ERASMUS UNIVERSITEIT ROTTERDAM

On the performance of different forecast combination approaches: How well do technical indicators truly predict commodity prices?

Bachelor Thesis Ecnometrics and Operations Research

Author:

Virgjil Karaja (566701)

Supervisor:

Dick Van Dijk

Second Assesor:

Terri van der Zwan

June 30, 2023

The views stated in this thesis are those of the author and not necessarily those of the supervisor, second assessor, Erasmus School of Economics or Erasmus University Rotterdam.

Abstract

Are technical indicators the missing ingredient to timing commodity prices? Our paper seeks to answer the question by applying a total of 18 models generated from six predictive methods and three forecast combination approaches. We believe forecast combination to lead to better predictions of commodity index log price changes and we use various performance criteria to check the validity of this claim. Moreover, we simulate a simple portfolio strategy to see whether our models amass to monetary gains for the individual investor. We make use of the variables mentioned in Wang, Liu, and Wu (2020) and conduct our analysis over the sample, January 1982 - April 2023, where the period from January 1991 on-wards is used as our out-of-sample set. We find that our models lead to both financial gain and more forecasting power, especially those robust to overfitting, combined using equal-weights.

Table of contents

1 Introduction

In light of post-Covid financial volatility and the spikes in energy prices, market participants want to hedge their portfolio risk. As shown by Junttila, Pesonen, and Raatikainen (2018), investing in gold provides a good diversification against the stock market risk. A similar story holds for commodity markets as a whole, as they are up around 40% from pre-pandemic levels. Hence, the addition of commodity investments can be beneficial to the risk-averse investor. On the other hand, commodity prices are an indicator of the state of the economy (Groen and Pesenti, 2011), which is why they are a matter of relevance to policy makers as well. Having trustworthy forecasts of the levels of commodity prices can improve upon inflation targeting strategies due to their underlying relations as claimed by Gospodinov and Ng (2013).

A plethora of papers have been dedicated to modelling and predicting commodity prices. The common trait of most existing papers is their use of fundamental macroeconomic variables (Ahumada and Cornejo, 2016; Alquist and Kilian, 2010; Bessembinder and Chan, 1992; Chen, Rogoff, and Rossi, 2010). Small positive linkage between real-life macroeconomic indicators and commodity prices is found only over short-forecast horizons, when using an extensive set of fundamental variables (Gargano and Timmermann, 2014). We decide to build upon the work of Wang et al. (2020) and use information coming from technical indicators towards predicting commodity prices. We replicate the main result from their paper in Appendix A and we improve their approach by considering several more complex models and different weighting schemes throughout our paper. There are others who bridged the gap into the universe of technical indicators (Shen, Szakamary, and Sharma, 2007,1), using various technical rules in commodity futures trading. However, what we consider the most interesting is on whether the information provided by various technical rules can be applicable in forecasting commodity spot and futures price changes directly.

Wang et al. (2020) were the first to consider technical indicators in the context of predicting future price changes in commodities and find evidence on their predictive ability and financial profitability, all while using simple ordinary least squared regressions and naive forecast combinations. We decide to improve upon the naive methods they apply, thus assessing truly how well do technical indicators predict commodity future changes. We employ six methods: a linear regression method(LIN), elastic net(ENET), random forest(RF), principal component regression(PCR), partial least squares(PLS) and best subset selection via a modernized lens(MIO); over three combination approaches: equally-weighted(EW), minimizing mean squared prediction error(SUM) and partially-egalitarian LASSO(pLASSO); for a total of 18 models, which we motivate in Section 2.

We contribute to existing literature in two ways. Firstly, we employ complex methods and weighting schemes on the context of predicting commodity price changes, which to our knowledge, has not been done before in relevant literature. We asses the forecasting abilities of said models individually against

the historical mean as a benchmark over an out-of-sample setting, using R^2_{oos} and CSSED, plus in a multi-forecast environment using the model confidence set of Hansen, Lunde, and Nason (2011) where we provide a ranking of each forecasting model for each commodity type. Furthermore, we evaluate the financial benefits of the model and whether they are exploitable from a mean-variance optimizing investor. Secondly, our paper serves as an empirical comparison between different forecast combination techniques, while also weighing more complex machine learning regressions against simpler one-variable regressions again on the context of commodity price change forecasting.

Our data consists of eight different commodity indices and 105 technical variables generated in the same manner as in Wang et al. (2020), but over a longer sample, January 1982 up to April 2023. We set the period from January 1991 until April 2023 as our out-of-sample set to perform one-step ahead expanding window forecasts over. We find that their simple model is outperformed by the models we propose, with the exception of MIO, suggesting that including all 105 technical indicators enhances predictive power. The RF and dimension reduction(PCR and PLS) variants provide the most consistent forecasts, as shown by them being ranked the highest the most often by our comparison measure, MCS. When regarding different combination methods, pLASSO predicts the worst implying the importance of considering all five technical rule, while EW performs the best (Claeskens, Magnus, Vasnev, and Wang, 2016) for almost all commodity indices, followed closely by SUM. From a financial point of view, Ew loses its advantage over SUM and they give akin risk-adjusted returns and guaranteed utility, which are considerably positive and of use by the individual investor. On the other hand, pLASSO falls short again and fails to outperform a simpler portfolio using knowledge of only the risk premium and the historical average.

The remainder of the paper is organized in the following manner. In Section 2, we provide the relevant literature connected to the topic at hand and the motivation behind the techniques we apply. Section 3 goes over the methodology, starting with the different models and combinations, continuing with the predictive performance measures, ending with the portfolio creation and financial benefit criteria. We provide our main findings in Section 5 before concluding in Section 6. Appendices A through G contain useful supporting information such as Figures, Tables, plus other tests and measurements.

2 Literature Review

As mentioned, Gargano and Timmermann (2014) find only a small linkage between macoreconomic indicators and commodity prices. In addition, Wang et al. (2020) show that technical indicators dominate fundamental ones on both predictive power and financial profitability, hence we choose to focus solely on technical indicators in our analysis. We test the short-term forecasting ability of our models and assume it will generalize to longer-horizons as shown by Wang et al. (2020). The set of 105 technical variables from five rules is rather large so we do not see much merit in increasing the number of variables, but rather opt for a longer sample period.

The empirical (Rapach, Strauss, and Zhou, 2009a; Wang, Liu, and Wu, 2017) and theoretical (Timmerman, 2006) advantages of using the equally-weighted averaging are known and are the norm when conducting forecast combinations. Stock and Watson (1999, 2003) claim the existence of the so-called forecast combination puzzle, where the naive average empirically outperforms the more complex, "optimal" weighting schemes. Claeskens et al. (2016) provide various intuitive explanations behind the phenomena, one of which is associated with the error that comes from finding optimal weights. For that reason, we suggest finding weights as to minimize the prediction error associated with the set of forecasts as done in Hsiao and Wan (2014). We implement the conventional normalization condition of weights that sum up to one, by Markowitz (1952). We also consider the alternative constraint of Hsiao and Wan (2014), where they set the dot product of the weight vector with itself to be equal to 1, and provide the results in Appendix G. We believe that with combinations that specifically target the prediction errors, we can outperform the naive weights. However, these combination approaches would lead to subpar results in the cases when some of the forecasts to be combined are better off as discarded, as Diebold and Shin (2019) argue is often the case. To avoid that unpleasant result, we propose using their pLASSO approach, to check if the information of certain technical rules damages the overall predictability of ln commodity price changes.

The OLS estimation is known for its benefits, but also the various shortcomings it suffers from, explicitly its failure in dealing with correlated regressors and inefficiency when modelling non-normal data. To avoid the multicolinearity pitfall, we choose to run multiple one variable regressions and aggregate over the forecasts of the same technical rule instead of including the all at once or running one regression per rule type. We denote this method as LIN. An argument can be made for using other weighting schemes, but we believe this to be the nicest approach to deal with indicators of the same family as it retains the predictive abilities of each variable. We believe the forecasting capabilities to vary more across technical rules than within and that is why we focus our different combination approaches on handling forecast series from different families and end up with five forecast series to be combined, for LIN and all the other methods.

Including the regressors one-by-one is not ideal and leads to omitted-variable bias (Heij, de Boer, Franses, Kloek, van Dijk, et al., 2004). Hence, we propose two models that perform well in correlated environments, the elastic net of Zou and Hastie (2005) and the random forest of Breiman (2001). The other route to take when dealing with a large number of regressors is dimensionality reduction. To that extent, we introduce principal component regression (PCR) as done by Massy (1965) and partial least squares (PLS) regression created by Gerlach, Kowalski, and Wold (1979). While we suspect all indicators to have some explanatory power, on which we motivate the construction of our models, we want to allow the possibility of removing indicators that are not deemed as that important. To that extent, the last method we employ is the best subset selection method via a modernized optimization lens (MIO) from the paper of Bertsimas, King, and Mazumder (2016).

Our goal is to determine both forecasting ability and possible financial exploits and our paper tackles both issues. In both cases we choose to use the historical mean as our benchmark comparison following the remarks of Campbell and Thompson (2008). As measures of forecasting ability, we use R^2_{oos} and CSSED, following Wang et al. (2020). Furthermore, we employ model confidence set (MCS) procedure of Hansen et al. (2011), since we have various models to be compared. To asses possible financial gains we consider the mean-variance optimizing investor of Rapach, Strauss, and Zhou (2009b). We create a simple portfolio managing strategy as done by Wang et al. (2020) and rely on $SR_{\%}$ and CER to measure profitability of each model, following Rapach et al. (2009b).

3 Data

Our data consists of two main parts, the commodity prices and the technical indicators we use as our explanatory variables. We briefly mentioned the 3-month Treasury Bill rate used as our aprproximate for the risk-free rate.

3.1 Commodity prices

We make use of monthly data from the World Bank's Website[1] on eight different commodity indices. The indices we use in our analysis are: energy, non-energy, food, beverages, agriculture, metals&minerals, precious metals and raw materials. This indices are weighted averages of the commodities included in the respective category. For more detailed information, we refer you to the World Bank's Website. When computing the technical indicators we use the prices of the commodity indices as given in the website, however, for the predictive regressions we opt for using ln price changes as the dependent variable to remove the non-stationarity that is present on the original prices. We denote the dependent variable as y_t for each type of commodity index and compute it as:

$$y_t = ln(P_{t+1}) - ln(P_t). \tag{1}$$

Our sample ranges from January 1982 until April 2023, where we make a distinction between an in-sample and an out-of-sample period. The former takes place from January 1982 until December 1990, meaning that the latter takes place over the remainder of the total dataset. In Appendix B we provide in Table 6 the summary statistics of the eight commodity index prices and in Table 7 the summary statistics for their log price changes alongside their historical graphs, 2a and 2b. Moreover, we use the 3-month Treasury Bill rate obtained from the FRED's Website[2] as our proxy for the risk-free rate motivated by Dolatabaldi, Narayan, Nielsen, and Xu (2018). The rate from the website is in annualized percentage points, hence we divide by 1200 to turn it into a monthly decimal format.

