner among the regularization models. The MOM-E model is closest to being the best with the inclusion in the SSM for seven out of eight models, plus relatively high R^2_{OoS} values in Table 3[15].

To compare the overall performance of the LASSO models to the elastic net models we can see that for the commodities energy, non-energy, agriculture and beverage the LASSO models are less included than the elastic net models. On the other hand, for the raw materials, metals & minerals and food indices all the LASSO models are included in the SSM. All the more, the SSM made for the precious metals index only includes two LASSO models. Once again when comparing the forecasting performance through the MCS procedure for the LASSO and elastic net models, there is no obvious victor.

	Energy		Non-energy		Agriculture		Beverage		Food		Raw materials		Metals & minerals		Precious metals	
	$\hat{M}_{0.90}$	$\hat{M}_{0.75}$	$\hat{M}_{0.90}$	$\hat{M}_{0.75}$	$\hat{M}_{0.90}$	$\hat{M}_{0.75}$	$\hat{M}_{0.90}$	$\hat{M}_{0.75}$	$\hat{M}_{0.90}$	$\hat{M}_{0.75}$	$\hat{M}_{0.90}$	$\hat{M}_{0.75}$	$\hat{M}_{0.90}$	$\hat{M}_{0.75}$	$\hat{M}_{0.90}$	$\hat{M}_{0.75}$
MOM-L	0.919	0.921	×	×	0.928	0.922	×	×	1	1	1	1	0.979	0.979	×	×
FR-L	1	1	×	×	×	×	×	×	1	1	0.352	0.318	0.906	0.907	×	×
MV-L	×	×	×	×	×	×	×	×	1	1	1	1	0.306	0.290	×	×
OSLT-L	×	×	×	×	1	1	×	×	1	1	1	1	1	1	0.797	0.790
SR-L	×	×	×	×	1	1	×	×	1	1	1	1	1	1	1	1
MOM-E	1	1	1	1	1	1	1	1	0.652	0.625	1	1	1	1	×	×
FR-E	1	1	0.479	0.491	×	×	0.822	0.822	1	1	1	1	0.673	0.658	×	×
MV-E	0.4848	0.4858	1	1	0.491	0.499	0.370	0.370	0.632	0.640	1	1	0.366	0.3432	×	×
OSLT-E	0.992	0.994	1	1	0.998	0.998	×	×	1	1	0.651	0.644	1	1	×	×
SR-E	1	1	×	×	1	1	×	×	0.996	0.997	1	1	1	1	×	×
C-Net	0.5764	0.563	0.348	0.330	×	×	1	1	1	1	0.995	0.995	×	×	×	×

Table 5: P-values per regularization method forecast model generated with the MCS procedure for the confidence levels 90% and 75%, while evaluating the period from January 2002 until December 2022.

5 Conclusion

In the literature on forecasting commodity prices, there has been written a lot about models containing macroeconomic variables. Less of this literature considers the usage of technical indicators to forecast commodity prices. That is why our research contributes to the literature by analysing models with technical variables based on several indicators. Moreover, we utilize regularization methods to concentrate on the essential technical indicators to improve general forecasting performance. Thus taking another perspective in the literature on forecasting com-

[15]Note that Table 3 provides R^2_{OoS} values for another evaluation period than is used to construct Table 5.

modity prices.

The issue was stated as follows: How do regularization methods impact the performance of forecasting commodity prices when using both conventional variables and technical indicators? While evaluating the various models we came across some interesting findings. Even when the sample is extended until December 2022 the technical models all outperform the benchmark model by means of percentage out-of-sample R^2. Also when using the Clark and West statistic to test the null hypothesis ($H_0 : MSPE_{bench} \leq MSPE_{model}$) we see a rejection at a confidence level of 1% for almost all technical models and some macroeconomic models. The best forecasting performance in terms of percentage out-of-sample R^2 belongs to the forecast of the raw materials index by the SR model.

The model confidence set procedure is a more robust evaluation method that can tell when certain models are inferior to one another. For the confidence levels 90% and 75%, it showed that for many commodities the models were not significantly inferior to each other. However, the superior set model for the raw materials index contained only the MOM, OSLT and SR models.