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

[2] https://fred.stlouisfed.org/series/DTB3

3.2 Technical Indicators

We follow the work of Wang et al. (2020) and use the technical indicators they propose as our potential regressors. We refer you to their paper to understand the formulas used to generate the technical indicators and only provide the precise formulas in Appendix C. There is a total of 105 variables, generated by five different famous technical rules, momentum, moving average, filter, oscillator and support-resistance. Each rule leads to 5, 20, 10, 20 and 50 technical indicators respectively, which are used as the exogeneous parameters that go into the models. The predictors are all made into binary variables. The momentum rule compares the current price to the k period lagged price, where we pick several values for k as explained in Appendix C. The filter rule stacks up the current price against the most recent maximum and minimum. The moving average rule compares the short-term moving average versus the long-term one. Furthermore, the oscillator rule generates buying and selling signals if the price movements have been too quick. Lastly, the support-resistance rule also compares the current price with the most recent local minima and maxima, but provides buying/selling signals in an opposite manner to the filter rule.

4 Methodology

4.1 Models

We suggest several models for modelling log price changes using various technical indicators. In our approach, we try to estimate one forecast per technical rule: moving average, momentum, filter, oscillator and support-resistance, using the technical indicators generated by each rule as potential predictors. We estimate our models over an expanding window, starting with the in-sample period and making one-step ahead forecasts for each period of our out-of-sample set.

LIN In our paper we model the log change in returns as displayed in Equation 2:

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

where $\epsilon\ N(0, \sigma_\epsilon^2)$. The explanatory variable, x_t, takes values from the set of different technical indicators we employ and lastly, the ϵ_{t+1} denotes the error terms which are assumes to be identically and independently distributed through time. We estimate forecasts over the out-of-sample period from January 1991 to April 2023 using OLS regressions in R.

In total we obtain 105 OLS estimators and 105 matching forecasts, which we average into five forecast series, one for each originating technical rule. We find the approach of grouping predictors as an intuitive approach that should lead to better estimates, as forecast combination is known to surpass individual forecasts in performance (Timmerman, 2006). Additionally, it has the added benefit of avoiding the multicolinearity pitfall of OLS regressions.

ENET The elastic net is a machine learning method that performs penalized regressions, first suggested by Zou and Hastie (2005). It performs regressions in a linear model, with the penalty $\alpha\|\beta\|_1 + (1-\alpha)\|\beta\|_2^2$,

meaning that as α approaches 0, the elastic net estimates approach those of the ridge (Hoerl and Kennard, 1970) and when it approach 1, it resembles LASSO (Tibshirani, 1996). The solution of the elastic net is found after performing the optimization problem in Equation 3:

$$\min \frac{1}{2n}\left\|y_i - X_j\beta\right\|_2^2 + \lambda((\frac{1-\alpha}{2})\|\beta\|_2^2 + \alpha\|\beta\|_1), \tag{3}$$

where y corresponds to the vector containing the i-th commodity log price changes and X_j is the matrix containing all the technical indicators generate by the j-th rule, since ENET can deal with correlated regressors at once. This way we obtain five forecast series, one for each technical rule, and we follow the same route for the other methods as well. The elastic net retains good performance in an environment where multicolinearity is present, plus it combines both the penalization of ridge with the variable selection of LASSO.

As there is no closed form solution to Equation 3, the elastic net is solved via numerical methods for which we use the 'glmnet' package and its extension 'glmnetUtils' in R. To solve the problem we need to know the values of the exogenous parameters, α and λ. To that extent we emphasize the validity of using cross-validation as proven by Picard and Cook (1984) and we settle for using 10-fold cross-validation as it is a good average; less biased than 5-fold, smaller variance than leave-one-out in our not so small dataset, less computationally expensive than bootstrapping and bagging (Breiman, 1996). The series of α's we search over is user-specified[3], while the series of λ'is chosen by the method itself as to improve the convergence rate[4].

RF The random forest is built upon several decision trees, where each new observation is categorized on the basis of its characteristics. We provide a visualization of such a tree in the Appendix B, Figure 3, adapted from Murphy (2012) and we refer you to their book for further explanation. However, the random forest is not useful for just classification, but it also provides forecasts of a dependent variable by taking the mean of the required characteristic over all the observations belonging to the same final node that the new observation is placed into. For a more detailed explanation, see Hastie, Tibshirani, and Friedman (2009).

The RF has the benefits of dealing with multicolinearity, plus being less susceptible to outliers (Breiman, 2001). Moreover, to reduce the high variance associated with a single decision tree (Hastie et al., 2009), the random forest is built using bagging (Breiman, 1996) where a number of separate bootstrap samples is taken from the original data to train the trees and finally an average is taken across the trees. Additionally, the random forest adds a randomness feature in variable selection to reduce the underlying correlation between the trees and the number of variables subsetted can be seen as a tuning parameter, making RF less prone to overfitting.

[3]We test 10 values for α ranging from 0 to 1 in equal intervals

[4]See 'cva.glmnet' in the package manual found in https://cran.r-project.org/web/packages/glmnet/index.html

We train one random forest model for each technical rule containing as regressors all of the technical indicators built from said rule and then provide forecasts over the out-of-sample period. To estimate the random forests, we use the 'ranger' package in R, with the default values for the bootstrapped trees(500), for the maximum tree depth(infinite), for the spit rule(variance), for the minimum node size(5), for the number of variables to use in each split(rounded down square root of the total number of regressors), etc.

PCR Principal Component Regression (PCR) (Massy, 1965) is a regression method that utilizes the benefits of principal component analysis. Applying PCR is a three step approach. Initially we perform PCA (Wold, Esbensen, and Gelaldi, 1987) over the set of p predictor variables. Next, we keep a number k of the principal components we find with usually k << p. Lastly, we perform a simple linear regression using the k principal components as the regressors. PCR does well when the regressors are correlated as it captures most of the total variance with the extra advantage of retaining the same set of final regressors (the principal components) for each commodity type, making it more resilient to overfitting.

We use the package 'caret' to estimate the principal component regressions. We compute one forecast for each technical rule using the indicators created by that rule as predictors. We use 10 fold-cross validation based on random selection for selecting p where the model that produces the smallest RMSE in-sample is used. The number of generated samples for each tested value of p is pre-select at 10 (similar reason as for using 10-fold cross-validation) while the other parameters are used at default values.

PLS Partial least squares (PLS) Gerlach et al. (1979) is another regression method that aims to transform the set of explanatory variables into a new set with smaller cardinality, but that still retains most of the variation in the variables. The general model that PLS assumes is displayed by Equations 4 and 5:

$$X = TP^T + E, \tag{4}$$

$$Y = UQ^T + F. \tag{5}$$

X and Y denote the matrix of predictors and dependent variables (in our case a vector) respectively, while T and U are their projections into a subspace with fewer final variables. Consequently, P and T are the loading matrices while E and F are the associated errors, assumed to be independently and identically distributed. The decomposition is done in such a way to maximize the pairwise covariance between T and U. Differently from PCA, orthogonality is implied onto the scores, not the loadings. PLS performs supervised regressions, where the final predictors are chosen to also capture the variability in the dependent variable. As a result, the models found by PLS often have a higher fit with fewer regressor and guarantee better in-sample performance, but they have a higher tendency to overfit when compared to PCA.

We use the package 'caret' to estimate the partial least squares regression. We compute the forecasts in the same manner as for the previous methods, one per technical rule. We tune the number of final predictors to be kep in the same manner as we do for the PCR and the other values are left at default values.

MIO The last method we propose, is the best feature selection approach via a modernized optimization lens (MIO) (Bertsimas et al., 2016). It aims to select the best k predictors from an initial set of p variables by solving a revamped mixed integer optimization problem. The formal formulation of their approach can be found below:

$$\min_{\beta,z} \frac{1}{2}\|y - XB\|_2^2 \tag{6}$$

$$s.t \quad -\mathcal{M}_{\mathrm{U}}\mathrm{z}_i \leq \beta_i \leq \mathcal{M}_{\mathrm{U}}\mathrm{z}_i, \; i = 1, ..., p \tag{7}$$

$$\mathrm{z}_i \in 0, 1, \; i = 1, ..., p \tag{8}$$

$$\sum_{i=1}^{p} \mathrm{z}_i \leq \mathrm{k}. \tag{9}$$

$\mathcal{M}_{\mathrm{U}}$ is picked such that $\mathcal{M}_{\mathrm{U}} \geq \left\|\hat{\beta}\right\|_\infty$ and $\mathrm{z} \in \{0,1\}^p$. This formulation basically means that if $\mathrm{z}_i = 1$, then $|\beta_i| \leq \mathcal{M}_{\mathrm{U}}$ and otherwise $\beta_i = 0$, hence $\|\mathrm{z}_i\|_1$ corresponds with the number of predictors in the final model. To understand how $\mathcal{M}_{\mathrm{U}}$ is picked and the related propositions we refer you to the original paper of Bertsimas et al. (2016). To put it simply, MIO minimizes the objcetive function of Equation 6, while restricting the final weights to be lower than a pre-select $\mathcal{M}_{\mathrm{U}}$ in absolute value.

In the context of our research, it provides near optimal solutions to the feature selection problem in reasonable running time and overcomes the drawback of convex-based optimization methods. To solve the MIO models we use the 'bestsubset' R package, which makes use of the Gurobi 10.0 solver, applied to a R environment. We compute forecasts over the out-of-sample period for each technical rule by passing down its corresponding indicators as the set of potential predictors. Due to the non-polynomial running time of the algorithm, tuning is not a favourable option hence we hand-picked the number of features to be selected in the final model and we setting it equal to the square root of available predictors rounded up, following the approach of Bertsimas et al. (2016). Additionally, solving to optimality could take up to several hours, hence we set an artificial timer limit of 120 seconds[5].

4.2 Forecast Combinations

By construction, we generate, using the various models, five series of forecast for each commodity type. Now, we explain the forecast combination methods we apply to join the five series into one final, better prediction.

EW One of the combinations we consider, is the simple $\frac{1}{N}$ average. Timmerman (2006) shows that equal weights are optimal in situations when the individual forecast errors have the same variance and identical pair-wise correlations, which not need be the case, but provides a decent benchmark. We generate the equally-weighted forecasts (EW) using the formula in Equation 10:

$$\hat{y}_t^{j\_EW} = \frac{1}{N}\sum_{i=1}^{N} \hat{y}_{i,t}^{j}, \tag{10}$$

[5]The parameters set in the R function do not correspond to the actual running time, but they just ensure that a forecast is made within two minutes on each iteration

where i denotes the technical rule and j denotes the method used.

SUM Another combination method we suggest is designed to minimize the mean squared prediction error (MSPE) of our predictions. We generate prediction errors of $N \times 1$ (N = 5), where each element of the vector is computed as shown in Equation 11:

$$\mathrm{v}_{i,t}^{j} = y_t - \hat{y}_{i,t}^{j} \tag{11}$$

Having computed the prediction vectors, we minimize the expression in Equation 12, corresponding to the MSPE of the forecasts:

$$\min\ w^T \Sigma w, \tag{12}$$

where w^j is the vector containing the weights assigned to the forecasts of each technical rule and Σ is the covariance matrix of the prediction errors. To avoid data snooping, the covariance matrix of each period t in our out-of-sample set is computed using all the information available up to time t-1, according to the formula in Equation 13:

$$\Sigma_t = \mathrm{v}_{i,t-1}^{j} \mathrm{v}_{i,t}^{j^T}. \tag{13}$$

As argued by Hsiao and Wan (2014), we need a normalization condition to solve the minimization problem of Equation 12. We choose to use the conventional condition of Markowitz (1952), also explained by Timmerman (2006) and Newbold and Granger (1974). The normalization criteria is that of weights summing up to 1, displayed in Equation 14:

$$e^T w^j = \sum_{i=1}^{N} w_i^j = 1. \tag{14}$$

Minimizing 12 with respect to 14 leads to the optimal weights shown in Equation 15:

$$w^{j^*} = (e^T \Sigma e)^{-1} (\Sigma^{-1} e), \tag{15}$$

where e is a $N \times 1$ vector of ones. Having computed the optimal weights for each method, the final forecast (SUM) is as shown in Equation 16:

$$\hat{y}_t^{j\_SUM} = \sum_{i=1}^{N} w_i^{j^*} \hat{y}_{i,t}^{j}. \tag{16}$$

pLASSO The last combination method we implement is the partially-egalitarian LASSO (pLASSO). Diebold and Shin (2019) argue that from the intial set of forecasts, it is often the case that some are better off being discarded, and the final optimal weights should be shrunk towards the simple average. To that avail, they come up with the pLASSO, a two-step method that combines variable selection with shrinkage towards simple average weights. Mathematically, it aims to solve the double penalization problem of Equation 17:

$$\hat{\beta}_{pLASSO}^{j} = arg \min_{\beta^j} \left(\left\| y_i - F^j \beta^j \right\|_2^2 + \lambda_1 \left\| \beta^j \right\|_1 + \lambda_2 \left\| \beta^j - \frac{1}{p(\beta^j)} \right\| \right), \tag{17}$$

where $p(\beta^j)$ denotes the number of non-zero elements in β^j. The first penalty is the standard LASSO penalty, which shrinks to zero and performs variable selection, while the second penalty shrinks towards

the mean of the remaining coefficients. The authors do not provide a clear-form solution[6], but provide a simple two-step implementation.

1. Using a normal LASSO regression, as in Equation 18, select k forecasts out of the five forecasts of each technical rule. Then before moving to Step 2, we filter our β^j and F^j to contain only the forecast that were not discarded by the LASSO.

2. Using egalitarian Ridge (eRidge) shrink the forecasts with a non-zero coefficient estimate towards $\frac{1}{k}$. eRidge can be computed using the formula of Equation 19, which is essentially a normal Ridge regression, but the dependent variable is no longer y_t, but $(y_t - \bar{f_t^j})$ with $\bar{f_t^j}$) being the mean of the $N \times 1$ vector of technical forecasts. We refer you to Figure 5 in Appendix G (adapted from Diebold and Shin (2019)) as to why this substitution is equivalent to eRidge.

To avoid the look-ahead bias, we estimate the coefficient at each period t in our out-of-sample set using the information available only at time t-1, similar to an expanding window, but where the forecasts from our models are treated as the actuals.

$$\hat{\beta}^j_{LASSO,T} = arg\min_{\beta^j} (\left\| y_i - F^j\beta^j \right\|_2^2 + \lambda_1 \left\| \beta^j \right\|_1) \tag{18}$$

$$\hat{\beta}^j_{pLASSO,T} = arg\min_{\beta^j} ((\left\| y_i - \bar{F^j} - F^j\beta^j \right\|_2^2 + \lambda_2 \left\| \beta^j - \frac{1}{p(\beta^j)} \right\|) \tag{19}$$

After computing the pLASSO, we estimate the one-step ahead forecasts as depicted by Equation 20, where the final coefficients function as weights:

$$\hat{y}_t^{j\text{-}pLASSO} = \sum_{i=1}^{p(\beta^j_{pLASSO,T})} \beta^j_{pLASSO,T,i}\, \hat{y}^j_{i,t}. \tag{20}$$

4.3 Comparison Measures

We want to compare the forecasting performance of the various models we propose. To that extent we use three methods, the out-of-sample R squared, the cumulative sum of squared predictive error difference and the model confidence set.

R^2_{oos} We follow suit the literature of Wang et al. (2020) and Campbell and Thompson (2008) and use the percentage out-of-sample R^2 (R^2_{oos}) as one of our measures for validating forecasting performance. It gives a direct comparison between forecasting models and assigns a value to the difference in their effectiveness. The metric is calculated using the formula in Equation 21, making use of the MSPE (given at Equation 22):

$$R^2_{oos} = 100 \times (1 - \frac{MSPE_j}{MSPE_{bench}}),, \tag{21}$$

$$MSPE^j = \frac{1}{T-M} \sum_{t=M+1}^{T} (\hat{y}_t^j - y_t)^2, \tag{22}$$

where j denotes the method used and T-M is the number of observations in the out-of-sample period. As our benchmark we take the historical average of price changes, $\hat{y}^{bench}_{t+1} = \frac{1}{t}\sum_{i=1}^{t} r_i$ due to its good empirical

[6]The objective function is not continuous at points where the coefficient estimate is 0

performance, as displayed by Campbell and Thompson (2008). Hence, a positive R^2_{oos} would imply that the model outperforms the benchmark and a negative one would imply the opposite, but despite its clear interpretation, the R^2_{oos} suffers in giving a visual perspective on how the forecasts from the model differentiate from those of the benchmark.

To test the statistical significance of the R^2_{oos}'s we utilize the method of Clark and West (2007) (CW), which is an adaption of Diebold and Mariano (1995) for when dealing with nested models. They compute a series of L statistics as shown in Equation 23, and regress it on a constant where the estimate for the constant is the CW statistic. A significant CW statistic indicates that the MSPE's of the two models are not the same.

$$\mathrm{L}_t = (y_t - \hat{y}_t^{bench})^2 - (y_t - \hat{y}_t^{j})^2 + (\hat{y}_t^{bench} - \hat{y}_t^{j})^2 \tag{23}$$

CSSED The cumulative squared predictive error difference (CSSED) is a measure for evaluating forecasting performance over time. Differently from R^2_{oos}, it provides a visual interpretation of how the forecasting ability of two models changes over time. We use the formula of Equation 24 to compute the CSSED at time t for each of our 18 models, against the benchmark:

$$CSSED_T^j = \sum_{t=M+1}^{T} (e^2_{bench,t} - e^2_{j,t}). \tag{24}$$

An increase in CSSED from time t-1 to t, implies that the model gives more precise forecasts than the benchmark at time t and vice-versa.]

MCS The model confidence set (MCS) is an approach with the advantage that it can compare the statistical significance of several forecasting methods at once. It creates a set of models that are the best from the total set of model with a certain probability. The steps of MCS are complex and explained in detail in Hansen et al. (2011), however we give a brief overview of the general idea and algorithm. The MCS starts by setting a set of superior models as in Equation 25:

$$\mathcal{M}^* := \{i \in M_0 : \mathbb{E}(d_{ij,t}) \leq 0, \forall j \in M_0\}, \tag{25}$$

where $d_{ij,t}$ is the loss differential between model i and j, at time t and M_0 is the set containing all models. The algorithm for the method is a series of consecutive tests where the null hypothesis of Equation 26 is considered:

$$H_{0,M} : \mathbb{E}(d_{ij,t}) = 0 \quad \forall i, j \in M. \tag{26}$$

The algorithm starts with the full set M_0 and tests hypothesis 26 recursively to determine which models are kept and which are discarded from the superior set at a confidence level of α. MCS also provides a p-value associated with each model in the set where the higher the p-value, the more likely the model is part of the superior set.

We use the 'MCS' R package for conducting the MCS procedure. We set the significance level at 15% following Hansen et al. (2011) and use the maximum as the t-statistics of $d_{ij,t}$ as our test statistic, plus

we implement 5000 bootstrap samples to overcome the non-standard distribution of our chosen statistic. We use the squared error terms as our loss function and set the other parameters at their default values.

4.4 Financial Gain

We go into a lot of detail analyzing the forecasting performance of the model, but with that alone we cannot say whether these strategies are financially beneficial for the individual investor. That is why we propose a simple portfolio strategy to determine whether the models do lead to higher returns than the historical average benchmark.

4.4.1 Portfolio Creation

We imagine an investor with a mean-variance preference who allocates his wealth between the risk-free asset and the commodity index. The utility of such an investor would be:

$$\mathrm{U}_t = \mathbb{E}(w_t R_t + r_{f,t}) - \frac{1}{2}\gamma var(w_t R_t + r_{f,t}), \tag{27}$$

where w_t is the weight given to the commodity index, R_t is the commodity excess returns at time discounted by the risk-free rate, $r_{f,t}$ (approximated by the 3-month T-Bill) all at time t. The investors aversion to risky assets is given by γ, which we set at 2^7 following Wang et al. (2020). Using our log commodity price change forecasts ex-ante and volatility estimates over a 5-year rolling window, we obtain the optimal weights in Equation 28:

$$w_t^{j^*} = \frac{1}{\gamma}(\frac{\hat{R}_{t+1}^j}{\hat{\sigma}_{t+1}^2}). \tag{28}$$

Having computed the optimal portfolio weights, we can compute each period's returns over our out-of-sample set with the exception of the last observation, April 2023, because the weights at time t are computed using forecasts of time t+1. The portfolio returns are determined using Equation 29:

$$R_{p,t+1}^j = w_t^{j^*} R_{t+1} + r_{f,t+1} - \tau|w_{t+1}^{j^*} - w_t^{j^*}|, \tag{29}$$

where τ denotes the transaction costs. We set the transaction costs at 0.033% as done by Locke and Venkatesh (1997).

4.4.2 Portfolio Evaluation

The measures we rely on to compute the financial gain of our portfolios are the sharpe ratio gain ($SR_{\%}$) and the certainty equivalent returns gains (CER). The $SR_{\%}$ captures the extra risk-adjusted returns that a strategy based on the commodity type prices can earn when compared to a strategy focused around the historical benchmark. On the other hand, it fails to distinguish between the benefits of using the forecasts to invest on commodities versus depending on the risk-free asset alone. To avoid over-relying on $SR_{\%}$, we make use of the CER, since it manages to capture the utility gains on top of the risk premium for an individual investor. We compute $SR_{\%}$ as the difference in SR (Sharpe, 1966) multiplied by a factor of $\sqrt{12}$ to get the annualized value. Similarly, we report the CER as the difference in the investor's utility

[7] Rapach et al. (2009b) show that results remain qualitatively similar despite the value of γ

(U) multiplied by a factor of 1200 to get the annualized percentage value. We compute the SR and $SR_{\%}$ in Equations 30 and 31 respectively, while for the investor's utility we use the formula of Rapach et al. (2009b), Equation 32 and for CER we use the formula in Equation 33.

$$SR^j = \frac{\bar{R}_p^j}{\sigma_{R_p^j}}, \tag{30}$$

$$SR_{\%}^j = \sqrt{12}(SR^j - SR^{bench}), \tag{31}$$

$$\mathrm{U}^j = \bar{R}_p^j - \frac{1}{2}\gamma\sigma_{R_p^j}^2, \tag{32}$$

$$CER_{\%}^j = 1200(U^j - U^{bench}). \tag{33}$$

5 Results

5.1 Predictive ability comparisons

5.1.1 The out-of-sample R squares

We display the R_{oos}^2's that we obtain in Table 1. We see that all models provide positive R_{oss}^2's, statistically significant in most cases, with the exception of the MIO and several pLASSO combinations. MIO is the only feature selection approach that discards some of the predictors we pass it. An obvious insight from the occurrence is that all technical indicators add explanatory value when included in the regressions and discarding them could worsen the forecasts to the point where the historical average benchmark is a better alternative (as it did). While including more or all indicators could lead to over-fitting worries, again we show that methods robust to over-fitting, like RF and PCR perform well. In fact, the methods we suggest outperform the already outperforming combination of Wang et al. (2020). Their model, denoted by LIN_EW, gets dominated in terms of R_{oos}^2, by one or more complex models, for each commodity index and for each combination approach with the exception of pLASSO. We can safely say that forecast combinations considerably improve upon the single forecasts of the historical benchmark, supporting the claims of Timmerman (2006).