The selection of the essential indicators through implementing the LASSO and elastic net methods on the technical variables resulted in higher percentage out-of-sample R^2 values. However, not all models profited from the variable selection. Especially the SR and FR rules had many similar indicators which caused problems for the regularization methods. Nevertheless, the overall forecasting performance increased. To compare which regularization method provided the most benefit to the forecasting performance, we again used the MCS procedure. Observing the results, we can not declare a clear winner between the LASSO models and the elastic net models.

The C-Net model outperformed the benchmark model by means of percentage out-of-sample R^2. Also, for all commodities, the null hypothesis was rejected. Still, when compared with other LASSO and elastic net models through the MCS procedure, the C-Net model was not included in all superior set models. Meaning, the elastic net combination of both macroeconomic and technical variables did not outperform all the regularization models.

In conclusion, the impact of regularization methods on the performance of forecasting commodity prices is positive. By selecting the essential indicators the general forecasting performance is enhanced. Despite that, the combination elastic net model did outperform the benchmark model but not all the univariate technical models. Therefore one can say the elastic net combination of both macroeconomic and technical variables is not the optimal method.

Suggestions for further research are based on both computation time problems and the expansion of methods with machine learning. The computation time of running some of the regularization methods was relatively long. Therefore, a suggestion for further research is to use the monthly recursive window and the usage of a bigger data set to increase the size of the training set. The other suggestion for further research is to incorporate more machine learning methods into the issue of forecasting commodity prices.

References

Akram, Q. F. (2009). Commodity prices, interest rates and the dollar. *Energy economics*, *31*(6), 838–851.

Borensztein, E. & Reinhart, C. M. (1994). The macroeconomic determinants of commodity prices. *Staff Papers*, *41*(2), 236–261.

Chen, Y.-C., Rogoff, K. S. & Rossi, B. (2010). Can exchange rates forecast commodity prices? *The Quarterly Journal of Economics*, *125*(3), 1145–1194.

Clark, T. E. & West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of econometrics*, *138*(1), 291–311.

Dickey, D. A. & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, *74*(366a), 427–431.

Diebold, F. X. & Shin, M. (2019). Machine learning for regularized survey forecast combination: Partially-egalitarian lasso and its derivatives. *International Journal of Forecasting*, *35*(4), 1679–1691.

Fuertes, A.-M., Miffre, J. & Rallis, G. (2010). Tactical allocation in commodity futures markets: Combining momentum and term structure signals. *Journal of Banking & Finance*, *34*(10), 2530–2548.

Gargano, A. & Timmermann, A. (2014). Forecasting commodity price indexes using macroeconomic and financial predictors. *International Journal of Forecasting*, *30*(3), 825–843.

Granger, C. W. & Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of forecasting*, *3*(2), 197–204.

Groen, J. J. & Pesenti, P. A. (2011). Commodity prices, commodity currencies, and global economic developments. In *Commodity prices and markets* (pp. 15–42). University of Chicago Press.

Hansen, P. R., Lunde, A. & Nason, J. M. (2003). Choosing the best volatility models: the model confidence set approach. *Oxford Bulletin of Economics and Statistics*, *65*, 839–861.

Hansen, P. R., Lunde, A. & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*(2), 453–497.

Heij, C., de Boer, P., van Dijk, H. K., Kloek, T. & Franses, P. H. (2004). *Econometric Methods With Applications in Business and Economics.* Athenaeum Uitgeverij.

Hoerl, A. E. & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67.

Levy, R. A. (1967). Relative strength as a criterion for investment selection. *The Journal of finance*, *22*(4), 595–610.

Moskowitz, T. J., Ooi, Y. H. & Pedersen, L. H. (2012). Time series momentum. *Journal of financial economics*, *104*(2), 228–250.

Parzinger, M., Hanfstaengl, L., Sigg, F., Spindler, U., Wellisch, U. & Wirnsberger, M. (2022). Comparison of different training data sets from simulation and experimental measurement with artificial users for occupancy detection—using machine learning methods random forest and lasso. *Building and Environment*, *223*, 109313.

Rapach, D. E., Strauss, J. K. & Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies*, *23*(2), 821–862.

Rapach, D. E. & Zhou, G. (2020). Time-series and cross-sectional stock return forecasting: New machine learning methods. *Machine learning for asset management: New developments and financial applications*, 1–33.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.