Our results show that the MSPE minimizing forecast combination and the equally-weighted combinations have similar forecasting results across all methods for most of the commodity indices, with the energy index being captured better by EW models, while the food index is better predicted by the SUM models. Considering that their performance is similar and that EW is a lot less computationally-demanding, then the naive average is to be preferred. We reach the same so-called "forecast combination puzzle", analyzed by Claeskens et al. (2016), where the naive combination seems like the best alternative. Nevertheless, the highest R_{oos}^2 of 14.41 is attained by RF_SUM. It seems that the nature of the approach to minimize the forecast-related errors, leads to its good performance, but is suffers from the estimation error that comes from the recursive computation of the covariance matrix of errors.

Table 1: The R^2_{oos} forecasting results

Panel A: Equally-weighted forecast combinations								
	Eng	Non-en	Agrc	Bev	Food	Raw.Mats	Met&Min	Prec.Met
LIN_EW	2.16**	6.34**	5.06**	3.09*	3.93*	6.37**	4.97**	1.25
ENET_EW	3.89**	10.33**	10.52**	4.22**	7.74**	11.68**	7.85**	1.65*
RF_EW	3.76**	11.13**	11.94**	3.22**	7.58**	12.28**	9.51**	2.42**
PCR_EW	4.26**	11.44**	11.49**	3.73**	7.22**	12.38**	7.27**	2.86**
PLS_EW	4.79**	9.87**	10.99**	3.46**	7.98**	11.69**	9**	2.6**
MIO_EW	3.7**	7.56**	3.86**	-0.35*	1.2*	5.76**	4.09**	-0.08
Panel B: MSPE minimizing weights combination								
	Eng	Non-en	Agrc	Bev	Food	Raw.Mats	Met&Min	Prec.Met
LIN_SUM	1.19	7.6**	11.7**	1.15*	7.75**	10.33**	7.14**	-0.85
ENET_SUM	-1.68**	7.21**	9.07*	0.93*	11.23**	9.24**	4.76**	-6.88
RF_SUM	1.72**	6.16**	14.41**	1.31**	6.96**	12.05**	6.58**	0.056
PCR_SUM	1.79**	14.66**	11.71**	1.61**	9.42**	11.41**	6.98**	0.77*
PLS_SUM	2.23**	4.77**	10.09**	2.06**	6.49**	7.43**	8.24**	-1.11*
MIO_SUM	2.38**	2.79*	0.52**	-4.51	-0.83*	3.4**	2.86**	-1.87*
Panel C: Partially-egalitarian LASSO combination								
	Eng	Non-en	Agrc	Bev	Food	Raw.Mats	Met&Min	Prec.Met
LIN_pLASSO	0.11*	6.34**	10.88**	2.69*	7.94**	10.17**	6.17**	-0.26
ENET_pLASSO	-1.15	1.57	6.20**	1.27	6.90**	7.05*	2.77**	-14.19
RF_pLASSO	-3.98*	-3.08*	2.89*	-2.20	-4.59	-1.05	-2.13*	-16.20*
PCR_pLASSO	-0.30	3.83	5.99**	0.46	1.66*	4.54**	1.94**	-0.33
PLS_pLASSO	0.83*	1.80	3.10*	0.76	-0.85	2.01	0.22	-0.67
MIO_pLASSO	1.61**	-3.02	0.49	0.70	-1.24	1.57	2.66*	-0.39

Notes: The table contains the R^2_{oos}'s computed over the out-of-sample period beginning from January 1991 until April 2023. The portfolio names are constructed as the forecasting method plus the forecast combination approach and are categorized into panels depending on said combination approach. The R^2_{oos}'s are computed using Equation 21 against the historical average benchmark. Lastly, we use the Clark and West (2007) method to test statistical significance and * denotes significance at the 5% significance level and ** at the 1% level. A significant, positive R^2_{oos}, means that the model is statistically better than the benchmark and vice-versa.

On the other hand, the partially-egalitarian LASSO combination is dominated by the other two combinations, also being the only combination that gives negative R^2_{oos}'s. Our suspicion as to why that happens is similar to the explanation behind MIO's poor performance. We argue that the initial step where some of the forecasts are discarded by the initial LASSO, inevitably leads to the loss of valuable information. Each forecast that is passed to the first step of the method, contains all the information associated by a given technical rule. Thus, we claim the poor performance of the pLASSO as proof that all five technical indicators contain high explanatory power towards modelling the log price changes of commodity prices, affirming the initial result of Wang et al. (2020). In addition, we provide the MSPE values, which lead

to the R^2_{oos}'s, as computed in Table 1, in Appendix D.

5.1.2 The out-of-sample cumulative sum of squared predictive error difference

For each commodity type, we depict the CSSED^j for each model j over our out-of-sample period in one graph. One of these graphs can be found in Figure 1 corresponding to the agriculture index. We put the graphs for the other commodity types in the Appendix G (Figures 4a, 4b, 4c, 4d, 4e, 4f and 4g), but the main conclusions represent those obtained by looking at the R^2_{oos}'s. The pLASSO combinations are more often than not in the bottom bunch, sometimes in the negative meaning that the historical benchmark is a better alternative. In the agriculture case, MIO_pLASSO displays performance basically on par with the historical mean. We can see the pattern of the models generated by SUM and EW mixed near the top, with RF_SUM performing the best over time. This model and its equally-weighted equivalent, RF_EW, perform the best over time for most commodity indices. In the categories where this models do not perform that well, PCR and PLS variants have the highest CCSED's. A common pattern that can be seen through the different models is a sudden jump in performance around the 2008 financial crisis and the 2019 Covid outbreak, implying that these complex models greatly exceed the benchmark in foreseeing the unforeseeable.

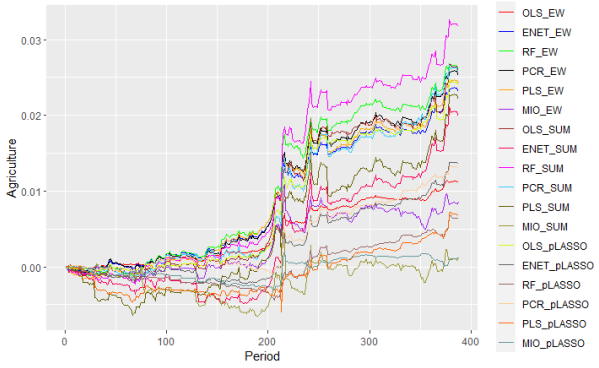


Figure 1: The CSSED computed over-time against the historical average benchmark for the Agriculture commodity index

5.1.3 The model confidence set results

The model confidence set procedure further affirms our beliefs, see Table 2. The SUM generated moels have a few "podium" positions across the different commodity indices with RF_SUM and PLS_SUM in particular being in the top three for 4 of the commodity indices. However, the EW combinations capture the most top three positions across the board. In particular RF_SUM has the most number one postions, while PCR_EW is on the "podium" in all, but two commodity types. The naive LIN_EW combination of Wang et al. (2020) performs average at best, being dominated by the more complex methods. Additionally, the two dimension reduction techniques, have very similar performance despite the combination method. The added supervision that comes as benfit of PLS does not visibly lead to better results, suggesting that the principal components that capture the most variability manage to

capture the variance of the dependent variables as well. Moreover, we provide the results of the MCS when employed in the context of each combination type in Appendix E.

Table 2: The MCS procedure results

Panel A: Equally-weighted forecast combinations								
	Agric	Bev	Eng	Food	Met&Min	Non-en	Prec.Met	Raw.Mats
LIN_EW	-	5(0.99)	8(0.74)	10(0.3)	9(0.77)	11(0.18)	12(0.35)	-
ENET_EW	7(0.87)	5(0.97)	3(1)	6(0.59)	6(0.95)	3(0.92)	6(0.78)	7(0.96)
RF_EW	6(0.91)	4(1)	6(0.95)	9(0.32)	1(1)	4(0.92)	2(1)	1(1)
PCR_EW	3(0.95)	3(1)	2(1)	8(0.32)	8(0.83)	2(0.98)	1(1)	2(1)
PLS_EW	5(0.92)	2(1)	1(1)	7(0.56)	2(1)	5(0.89)	3(1)	6(1)
MIO_EW	-	12(0.36)	4(1)	-	-	9(0.4)	15(0.22)	-
Panel B: MSPE minimizing weights combination								
	Agric	Bev	Eng	Food	Met&Min	Non-en	Prec.Met	Raw.Mats
LIN_SUM	2(0.95)	11(0.6)	12(0.22)	4(0.72)	5(0.95)	7(0.55)	-	5(1)
ENET_SUM	11(0.31)	14(0.29)	-	1(1)	12(0.53)	6(0.74)	14(0.26)	9(0.63)
RF_SUM	1(1)	8(0.73)	13(0.16)	12(0.27)	11(0.6)	-	5(0.79)	3(1)
PCR_SUM	4(0.93)	9(0.66)	11(0.38)	2(0.84)	4(0.96)	1(1)	13(0.29)	4(1)
PLS_SUM	9(0.79)	7(0.92)	7(0.91)	11(0.28)	3(1)	8(0.51)	16(0.22)	10(0.26)
MIO_SUM	-	-	5(1)	-	10(0.73)	-	11(0.36)	-
Panel C: Partially-egalitarian LASSO combination								
	Agric	Bev	Eng	Food	Met&Min	Non-en	Prec.Met	Raw.Mats
LIN_pLASSO	8(0.86)	6(0.97)	-	3(0.82)	7(0.93)	-	9(0.48)	8(0.9)
ENET_pLASSO	-	13(0.34)	-	5(0.72)	12(0.45)	-	4(0.94)	11(0.2)
RF_pLASSO	-	15(0.2)	-	-	-	-	-	-
PCR_pLASSO	10(0.39)	-	-	13(0.24)	13(0.48)	10(0.2)	7(0.65)	-
PLS_pLASSO	-	10(0.6)	10(0.46)	-	-	-	10(0.48)	-
MIO_pLASSO	-	-	9(0.74)	-	15(0.37)	-	8(0.59)	-

Notes: The table contains the results from the MCS procedure of Hansen et al. (2011) over our 18 models. Each entry in the table indicates the ranking the methods were given by the MCS procedure alongside its associated p-value in brackets. A '-' shows that the model is not in the set of superior models at α = 15%. We use the differential to all other models leading to the null hypothesis of Equation 26. One can also use a different null hypothesis using the average differential to every other model in the confidence set, agin suggested in Hansen et al. (2011). That approach would lead to similar results, but not necessarily the exact same.

The pLASSO models perform the worst with them not being included in most superior sets, because of the information discarding behaviour we mention. Out of the pLASSO models, only the LIN variant displays "superior" predictive ability and even on those cases it is on the bottom end of the sets. On the contrary, SUM models showcase average superior performance. They take the middle spots for most indices, but RF_SUM and PCR_SUM are above the rest as they manage to secure high-ranking spots for 4 of the 8 commodity price changes. A noteworthy observation is in regards to the poor performance

of the model of Wang et al. (2020) (LIN_EW) that gets always dominated by all other equally-weighted combinations besides the MIO.

Table 3: The $SR_{\%}$ financial gains

Panel A: Equally-weighted forecast combinations								
	Agric	Bev	Eng	Food	Met&Min	Non-en	Prec.Met	Raw.Mats
LIN_EW	0.38	-0.79	0.36	0.21	0.37	0.57	0.01	0.6
ENET_EW	0.54	-0.79	0.26	0.34	0.33	0.62	0.07	0.71
RF_EW	0.54	-0.82	0.26	0.26	0.49	0.64	0.21	0.75
PCR_EW	0.53	-0.77	0.34	0.31	0.36	0.65	0.21	0.7
PLS_EW	0.51	-0.8	0.3	0.28	0.39	0.66	0.18	0.73
MIO_EW	0.14	-0.95	0.07	0.05	0.2	0.4	0.09	0.43
Panel B: MSPE minimizing weights combination								
	Agric	Bev	Eng	Food	Met&Min	Non-en	Prec.Met	Raw.Mats
LIN_SUM	0.57	-0.97	-0.05	0.39	0.52	0.57	-0.1	0.67
ENET_SUM	0.42	-0.82	0.15	0.39	0.42	0.63	0.18	0.5
RF_SUM	0.61	-0.84	0.32	0.35	0.48	0.64	0.21	0.73
PCR_SUM	0.5	-0.86	0.26	0.39	0.42	0.66	0.25	0.67
PLS_SUM	0.5	-0.86	0.26	0.39	0.42	0.66	0.25	0.67
MIO_SUM	0.19	-1.03	0.04	-0.04	0.22	0.37	0.05	0.47
Panel C: Partially-egalitarian LASSO combination								
	Agric	Bev	Eng	Food	Met&Min	Non-en	Prec.Met	Raw.Mats
LIN_pLASSO	0.45	-0.85	0.05	0.34	0.42	0.53	0.12	0.62
ENET_pLASSO	0.41	-0.94	-0.06	0.42	0.28	0.59	-0.11	0.39
RF_pLASSO	0.4	-1.07	-0.65	-0.51	-0.1	0.52	-0.29	0.28
PCR_pLASSO	0.41	0.01	-0.02	0.29	0.58	0.52	1.9	0.34
PLS_pLASSO	0.35	-0.98	0.12	0.19	0.43	0.74	0.39	0.35
MIO_pLASSO	0.18	-0.99	0.01	-0.06	0.26	0.46	0.45	0.21

Notes: The table displays the $SR_{\%}$'s for the different models over the out-of-sample period from January 1991 until March 2023 (the last month does not have a t+1 forecast). The $SR_{\%}$'s are computed using Equation 31 and then multiplied by $\sqrt{12}$ to obtain the annualized Sharpe ratio values.

5.2 Financial Gains

5.2.1 Sharpe ratio gains over the out-of-sample period

We can see from the $SR_{\%}$'s depicted in Table 3, that our models generate considerable risk-adjusted returns to the historical average benchmark with the beverages commodity being the single exception. The EW and SUM models give very similar positive returns ranging between 0.01-0.73 hinting towards the fact that MSPE minimizing weights do not necessarily provide financial gains to the simple equal-

weighting scheme.

A similar remark can be made for the older model, LIN_EW. At the same time as it gets outperformed slightly in most cases, the difference is not as considerable as it was in terms of predictive power. A surprising result is the decent performance of the pLasso combinations. While the gap to the other combinations is clear in terms of out-of-sample predictive power, it seems that the risk-free adjusted returns generated by this combination type are similar to those of EW and and SUM. It is the only combination approach that leads to negative $SR_{\%}$'s for other commodities than beverages, but, simultaneously, it reaches the highest sharpe ratio gain of 1.9. An explanation for this occurrence, is that the risk-free rate dominates the forecasts from these methods. That means that the weights on the index are often set to 0 and it opts often to invest only on the risk-free asset leading to the idea that it generates positive risk-adjusted returns relative to the benchmark.

5.2.2 Certainty equivalent returns over the out-of-sample period

In Table 4 we provide the CER's over the out-of-sample period. The first thing we point out is that we were right in assuming that the $SR_{\%}$'s of the pLASSO models were misleading. We show that when accounting for the risk-free returns, the appeal of these portfolios drops tremendously. The additional value of the information from only the forecasts of the pLasso methods is below average at best when compared to EW and SUM. The other two combinations lead to high positive returns, ranging from 19 basis points to as high as 982. The MIO methods lead to the smallest CER, despite the combination method, again emphasizing the importance of keeping the information provided by each technical indicator. Furthermore, LIN_EW gets outperformed for almost all indices by the more complex methods like RF, PCR and PLS. To distinguish between EW and SUM in terms of more financial profitability, the differences in CER's do not occur frequently enough or at big enough amplitudes to draw a conclusion. The only conclusive remark we make is that the benefits of basing financial decisions on technical indicators do exist and are quite significant at that.

Table 4: The CER financial gains

Panel A: Equally-weighted forecast combinations								
	Agric	Bev	Eng	Food	Met&Min	Non-en	Prec.Met	Raw.Mats
LIN_EW	3.67	3.62	2.86	3.83	6.73	5.4	1.09	4.46
ENET_EW	5.79	5.22	5.84	5.94	6.72	6.03	1.7	5.74
RF_EW	5.88	5.13	6.79	5.4	9.45	6.33	4.69	5.93
PCR_EW	5.88	5.94	7.87	6	7.31	6.32	4.29	5.7
PLS_EW	5.7	5.26	7.58	5.91	8	6.54	4.21	5.93
MIO_EW	2.52	3.05	3.67	3.32	4.73	4.55	2.68	3.54
Panel B: MSPE minimizing weights combination								
	Agric	Bev	Eng	Food	Met&Min	Non-en	Prec.Met	Raw.Mats
LIN_SUM	5.77	2.67	1.26	6.52	9.82	6.09	0.19	5.52
ENET_SUM	4.72	4.27	4.94	6.22	8.44	6.03	4	4.03
RF_SUM	5.75	5.01	7.96	6.08	9.24	6.31	4.41	5.76
PCR_SUM	5.57	4.2	6.9	6.48	8.17	6.38	4.86	5.43
PLS_SUM	4.98	4.85	8.96	5.39	8.81	5.85	3.2	4.43
MIO_SUM	2.65	1.78	3.24	2.34	5.07	4.28	2.16	3.67
Panel C: Partially-egalitarian LASSO combination								
	Agric	Bev	Eng	Food	Met&Min	Non-en	Prec.Met	Raw.Mats
LIN_pLASSO	4.35	3.45	1.69	4.59	7.35	4.81	-0.51	4.92
ENET_pLASSO	3.5	1.22	1.02	4.38	3.36	2.85	-0.66	2.44
RF_pLASSO	2	-0.7	-6.75	-2.67	-0.88	-1.09	-1.94	-0.98
PCR_pLASSO	3.32	-0.4	1.07	0.86	2.34	1.46	-0.29	2.25
PLS_pLASSO	2.79	0.14	2.89	-0.23	0.09	1.63	-0.37	1.73
MIO_pLASSO	-0.01	0.89	1.39	-1.47	2.85	-0.39	-0.08	0.77

Notes: The table displays the CER's for the different models over the out-of-sample period from January 1991 until March 2023 (the last month does not have a t+1 forecast). The CER's are computed using 33 and then multiplied by 1200 to obtain the annualized certainty equivalent returns in percentage points.

6 Conclusions

In this paper we construct 18 forecasting models combining six methods and three forecast combination approaches. We train our models over our training set, from January 1982 until December 1990, using a set 105 technical indicators as potential explanatory variables. We asses how they compare out-of-sample against the historical average in terms of forecasting ability and possible financial profitability. The criteria we base our comparison upon are R^2_{oos} and CSSED for predictive prowess, plus $SR_{\%}$ and CER for monetary gains. We also compare the forecasts of the models altogether, using the MCS procedure, giving an ordered ranking of the superior models for each commodity index.

We conclude that the EW combinations provide the most accurate forecasts when using the R^2_{oos} as the evaluating measure, followed closely by the SUM. The pLASSO performs considerably worse than the other two, falling behind the benchmark at times for certain commodities, implying that all five technical rules are useful for predicting the dependent variables. The best forecasting models are RF_EW, PCR_EW and PLS_EW which obtain consistently the highest R^2_{oos}'s and are ranked in the top three by the MCS procedure, for most indices. In addition, we claim that the simple model of Wang et al. (2020) retains its positive performance, however, it considerably falls behind more complex methods and combinations.

In terms of financial gain, almost all models give positive $SR_{\%}$ and lead to significant risk-adjusted returns. The difference between EW and SUM combinations is no longer detectable as they both provide similar $SR_{\%}$'s and CER's. It goes to show that using the error minimizing weights ex-ante, does not directly impact future portfolio returns and a investor would get similar returns just by a simple forecast average. On the other hand, pLASSO gives the highest $SR_{\%}$'s, but the lowest CER's, often smaller than zero. This means that the portfolios created by these methods are dominated by the risk-free rate and the investor does not obtain much useful information by the addition of the pLASSO forecasting knowledge in terms of increasing individual utility.

One possibility for extending our paper is to test the novelty of our approach on different datasets, moving on to the field of cryptocurrencies or back at the realm of stock and/or bonds. The CSSED indicates that our models heavily outperform the benchmark during times of financial crises and higher market volatility and we want to see whether it is still applicable when using other, riskier datasets. It is interesting to consider the idea of creating forecasts for the eight commodity prices at once, making it smart to use a model that generates well to multi-dimensional spaces such as the support vector regression (SVR) deriven from the classification method of Vapnik and Chervonenkis (1974). The SVR gives the flexibility to define how much error is acceptable in the model, making it very useful for the individual risk-averse investor who has a stop-loss preference. Furthermore, we intuitively group the indicators variables into five groups based on their similarities, but one can rely on using deep learning to choose which variables to use per each regression. One alternative we provide is neural networks (NN) of Rosenblatt (1958), models that learn from the data itself, seeing patterns invisible to the naked eye. Lastly, other combination approaches can be tested, like the regression of Granger and Ramanathan (1984), and see if they lead to better forecasting accuracy or are more financially exploitable.

Appendices

A Replication of Wang et al. (2020)

In this section we replicate one of the main results from the paper of Wang et al. (2020). We restrict ourselves to the same sample period as used by them, January 1982 until December 2017 with the out-of-sample starting from January 1991. We depict the R^2_{oos}'s for each commodity index forecast generated by the five technical rules and the equally-weighted technical forecast in Table 5. We see some changes in the values obtained by specific technical rules, however, the equally-weighted portfolio is within a 0.5 value for each commodity index. We speculate the changes originate from a change in the base-year used in the World Bank's Website and we prove our claim when comparing the summary statistics of our dataset with the one used by Wang et al. (2020). Despite being restricted to the same sample period, the summary statistics change drastically especially in terms of non-normality. The summary statistics for the original paper's dataset are in their online Appendix [8], while for our paper we do not report the summary statistics for the restricted sample, but they resemble closely the values in Table 6.