Wang, Y., Liu, L. & Wu, C. (2020). Forecasting commodity prices out-of-sample: Can technical indicators help? *International Journal of Forecasting*, *36*(2), 666–683.

Zou, H. & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, *67*(2), 301–320.

A Technical Indicators

The first technical variable MOM is generated with five indicators corresponding with the momentum rule. This trading rule looks at the trend of the price and then provides a signal via the following.

$$S_{t,MOM} = \begin{cases} 1, & \text{if } P_t \geq P_{t-k} \\ 0, & \text{if } P_t < P_{t-k} \end{cases} \tag{13}$$

Where in Equation 13, $S_{t,MOM}$ represents the signal given on time t and P_t the price on time t. The variable P_{t-k} represents the lagged price with options $k = 1, 3, 6, 9, 12$ which gives us five momentum rule indicators.

The second technical variable FR is generated with twenty indicators corresponding to the filtering rule. This trading rule provides both buying and selling signals. The buying (selling) signal is provided when the price P_t has risen (fallen) above (below) its most recent low (high) more than a specified percentage (Wang et al., 2020). The filtering rule indicators can be written as follows.

$$S^{buy}_{t,FR} = \begin{cases} 1, & \text{if } P_t \geq \left(1 + \frac{\eta}{100}\right) * min(P_{t-1}, P_{t-2}, ..., P_{t-k}) \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

$$S^{sell}_{t,FR} = \begin{cases} 1, & \text{if } P_t \leq \left(1 - \frac{\eta}{100}\right) * max(P_{t-1}, P_{t-2}, ..., P_{t-k}) \\ 0, & \text{otherwise} \end{cases} \tag{15}$$

Where in Equation 14, the $S^{buy}_{t,FR}$ represents the buying signal and in Equation 15, $S^{sell}_{t,FR}$

represents the selling signal. In both equations, the possible values for k and η are as follows. $k = 1, 3, 6, 9, 12$ and $\eta = 5, 10$ resulting in ten buying and ten selling indicators.

The third technical variable MV is generated with ten indicators corresponding to the moving average rule. The rule provides a signal when the short-term moving average ($MV_{s,t}$) of prices is higher than the long-term moving average ($MV_{l,t}$). The short and long-term moving averages are computed as follows.

$$MV_{j,t} = \left(\frac{1}{j}\right) \sum_{i=0}^{j-1} P_{t-i} \quad , \quad j = s, l \tag{16}$$

Where s and l in Equation 16 respectively represent the short and long-term moving average. This gives as the following rule for the moving average signals ($S_{t,MV}$).

$$S_{t,MV} = \begin{cases} 1, & \text{if } MV_{s,t} \geq MV_{l,t} \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

The possible values for s and l are $s, l = 1, 3, 6, 9, 12$ but we always need the condition ($s < l$). Therefore, this rule produces ten moving average indicators.

The fourth technical variable OSLT is generated with twenty indicators corresponding to the oscillator trading rule. This trading rule provides a buying or selling signal based on the relative strength indicator (Levy, 1967). Which can be defined as follows.

$$RSI(m) = 100 \left(\frac{U_t(m)}{U_t(m) + D_t(m)}\right) \tag{18}$$

In Equation 18 the variables $U_t(m)$ and $D_t(m)$ represent respectively the upward movement and downward movement of the price over the most recent m months (Wang et al., 2020). These two movements can be computed with the following formulas.

$$U_t(k) = \sum_{j=0}^{k-1} I(P_{t-j} - P_{t-j-1} > 0)(P_{t-j} - P_{t-j-1}) \tag{19}$$

$$D_t(k) = \sum_{j=0}^{k-1} I(P_{t-j} - P_{t-j-1} < 0)|P_{t-j} - P_{t-j-1}| \tag{20}$$

In both equations, we find the indicator function $I(.)$ that is equal to one if the condition in the parentheses is met and otherwise has a value of zero. The variable k can have the values $k = 1, 3, 6, 9, 12$. The buying ($S^{buy}_{t,OSLT}$) and selling ($S^{sell}_{t,OSLT}$) signals for the oscillator indicators can then be formulated as follows.

$$S^{buy}_{t,OSLT} = \begin{cases} 1, & \text{if } RSI \leq 50 + \eta \\ 0, & \text{otherwise} \end{cases} \tag{21}$$

$$S^{sell}_{t,OSLT} = \begin{cases} 1, & \text{if } RSI \geq 50 + \eta \\ 0, & \text{otherwise} \end{cases} \tag{22}$$

Where in both equations η can have the following values, $\eta = 5, 10$. Together with the possible k values this makes ten buying indicators and ten selling indicators for the OSLT variable.