Table 5: The replications

	Agric	**Bev**	**Eng**	**Food**	**Met&Min**	**Non-en**	**Prec.Met**	**Raw.Mats**
MOM	5.19**	3.34**	1.97**	3.70**	5.08**	6.24**	1.50**	6.64**
FR	3.10**	3.02**	2.00**	2.92**	6.11**	5.38**	0.65*	3.61**
MV	3.78**	2.45**	1.07*	1.36*	4.19**	5.65**	0.91*	4.87**
OSLT	6.06**	3.89**	1.77**	4.16**	5.10**	6.99**	1.57**	6.96**
SR	5.50**	3.90**	1.89**	4.74**	5.66**	6.26**	0.84*	8.21**
EW-T	5.00**	3.52**	1.82**	3.61**	5.45**	6.51**	1.16*	6.46**

Notes: The table contains the R^2_{oos}'s computed over the out-of-sample of Wang et al. (2020), from January 1991 until December 2017. The first five portfolio names are constructed as the technical rule abbreviation and are constructed in the same manner as in the original paper, while EW-T corresponds to the equally-weighted average of all five technical rule portfolios. The R^2_{oos}'s are computed using Equation 21 against the historical average benchmark. Lastly, we use the Clark and West (2007) method to test statistical significance and * denotes significance at the 5% significance level and ** at the 1% level.

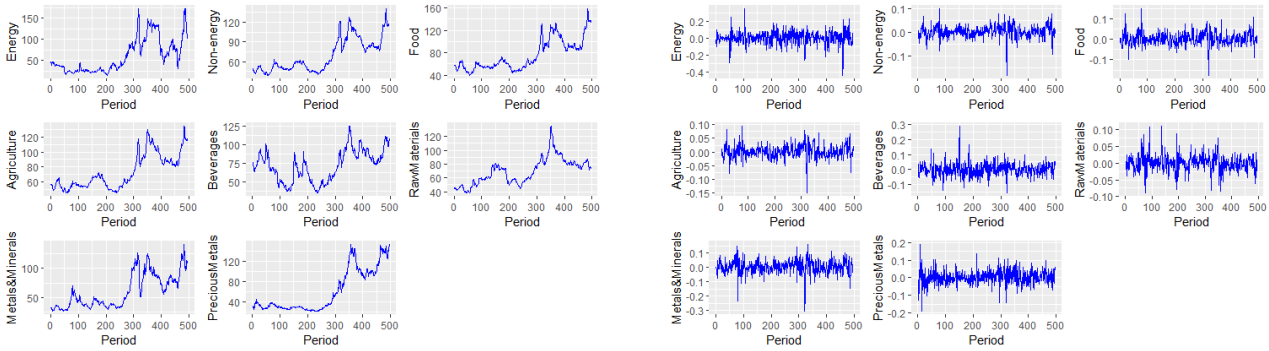
B Summary Statistics for commodity prices and Log price changes

Figures 2a and 2b depict the historical movements of commodity prices and ln price changes respectively. We see a similar increasing trend throughout time with certain peaks and troughs, but when taking the ln difference transformation, the series becomes stationary with a mean of almost 0, as shown in Table 7. In particular, the agriculture, non-energy and food indices display very similar movements and lead to some of the highest R^2_{oos}'s. Additionally, we see the correlation between a commodity's standard deviation

[8] Appendix A in https://www.sciencedirect.com/science/article/pii/S0169207019302286?casa_token=pNKEqCkLRIkAAAAA:Bc85z-TN1ZRQ13U2SrfWtTC6kL9iJUdp5fIoTEGKwzJj0e1rN95g2AGD4c6L9K6h87SG4bndkLjk#appSB

and its predictability, e.g raw materials has the lowest standard deviation leading to high R^2_{oos}'s, while precious metals is the most volatile and hence it consistently has the lowest R^2_{oos}'s across all 18 models. The standard deviations and other summary statistics of each commodity types are depicted in Table 6 and 7 for prices and ln price changes correspondingly.

Figure 2: Historical Movements of commodity prices and ln price changes



(a) Prices

(b) Ln price changes

Table 6: Summary Statistics for commodity price indices

Series	Mean	Median	Max	Min	St.Dev.	Skewness	Kurtosis	JB
Agriculture	72.86	63.98	134.078	44.87	22.68	0.79	2.44	59.87
Beverages	71.65	72.47	124.81	34.77	20.58	0.13	2.23	14
Energy	59.86	43.41	173.48	15.2	39.35	0.98	2.88	81.12
Food	74.79	63.2	159.04	39.65	26.85	0.93	2.85	73.24
Metals&Minerals	59.97	48	141.28	26.05	29.25	0.74	2.27	57.28
Non-energy	68.65	57.91	141.07	38.97	24.94	0.8	2.44	60.3
Precious Metals	60.14	35.33	153.29	21.84	40.81	0.86	2.24	74.87
Raw Materials	68.82	71.39	134.56	37.69	18.94	0.55	3.12	26.11

Notes: The table contains summary statistics for the price each commodity index as given in the World Bank's Website. The statistics are computed over the sample January 1982 until April 2023.

Table 7: Summary Statistics for commodity price indices

Series	Mean	Median	Max	Min	St.Dev.	Skewness	Kurtosis	JB
Agriculture	0.00	0.00	0.09	-0.15	0.02	-0.23	6.84	316.65
Beverages	0.00	0.00	0.29	-0.16	0.04	0.68	7.65	499.78
Energy	0.00	0.00	0.35	-0.44	0.08	-0.56	6.66	309.9
Food	0.00	-0.00	0.15	-0.18	0.03	0.11	6.54	266.65
Metals&Minerals	0.00	0.00	0.16	-0.3	0.05	-0.75	7.64	504.5
Non-energy	0.00	0.00	0.1	-0.18	0.03	-0.71	8.69	727.18
Precious Metals	0.00	-0.00	0.19	-0.19	0.04	-0.03	5.01	85.58
Raw Materials	0.00	0.00	0.11	-0.09	0.02	0.28	6.02	199.62

Notes: The table contains summary statistics for the ln price changes, using Equarion 1, for each commodity index. The statistics are computed over the sample January 1982 until April 2023.

C Technical indicator computation

We compute 105 indicators from five technical rules: momentum(MOM), filter(FR), moving average(MV), oscillator(OSC), support-resistance(SR); and explain how each of them are obtained.

MOM The momentum rule compares the current price with a k-period lagged price, according to Equation 34:

$$S_{t,MOM} = \begin{cases} 1, & \text{if } P_t \geq P_{t-k} \\ 0, & \text{if } P_t < P_{t-k}, \end{cases} \tag{34}$$

where k is the look-back period. We use values of k equal to 1, 3, 6, 9 and 12 resulting in five momentum indicators.

FR The filter rule compares the current price with the most recent local minima and maxima. It gives a buying/selling signal when the price is above/below a given percentage of the most recent low/high, given respectively in Equations 35 and 36:

$$S^{buy}_{t,FR} = \begin{cases} 1, & \text{if } P_t \geq (1 + \frac{\eta}{100}) * \min(P_{t-1}, P_{t-1}, ..., P_{t-k}) \\ 0, & \text{otherwise,} \end{cases} \tag{35}$$

$$S^{sell}_{t,FR} = \begin{cases} 1, & \text{if } P_t \leq (1 - \frac{\eta}{100}) * \max(P_{t-1}, P_{t-1}, ..., P_{t-k}) \\ 0, & \text{otherwise,} \end{cases} \tag{36}$$

where we choose the same values for the look-back period k (1,3,6,9,12) and the threshold, η, we set to 5 and 10, making for a total of 20 filter indicators.

MV The moving average rule gives buying/selling signals based on the short-term moving average over the last s days, $MA_{s,t} = \frac{1}{s}\sum_{i=0}^{s-1} P_{t-i}$ and the long-term moving average over the last l days, $MA_{l,t} =$

$\frac{1}{l}\sum_{i=0}^{l-1} P_{t-i}$. The buying signal is computed using the formula in Equation 37:

$$S_{t,MA} = \begin{cases} 1, & \text{if } MA_{s,t} \geq MA_{l,t} \\ 0, & \text{otherwise,} \end{cases} \tag{37}$$

where we let s,l = 1,3,6,9,12 with s¡l leading to a total of 10 moving average indicators.

OSLT The oscillator rule generates buying and selling signals if the price movements have been too rapid, based on the relative strength indicator (RSI) of Levy (1967), denoted in Equation 38:

$$RSI(k) = 100\,(\frac{U_t(k)}{U_t(k) + D_t(k)}), \tag{38}$$

where $U_t(k)$ and $D_t(k)$ are the upwards and downward price movements. We compute them using Equations 39 and 40:

$$U_t(k) = \sum_{j=0}^{k-1} \mathbb{1}(P_{t-j} - P_{t-j-1} \geq 0)(P_{t-j} - P_{t-j-1}), \tag{39}$$

$$D_t(k) = \sum_{j=0}^{k-1} \mathbb{1}(P_{t-j} - P_{t-j-1} < 0)|P_{t-j} - P_{t-j-1}|. \tag{40}$$

The oscillator indicators expect a reversal in trend, and are given in Equations 41 and 42:

$$S^{buy}_{t,OSLT} = \begin{cases} 1, & \text{if RSI } \leq 50 + \eta \\ 0, & \text{otherwise,} \end{cases} \tag{41}$$

$$S^{sell}_{t,OSLT} = \begin{cases} 1, & \text{if RSI } ¿ \ 50 + \eta \\ 0, & \text{otherwise,} \end{cases} \tag{42}$$

where we try the values 1,3,6,9,12 for k and 5, 10 for η making for a total of 20 oscillator indicators.

SR The final support-resistance rule generates buying/selling indicators by comparing the current price withe the local maxima/minima over the last k days. The buying and selling signals are as given in Equations 43 and 44 correspondingly:

$$S^{buy}_{t,SR} = \begin{cases} 1, & \text{if } P_t \geq (1 + \frac{\eta}{100}) * \max(P_{t-1}, P_{t-1}, ..., P_{t-k}) \\ 0, & \text{otherwise,} \end{cases} \tag{43}$$

$$S^{sell}_{t,FR} = \begin{cases} 1, & \text{if } P_t \leq (1 - \frac{\eta}{100}) * \min(P_{t-1}, P_{t-1}, ..., P_{t-k}) \\ 0, & \text{otherwise,} \end{cases} \tag{44}$$

where we set k = 1,3,6,9,12 and η = 1,2,3,4,5; resulting in 50 support-resistance indicators.

D The MSPE's of the out-of-sample forecasting

In the interest of reproducibility, we provide the actual MSPE values obtained for all 18 models across the eight commodity types. The results are given in Table 8 and obviously the match with the results for the R^2_{oos}, given in Table 1.

Table 8: The MSPE forecasting results

Panel A: Equally-weighted forecast combinations								
	Eng	Non-en	Agrc	Bev	Food	Raw.Mats	Met&Min	Prec.Met
LIN_EW	0.0062	0.00065	0.00054	0.0019	0.00087	0.00052	0.00219	0.00138
ENET_EW	0.00606	0.00062	0.00051	0.00188	0.00083	0.00049	0.00212	0.00137
RF_EW	0.00606	0.00061	0.0005	0.0019	0.00083	0.00049	0.00208	0.00136
PCR_EW	0.00603	0.00061	0.0005	0.00189	0.00084	0.00049	0.00213	0.00136
PLS_EW	0.006	0.00062	0.0005	0.00189	0.00083	0.00049	0.00209	0.00049
MIO_EW	0.00607	0.00064	0.00055	0.00197	0.00089	0.00053	0.0022	0.0014
Panel B: MSPE minimizing weights combination								
	Eng	Non-en	Agrc	Bev	Food	Raw.Mats	Met&Min	Prec.Met
LIN_SUM	0.00623	0.00064	0.00050	0.00194	0.00083	0.00050	0.00214	0.00141
ENET_SUM	0.00641	0.00064	0.00052	0.00194	0.00080	0.00051	0.00219	0.00149
RF_SUM	0.00619	0.00065	0.00049	0.00194	0.00084	0.00049	0.00215	0.00139
PCR_SUM	0.00618	0.00059	0.00050	0.00193	0.00082	0.00050	0.00214	0.00138
PLS_SUM	0.00616	0.00066	0.00051	0.00192	0.00084	0.00052	0.00211	0.00141
MIO_SUM	0.00615	0.00067	0.00056	0.00205	0.00091	0.00054	0.00224	0.00142
Panel C: Partially-egalitarian LASSO combination								
	Eng	Non-en	Agrc	Bev	Food	Raw.Mats	Met&Min	Prec.Met
LIN_pLASSO	0.00629	0.00065	0.00051	0.00191	0.00083	0.00050	0.00216	0.00139
ENET_pLASSO	0.00637	0.00068	0.00053	0.00194	0.00083	0.00052	0.00224	0.00159
RF_pLASSO	0.00655	0.00071	0.00055	0.00200	0.00094	0.00057	0.00235	0.00162
PCR_pLASSO	0.00632	0.00066	0.00053	0.00195	0.00089	0.00053	0.00226	0.00140
PLS_pLASSO	0.00624	0.00068	0.00055	0.00195	0.00091	0.00055	0.00230	0.00140
MIO_pLASSO	0.00620	0.00071	0.00056	0.00195	0.00091	0.00055	0.00224	0.00140

Notes: The table contains the MSPE computed over the out-of-sample period beginning from January 1991 until April 2023. The portfolio names are constructed as the forecasting method plus the forecast combination approach and are categorized into panels depending on said combination approach. The MSPE's are computed using Equation 22 against the actual values and we round up to five decimal places.

Table 9: The MCS procedure results within each combination method

Panel A: Equally-weighted forecast combinations								
	Agric	Bev	Eng	Food	Met&Min	Non-en	Prec.Met	Raw.Mats
LIN_EW	-	5(0.77)	6(0.34)	-	5(0.26)	-	-	-
ENET_EW	4(0.29)	1(1)	3(0.93)	2(0.98)	3(0.57)	4(0.63)	4 (0.16)	3(0.7)
RF_EW	1(1)	4(0.8)	5(0.63)	3(0.92)	1(1)	2(1)	2(0.99)	1(1)
PCR_EW	2(0.97)	2(0.95)	2(0.98)	4(0.28)	4(0.31)	1(1)	1(1)	2(1)
PLS_EW	3(0.88)	3(0.95)	1(1)	1(1)	2(1)	3(0.97)	4(0.65)	
					3(0.65)			
MIO_EW	-	-	4(0.89)	-	-	5(0.6)	-	-
Panel B: MSPE minimizing weights combination								
	Agric	Bev	Eng	Food	Met&Min	Non-en	Prec.Met	Raw.Mats
LIN_SUM	2(0.71)	4(0.99)	5(1)	3(0.34)	2(0.98)	2(0.69)	4(0.68)	3(0.8)
ENET_SUM	-	5(0.99)	6(0.31)	1(1)	6(0.46)	3(0.58)	6(0.16)	4(0.38)
RF_SUM	1(1)	3(1)	4(1)	-	4(0.93)	5(0.22)	2(1)	1(1)
PCR_SUM	3(0.56)	1(1)	3(1)	2(0.58)	3(0.97)	1(1)	1(1)	2(1)
PLS_SUM	4(0.41)	2(1)	2(1)	-	1(1)	4(0.36)	3(0.87)	-
MIO_SUM	-	-	1(1)	-	5(0.77)	-	5(0.68)	-
Panel C: Partially-egalitarian LASSO combination								
	Agric	Bev	Eng	Food	Met&Min	Non-en	Prec.Met	Raw.Mats
LIN_pLASSO	1(1)	1(1)	3(0.82)	1(1)	1(1)	1(1)	1(1)	1(1)
ENET_pLASSO	-	-	5(0.33)	2(0.41)	3(0.21)	4(0.16)	5(0.58)	-
RF_pLASSO	-	-	-	-	-	3(0.47)	-	-
PCR_pLASSO	-	-	4(0.52)	-	4(0.16)	2(0.82)	2(1)	-
PLS_pLASSO	-	2(0.38)	2(0.94)	-	-	3(0.47)	4(0.9)	-
MIO_pLASSO	-	-	1(1)	-	2(0.26)	-	3(1)	-

Notes: The table contain the results from the MCS procedure of Hansen et al. (2011) of the six forecasting methods within each forecast combination type. Each entry in the table indicates the ranking the methods was given by the MCS procedure alongside its associated p-value in brackets. A '-' shows that the model is not in the set of superior models at $\alpha = 15\%$. We use the differential to all other models leading to the null hypothesis 26. One can also use a different null hypothesis using the average differential to every other model in the confidence set, also suggested by Hansen et al. (2011). That approach would lead to similar results, but not necessarily the exact same.

E The MCS approach implemented within each forecast combination type

We implement the MCS approach of Hansen et al. (2011) within each combination approach to compare the different methods. The results are in Table 9. We clearly see the LIN_EW model of Wang et al.

(2020) being dominated throughout the eight commodities by all, but the MIO methods. In particular the RF_EW model is most consistently the best, followed closely by PCR_EW and PLS_EW. When moving to the SUM and pLASSO combinations, the linear regression method holds its own more, in fact it is the best out of all pLASSO methods. Within the SUM models, we see a similar story as with the EW, in regards to the best performing methods, but in a slightly different order. To be precise, PCR_EW is most often the best model, followed by PLS_EW and RF_EW. On the other hand, it seems that the benefits of more complex methods get undermined when combined with pLASSO as the usually better performing methods, RF, PCR and PLS, are more often not in the superior set rather than in. It goes to show that allowing for the possibility of discarding information from certain technical rules leads to worse forecasting ability in general and since LIN_pLASSO makes use of the most forecast series (by construction), it is the least affected.

F Another combination: Minimum eigenvector approach(MIN)

We choose to also implement the MIN combination approach, suggested by Hsiao and Wan (2014). Their approach is similar in estimation to the SUM approach we implement, but there is a clear difference. They do not compute the weights using the restriction of Markowitz (1952), given in Equation 14. Instead, they place the restriction on the dot product of the weight vector, as seen in Equation 45:

$$w^{j^T} w^j = \sum_{i=1}^{N} w_i^{j^2} = 1, \tag{45}$$

where w^j is the weight vector corresponding to the j-th method. We refer you to the paper of Hsiao and Wan (2014) to see the benefits of this approach and why they hold, but summarized the constraint of Equation 45 is less restrictive on the set feasible solutions than that of Equation 14.

The downside of the approach is that it inherently assumes equal predictive accuracy across the forecast series to be combined. Table 5 gives an indication towards that not being the case, as different rules obtain different R^2_{oos}'s. The forecasting results for the MIN approach are given in Table 10 from where we see underwhelming performance for each commodity type, with most models being outperformed by the benchmark. We opt to not conduct the CW test as most values are very close to 0 or negative and as such we do not expect to find proof of significant predictive ability. For the same reason we do not build any portfolios based on the MIN combination models.

Table 10: The results for the minimum eigenvector combination approach

Panel A: The out-of-sample R^2								
	Eng	Non-en	Agrc	Bev	Food	Raw.Mats	Met&Min	Prec.Met
LIN_MIN	0.41	-0.98	-0.41	0.12	-0.19	0.27	0.26	-0.56
ENET_MIN	-0.18	-1.68	-2.34	-0.97	-1.54	-0.15	-0.47	-0.43
RF_MIN	-2.88	-6.33	-4.41	-4.85	-3.59	-0.91	-2.73	-2.90
PCR_MIN	-0.26	-3.83	-2.36	-1.94	-1.03	-0.73	-1.37	0.07
PLS_MIN	-0.62	-2.78	-3.14	-3.14	1.71	0.70	-2.01	-0.91
MIO_MIN	0.71	-3.80	-3.07	-2.66	-3.21	-0.70	-2.68	-1.76
Panel B: The out-of-sample MSPE								
	Eng	Non-en	Agrc	Bev	Food	Raw.Mats	Met&Min	Prec.Met
LIN_MIN	0.00628	0.00070	0.00057	0.00196	0.00090	0.00056	0.00230	0.00140
ENET_MIN	0.00631	0.00070	0.00058	0.00198	0.00092	0.00056	0.00231	0.00140
RF_MIN	0.00648	0.00073	0.00059	0.00206	0.00093	0.00056	0.00236	0.00144
PCR_MIN	0.00632	0.00072	0.00058	0.00200	0.00091	0.00056	0.00233	0.00139
PLS_MIN	0.00634	0.00071	0.00058	0.00202	0.00089	0.00056	0.00235	0.00141
MIO_MIN	0.00626	0.00072	0.00058	0.00201	0.00093	0.00056	0.00236	0.00142

Notes: The table contains the R^2_{oos} and MSPE computed over the out-of-sample period beginning from January 1991 until April 2023 for the MIN combination portfolios. The portfolio names are constructed as the forecasting method plus the forecast combination approach. The R^2_{oos}'s are computed using Equation 21 against the historical average benchmark ,while the MSPE's are computed using Equation 22 against using the actual values and rounded up to five decimal places.

G Graphs and Figures

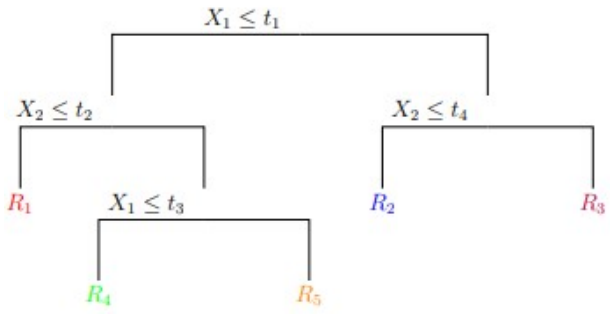
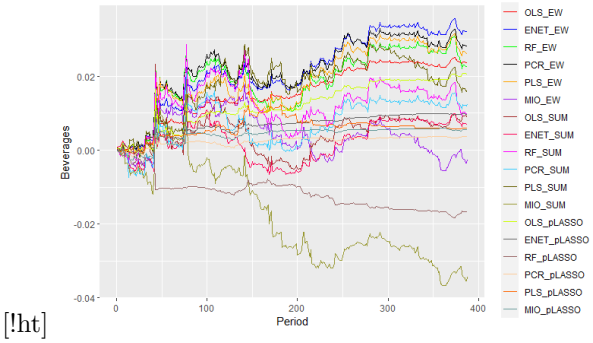


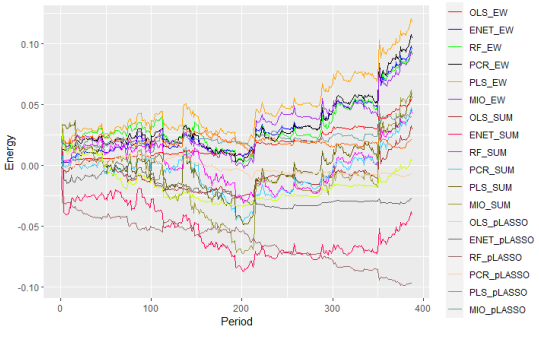
Figure 3: Graphical representation of a decision tree