The last technical variable SR is generated with fifty indicators corresponding to the support-resistance trading rule. The rule produces a buying and selling signal based on the current price level and the given support and resistance levels for prices. The buying ($S^{buy}_{t,SR}$) and selling ($S^{sell}_{t,SR}$) signal are given as follows.

$$S^{buy}_{t,SR} = \begin{cases} 1, & \text{if } P_t \geq \left(1 + \frac{\eta}{100}\right) * max(P_{t-1}, P_{t-2}, ..., P_{t-k}) \\ 0, & \text{otherwise} \end{cases} \tag{23}$$

$$S^{sell}_{t,SR} = \begin{cases} 1, & \text{if } P_t \leq \left(1 - \frac{\eta}{100}\right) * min(P_{t-1}, P_{t-2}, ..., P_{t-k}) \\ 0, & \text{otherwise} \end{cases} \tag{24}$$

Where in both equations the k and η variables have possible values, $k = 1, 3, 6, 9, 12$ and $\eta = 1, 2, 3, 4, 5$. Making a total of 25 buying signals and 25 selling indicators.

B Programming Code

For our research, we mostly used Eviews to estimate the models and forecast the commodity returns. We have attached a zip file to the paper where we included all the relevant codes we have run. In this zip file, we also have an Eviews workfile. This workfile contains all of the data that we needed to perform the research. The details of the data and the codes we run are described in a 'ReadMe' file that is also included in the zip file. The special cases of the codes we have executed are explained briefly below.

For the collection of the data we also refer to Section 2, here we describe where we found the data and how we transformed them into usable variables.

B.1 Technical indicators

For the creation of the technical indicators, we used Java programming code. The class we made is called 'TechnicalIndicators' and is included in the zip file that is sent along with the paper. Comments on the code are included in the code but below we describe briefly how the code works.

The code reads the commodity index prices from an Excel file and then puts them in several lists. These lists are then processed by five methods corresponding to each of the trading rules, explained in Appendix A. When the indicators are generated by the methods, they are written in a new Excel file. Thereafter, we can transfer the indicators to our Eviews workfile.

B.2 Model estimation and forecasting

For the estimation of our models and the forecasting of the commodity returns, we used Eviews. We have written multiple Eviews programs to execute different steps in the process of model estimation and forecasting. The Eviews programs are included in the zip file. The various programs are rather easy to understand because they all perform small or repetitive tasks and contain comments. Still, we briefly describe the functions of every program in the 'ReadMe' file included in the zip file.

B.3 MCS procedure

For the Model Confidence Set procedure, we used a code in R studios, called 'Model Confidence Set'. The code reads the loss functions of our models, generated with an Eviews program, from an Excel file. Then the code splits the loss functions from the models per commodity and performs the MCS procedure to generate the SSM for each of the commodities. We perform the MCS procedure for the confidence levels 90% and 75% both corresponding to an α of 0.1 and 0.25 respectively.

C Missing Indicators

- The filtering rule selling signal for the non-energy index with k = 1 and $\eta = 10$
- The filtering rule buying signal for the agriculture index with k = 1 and $\eta = 10$
- The filtering rule selling signal for the agriculture index with k = 1 and $\eta = 10$
- The filtering rule selling signal for the food index with k = 1 and $\eta = 10$
- The filtering rule selling signal for the raw materials index with k = 1 and $\eta = 10$
- The support-resistance rule selling signal for the non-energy index with k = 12 and $\eta = 5$
- The support-resistance rule selling signal for the agriculture index with k = 12 and $\eta = 5$
- The support-resistance rule selling signal for the food index with k = 12 and $\eta = 5$
- The support-resistance rule selling signal for the raw materials index with k = 9 and $\eta = 5$
- The support-resistance rule selling signal for the raw materials index with k = 12 and $\eta = 5$

D Correlation Matrices

The correlation matrix of the macroeconomic variables, shown in Table 6, portrays the correlation between the macroeconomic variables. We see that the correlation between TBL and LTY is the only relatively high correlation value. Therefore, one can say the correlation between the macroeconomic variables is not remarkably high.