Figure 4: The CSSED computed over-time against the historical average benchmark for the different commodity indeces



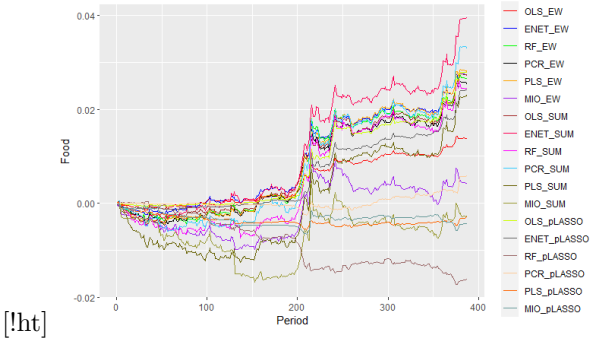
[!ht]

(a) Beverages Index



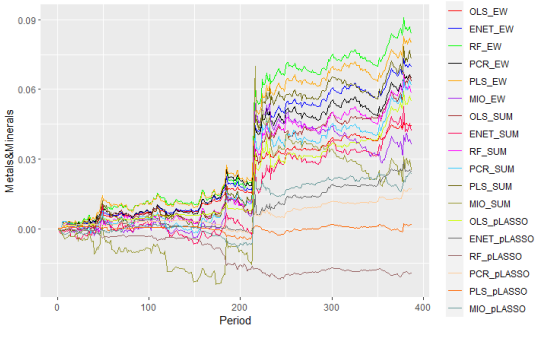
[!ht]

(b) Energy Index



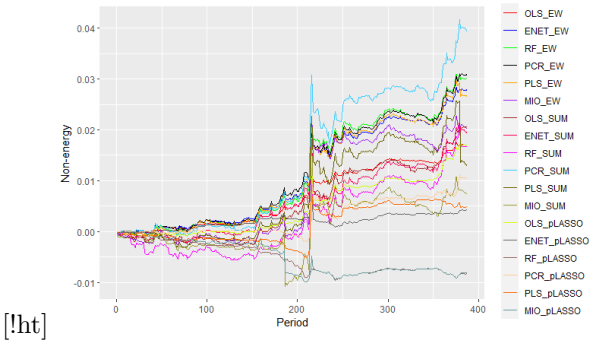
[!ht]

(c) Food Index



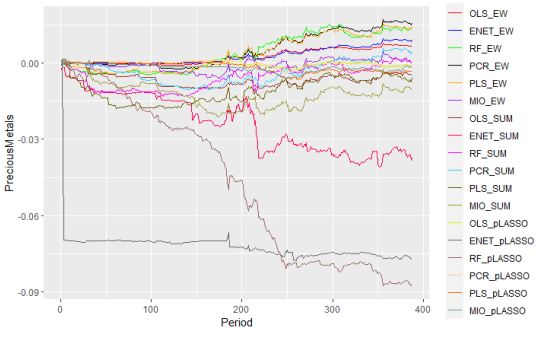
[!ht]

(d) Metals& Minerals Index



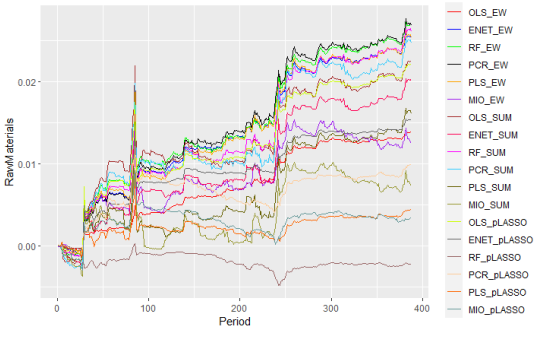
[!ht]

(e) Non-energy Index



[!ht]

(f) Precious Metals Index



[!ht]

(g) Raw Materials Index

$$\sum_{t=1}^{T}\left(y_t-\sum_{i=1}^{K}\beta_i f_{it}\right)^2+\lambda\sum_{i=1}^{K}\left|\beta_i-\frac{1}{K}\right|$$
$$=\sum_{t=1}^{T}\left(y_t-\bar{f}_t+\bar{f}_t-\sum_{i=1}^{K}\beta_i f_{it}\right)^2+\lambda\sum_{i=1}^{K}\left|\beta_i-\frac{1}{K}\right|$$
$$=\sum_{t=1}^{T}\left((y_t-\bar{f}_t)+\sum_{i=1}^{K}\left(\frac{1}{K}-\beta_i\right)f_{it}\right)^2$$
$$+\lambda\sum_{i=1}^{K}\left|\beta_i-\frac{1}{K}\right|$$
$$=\sum_{t=1}^{T}\left((y_t-\bar{f}_t)-\sum_{i=1}^{K}\delta_i f_{it}\right)^2+\lambda\sum_{i=1}^{K}|\delta_i|,$$

where

$$\delta_i=\beta_i-\frac{1}{K}\quad\text{and}\quad\bar{f}_t=\frac{1}{K}\sum_{i=1}^{K}f_{it}.$$

Hence, we obtain the egalitarian LASSO regression

$$y_t\rightarrow_{EgalLASSO} f_{1t},\ldots,f_{Kt},$$

by simply running the standard LASSO regression

$$(y_t-\bar{f}_t)\rightarrow_{LASSO} f_{1t},\ldots,f_{Kt}. \qquad (A.1)$$

Similarly, the egalitarian ridge can be implemented trivially by $(y_t-\bar{f}_t)\rightarrow_{Ridge} f_{1t},\ldots,f_{kt}$, in precise parallel with egalitarian LASSO implementation.

[width=8cm]

Figure 5: Derivation of the relationship between LASSO and eLASSO, taken from Diebold and Shin (2019)

H Code Documentation

The code consists of several R files, which conduct different parts of the analysis. Initially, we compute the 105 technical indicators using the data from the World Bank's Website. Furthermore, we compute the out-of-sample forecasts for each method and the equally-weighted combination while reporting the corresponding out-of-sample R squared ands MSPE's. Then, we compute the SUM, pLASSO and MIN combinations, again with the respective out-of-sample R squared and MSPE's. Next, we conduct the MCS procedure across the eight commodity types for all 18 models and also within each combination type seperately. We follow up by constructing the portfolios and ginding the sharpe ratio gains plus certainty equivalent returns. We also find the cumulative sum of squared prediction error difference for all models across the eight commodity types, before, finally estimating the CW statistic and the statistical significance of each model's forecasting power relative to the historical average benchmark.

References

H. Ahumada and M. Cornejo. Forecasting food prices: The case of corn, soybeans and wheat. *International Journal of Forecasting*, 32(3):838–848, 2016.

R. Alquist and L. Kilian. Time-varying risk premia and forecastable returns in futures markets. *Journal of Applied Econometrics*, 25(4):539–573, 2010.

D. Bertsimas, A. King, and R. Mazumder. Best subset selection via a modern optimization lens. *The Annals of Statistics*, 44(2):813–852, 2016.

H. Bessembinder and K. Chan. Time-varying risk premia and forecastable returns in futures markets. *Journal of Financial Economics*, 32(2):169–193, 1992.

L. Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.

L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

J. Campbell and S. Thompson. Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, 21(4):1509–1531, 2008.

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

G. Claeskens, J. Magnus, A. Vasnev, and W. Wang. The forecast combination puzzle: A simple theoretical explanation. *International Journal of Forecasting*, 32(3):754–762, 2016.

T. Clark and K. West. Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138(1):291–311, 2007.

F. Diebold and R. Mariano. Comparing predictive accuracy. *Journal of Business&Economic Statistics*, 13:253–263, 1995.

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

S. Dolatabaldi, S. Narayan, M. Nielsen, and K. Xu. Economic significance of commodity return forecasts from the fractionally cointegrated var model. *The Journal of Futures Markets*, 38(2):219–242, 2018.

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

R. Gerlach, B. Kowalski, and H. Wold. Partial least-squares path modelling with latent variables. *Analytica Chimica Acta*, 112(4):417–421, 1979.

N. Gospodinov and S. Ng. Commodity prices, convenience yields, and inflation. *The Review of Economics and Statistics*, 95(1):206–219, 2013.

C. Granger and R. Ramanathan. Improved methods of combining forecasts. *Journal of Forecasting*, 3(2): 197–204, 1984.

J. Groen and P. Pesenti. Commodity prices, commodity currencies, and global economic developments. *Commodity Prices and Markets*, 20:15–42, 2011.

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

T. Hastie, R. Tibshirani, and J. H. Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.

C. Heij, P. de Boer, P. H. Franses, T. Kloek, H. K. van Dijk, et al. *Econometric Methods with Applications in Business and Economics*. OUP Oxford, 2004.

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

C. Hsiao and S. Wan. Is there an optimal forecast combination? *Journal of Econometrics*, 178(2): 294–309, 2014.

J. Junttila, J. Pesonen, and J. Raatikainen. Commodity market based hedging against stock market risk in times of financial crisis: The case of crude oil and gold. *Journal of International Financial Markets, Institutions and Money*, 56:255–280, 2018.

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

P. Locke and P. Venkatesh. Futures market transaction costs. *Journal of Futures Markets*, 17(2):229–245, 1997.

H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

W. Massy. Principal components regression in exploratory statistical research. *Journal of the American Statistical Association*, 60(309):234–256, 1965.

K. P. Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

P. Newbold and C. Granger. Experience with forecasting univariate time series and the combination of forecasts. *Journal of the Royal Statistical Society*, 137(2):131–165, 1974.

R. Picard and D. Cook. Cross-validation of regression models. *Journal of the American Statistical Association*, 79(387):575–583, 1984.

D. Rapach, J. Strauss, and G. Zhou. Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies*, 23(2):821–862, 2009a.

D. Rapach, J. Strauss, and G. Zhou. Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies*, 23(2):821–862, 2009b.

F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):396–408, 1958.

W. F. Sharpe. Mutual fund performance. *The Journal of business*, 39(1):119–138, 1966.

Q. Shen, A. Szakamary, and S. Sharma. An examination of momentum strategies in commodity futures markets. *The Journal of Future Markets*, 27(3):227–256, 2007.

Q. Shen, A. Szakamary, and S. Sharma. Trend-following trading strategies in commodity futures: A re-examination. *Joural of Banking & Finance*, 34(2):409–426, 2010.

J. Stock and M. Watson. A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W.J. Granger*, pages 1–44, 1999.

J. Stock and M. Watson. How did leading indicator forecasts perform during the 2001 recession? *Federal Reserve Bank of Richmond Economic Quarterly*, 89(3):71–90, 2003.

R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B*, 58, 1996.

A. Timmerman. *Handbook of Economic Forecasting*, volume 1. ELSEVIER, 2006.

V. Vapnik and A. Chervonenkis. On the method of ordered risk minimization. i. *Automation and Remote Control*, 35(8):1226–1235, 1974.

Y. Wang, L. Liu, and C. Wu. Forecasting the real prices of crude oil using forecast combinations over time-varying parameter models. *Energy Economics*, 66:337–348, 2017.

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

S. Wold, K. Esbensen, and P. Gelaldi. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1-3):37–52, 1987.

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