	DP	TBL	LTY	INFL	IP	RET	KI	EX
DP	1.000	0.129	0.268	0.035	0.104	0.098	-0.138	0.069
TBL	0.129	1.000	0.766	0.090	0.107	0.008	0.144	0.011
LTY	0.268	0.766	1.000	0.040	0.116	-0.019	0.172	-0.020
INFL	0.035	0.090	0.040	1.000	0.223	-0.034	0.308	0.031
IP	0.104	0.107	0.116	0.223	1.000	0.074	0.106	-0.084
RET	0.098	0.008	-0.019	-0.034	0.074	1.000	-0.002	-0.211
KI	-0.138	0.144	0.172	0.308	0.106	-0.002	1.000	-0.083
EX	0.069	0.011	-0.020	0.031	-0.084	-0.211	-0.083	1.000

Table 6: Correlation matrix for the macroeconomic variables.

As an illustration of the correlation between technical indicators, we portray the correlation matrix between the first two indicators per trading rule corresponding to energy index prices. The results are shown in Table 7.

	MOM-I	MOM-II	FR-I	FR-II	MV-I	MV-II	OSLT-I	OSLT-II	SR-I	SR-II
MOM-I	1.000	0.403	0.541	0.549	0.720	0.432	-1.000	-0.389	0.864	0.625
MOM-II	0.403	1.000	0.360	0.692	0.616	0.805	-0.403	-0.905	0.380	0.585
FR-I	0.541	0.360	1.000	0.568	0.526	0.387	-0.541	-0.366	0.626	0.650
FR-II	0.549	0.692	0.568	1.000	0.784	0.675	-0.549	-0.701	0.540	0.611
MV-I	0.720	0.616	0.526	0.784	1.000	0.612	-0.720	-0.590	0.688	0.608
MV-II	0.432	0.805	0.387	0.675	0.612	1.000	-0.432	-0.777	0.394	0.544
OSLT-I	-1.000	-0.403	-0.541	-0.549	-0.720	-0.432	1.000	0.389	-0.864	-0.625
OSLT-II	-0.389	-0.905	-0.366	-0.701	-0.590	-0.777	0.389	1.000	-0.352	-0.612
SR-I	0.864	0.380	0.626	0.540	0.688	0.394	-0.864	-0.352	1.000	0.724
SR-II	0.625	0.585	0.650	0.611	0.608	0.544	-0.625	-0.612	0.724	1.000

Table 7: Correlation matrix for first two technical indicators per trading rule for the energy index.

Where MOM-I and MOM-II in Table 7 are for example the first two indicators that are based on the momentum trading rule. See Appendix A for the calculation of these technical indicators.

E Forecasting Results for Sample until December 2017

In Table 8 one can see the percentage out-of-sample R^2 (R^2_{OoS}) for the various univariate forecasting models. We take the period from January 1982 until December 1991 as in-sample period to forecast the returns for the out-of-sample period from January 1992 until December 2017. The numbers in bold represent a positive R^2_{OoS}, which implies the model outperforms the historical benchmark model. For example, The EX model has a positive R^2_{OoS} for seven out of eight commodities. Especially the non-energy index is forecasted relatively well by the EX model. The metals & minerals index has the IP, RET and EX models performing better than the benchmark model. Interesting to see is that all the technical models outperform the benchmark model, with many models having R^2_{OoS} values higher than 4%. Nevertheless, the goal was to replicate the forecasts of Wang et al. (2020). When looking at the R^2_{OoS} values from the Wang et al. (2020) paper we still see some differences. Most of the R^2_{OoS} values are close to the ones of Wang et al. (2020), with most of the differences being not more than 1 percentage point. A possible explanation for the dissimilarity is that the commodity indexes are based on a base year. The base year of these indexes has been changed after the paper of Wang et al. (2020) was written. By changing the base year the returns change as well. Therefore the R^2_{OoS} values are not perfectly similar.

	Energy	Non-energy	Agriculture	Beverage	Food	Raw materials	Metals & minerals	Precious metals
DP	-0.328	-0.280	-0.727	-1.387	-0.842	-1.454	-0.362	**0.252**
TBL	-0.621	-2.190	-1.659	-0.293	-2.289	-1.373	-1.938	-0.624
LTY	-0.526	-1.051	-1.041	-1.085	-1.176	-2.214	-0.886	-0.250
INFL	-0.612	-0.957	-1.213	-1.491	-1.029	-0.884	-0.380	**1.688***
IP	**0.203**	**2.135**	**0.839**	**0.508**	**0.482**	-0.225	**1.428**	-0.582
RET	-2.330	**3.061*****	**0.855***	-0.368	**0.901****	-0.443	**5.502*****	-0.304
KI	-0.225	**2.006****	**2.331****	**0.001**	**1.571****	**0.693****	-0.858	-4.540
EX	**1.828****	**5.367*****	**3.291*****	-0.595	**2.351*****	**0.147**	**3.952*****	**1.115****
EW-M	**0.357**	**2.751*****	**1.944*****	**0.041**	**1.698*****	-0.054	**1.803*****	**0.846***
MOM	**1.882*****	**6.222*****	**5.467*****	**3.548*****	**4.045*****	**6.941*****	**4.810*****	**1.514*****
FR	**2.050*****	**5.271*****	**3.381*****	**3.241*****	**3.198*****	**3.819*****	**5.877*****	**0.610**
MV	**0.746****	**5.674*****	**3.963*****	**2.570*****	**1.707*****	**5.041*****	**3.937*****	**0.934***
OSLT	**1.604*****	**6.978*****	**6.173*****	**4.050*****	**4.425*****	**7.205*****	**4.837*****	**1.648*****
SR	**1.835*****	**6.347*****	**5.633*****	**3.968*****	**4.865*****	**8.694*****	**5.631*****	**0.749***
EW-T	**1.700*****	**6.469*****	**5.184*****	**3.674*****	**3.857*****	**6.734*****	**5.231*****	**1.163****

Table 8: Percentage Out-of-sample R^2 for the models estimated with ordinary least squares, while evaluating the period from January 1992 until December 2017.

F Forecasting Results for Volatile Sample

In Table 9 one can see the percentage out-of-sample R^2 values for the standard technical models. The forecast evaluation period is from January 2018 until December 2022. The numbers in bold represent positive R^2_{OoS} values. The Clark and West test is not performed here.

	Energy	Non-energy	Agriculture	Beverage	Food	Raw Materials	Metals & minerals	Precious metals
MOM	**2.098**	**4.463**	**6.957**	-0.722	**5.680**	**7.296**	**1.415**	**2.079**
FR	**3.596**	**5.202**	**3.742**	-1.045	**4.020**	**2.086**	**4.113**	**0.889**
MV	**1.479**	**2.803**	**2.687**	-2.124	**2.355**	**2.990**	-0.621	-0.412
OSLT	**2.252**	**5.313**	**7.706**	-1.866	**5.721**	**7.278**	**1.483**	**3.064**
SR	**3.548**	**10.552**	**7.723**	**1.978**	**9.842**	**7.001**	**6.434**	**2.664**
EW-T	**2.635**	**5.962**	**6.039**	-0.506	**5.756**	**5.779**	**2.800**	**1.767**

Table 9: Percentage Out-of-sample R^2 for the models estimated with ordinary least squares, while evaluating the period from January 2018 until December 2022.

G Combination Elastic Net: Θ sets

	DP	TBL	LTY	INFL	IP	RET	KI	EX	MOM	FR	MV	OSLT	SR
Energy	✓	×	×	×	×	✓	×	×	✓	✓	✓	✓	✓
Non-energy	✓	✓	✓	✓	×	×	×	✓	✓	×	✓	✓	✓
Agriculture	×	✓	✓	✓	✓	×	✓	×	×	×	✓	✓	×
Beverage	×	✓	×	×	✓	×	×	×	×	✓	×	✓	✓
Food	×	×	✓	×	✓	×	✓	×	✓	×	×	✓	×
Raw materials	×	✓	✓	×	×	✓	×	×	✓	×	✓	✓	✓
Metals & minerals	×	✓	×	×	✓	✓	×	×	×	✓	✓	×	✓
Precious metals	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	✓

Table 10: Θ sets per commodity index, generated in the combination elastic net method.